Human-Centric Decision and Negotiation Support for Societal Transitions

Local Proceedings of the

24th International Conference on Group Decision and Negotiation &
10th International Conference on Decision Support System Technology

Vol II. People as active participants

Sérgio Pedro Duarte, Pascale Zaraté,
António Lobo, Boris Delibašić, Tomasz Wachowicz, Marta Campos Ferreira

University of Porto | June 2024, Porto, Portugal

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Faculdade de Engenharia da Universidade do Porto
Rua Dr. Roberto Frias s/n
4200-465 Porto
Portugal

Editors

Sérgio Pedro Duarte
University of Porto

Pascale Zaraté
University Toulouse Capitole - IRIT

António Lobo
University of Porto

Boris Delibašić
University of Belgrad

Tomasz Wachowicz
University of Economics in Katowice

Marta Campos Ferreira
University of Porto

Honorary Chairs

Adiel Teixeira de Almeida	Universidade Federal de Pernambuco, Brazil
Fatima Dargam	SimTech-Reach Consulting, Austria

Conference Chairs

Sérgio Pedro Duarte	University of Porto, Portugal
Pascale Zaraté	University Toulouse Capitole - IRIT, France

Program Committee Chairs

António Lobo	University of Porto, Portugal
Boris Delibašić	University of Belgrad, Serbia
Tomasz Wachowicz	University of Economics in Katowice, Poland

Organizing Committee

Carlos Rodrigues	University of Porto, Portugal
Marta Campos Ferreira	INESC TEC, University of Porto, Portugal
Sara Ferreira	University of Porto, Portugal
Telmo Botelho	Faculty of Engineering, University of Porto, Portugal

Program Committee

Abdelkader Adla	University of Oran, Algeria
Ana Paula Costa	Federal University of Pernambuco, Brazil
Antonio de Nicola	ENEA, Italy
Carolina Lino Martins	UFPE/UFMS, Brazil
Danielle Costa Morais	Federal University of Pernambuco, Brazil
Daouda Kamissoko	IMT Mines - Albi, France
Ewa Roszkowska	Bialystok University of Technology, Poland
François Pinet	INRAE, France
Fuad Aleskerov	HSE, Russia
Gert-Jan de Vreede	University of South Florida, USA
Haiyan Xu	Nanjing University of Aeronautics and Astronautics, China
Hannu Nurmi	University of Turku, Finland
Jason Papathanasiou	University of Macedonia, Greece
Jorge Freire de Sousa	University of Porto, Portugal
José Moreno-Jiménez	Universidad de Zaragoza, Spain
José Pedro Tavares	University of Porto, Portugal
Konstantinos Vergidis	University of Macedonia, Greece
Liping Fang	Toronto Metropolitan University, Canada

Luís Dias	University of Coimbra, Portugal
Mareike Schoop	University of Hohenheim, Germany
María Teresa Escobar	Universidad de Zaragoza, Spain
Marta Campos Ferreira	University of Porto, Portugal
Masahide Horita	University of Tokyo, Japan
Michael Filzmoser	Vienna University of Technology, Austria
Muhammed-Fatih Kaya	University of Hohenheim, Germany
Nannan Wu	Nanjing University of Inf. Science and Technology, China
Pavlos Delià s	International Hellenic University, Greece
Przemysław Szufel	Warsaw School of Economics, Poland
Rosaldo Rossetti	University of Porto, Portugal
Rudolf Vetschera	University of Vienna, Austria
Sandro Radovanovic	University of Belgrade, Serbia
Sara Ferreira	University of Porto, Portugal
Sean Eom	Southeast Missouri State University, USA
Shawei He	Nanjing University of Aeronautics and Astronautics, China
Teresa Galvão	University of Porto, Portugal
Thiago Sobral	University of Porto, Portugal
Tomasz Szapiro	Warsaw School of Economics, Poland
Tung X. Bui	University of Hawaii, USA

Reviewers

Eduarda Frej	Michael Filzmoser	Carla Fernandes	Hsiangchu Lai
Alexandre Leoneti	Kevin Li	Tânia Fontes Joana Hora	Sara Martins
Yu Han	Peng Sun	Marc Kilgour	Fábio Moreira
Yu Maemura	Rustam Vahidov	Sabine Theresia Koeszegi	Jinmuzi Zhang
Junjie Wang	Juliana Carvalho		

Conference organization

Associations and Organizations

Organization

Table of Contents

Preface

The collection of papers results from the papers presented at the joint conference that gathered the 24th International Conference on Group Decision and Negotiation (GDN 2024) and the 10th International Conference on Decision Support System Technology (ICDSST 2024), held in Porto, Portugal, from 3–5 June, 2024. Both conferences were under the main theme "Human-centric decision and negotiation support for societal transitions". The event was organized by a local team from the Faculty of Engineering of the University of Porto, in collaboration with both the EURO Working Group on Decision Support Systems (EWG-DSS) and the Group Decision and Negotiation section from INFORMS.

This joint conference happened ten years after the last conference the two groups collaborated to organize, which was held in Toulouse, France, with a special focus: Group Decision Making and Web 3.0. That year marked the beginning of the ICDSST conferences. The EWG-DSS series of International Conferences on Decision Support System Technology (ICDSST) was planned to consolidate the tradition of annual events organized by the EWG-DSS in offering a platform for European and international DSS communities, comprising the academic and industrial sectors, to present state-of-the-art DSS research and developments, to discuss current challenges that surround decision-making processes, to exchange ideas about realistic and innovative solutions, and to co-develop potential business opportunities.

The GDN Section of INFORMS emerges from evolving, unifying approaches to group decision and negotiation processes. These processes are complex and self-organizing involving multiplayer, multicriteria, ill-structured, evolving, dynamic problems. In defining the domain of group decision and negotiation, the term "group" is interpreted to comprise all multiplayer contexts. Thus, organizational decision making support systems providing organization-wide support are included. Group decision and negotiation refers to the whole process or flow of activities relevant to group decision and negotiation, not only to the final choice itself, for example scanning, communication and information sharing, problem definition (representation) and evolution, alternative generation and social-emotional interaction. Descriptive, normative and design viewpoints are of interest. Thus, Group Decision and Negotiation deals broadly with relation and coordination in group processes.

Human-Centric Decision and Negotiation Support for Societal Transitions

The topic selected for the event reflected the interests of both groups and, as usual, the current world scene and research trends. In a time when technology is rapidly evolving, decision-makers face two major challenges: (1) using technology to improve the decision process, and (2) ensuring that decisions really support the best interests of the actors involved. On the one hand, the evolution of machine learning and AI offers incredible benefits; on the other hand, we as technology creators must ensure that humans remain the main beneficiaries of new services, software, and policies.
The transition period society is going through has brought even more complexity to decision processes, by increasing uncertainty regarding the future. Whatever is our research focus (climate, energy, AI, automation, information and communication technology, etc.), change, transition, and challenges are recurrent. Add uncertainty to the mix and we have highly complex decision processes, with several interested actors and multiple levels of goals. This recurring uncertainty has impacts on economics, employment, demographics, politics, and other societal concerns.

Accordingly, the topics promoted discussions on the human and technological aspects of decision-making processes to build bridges between two domains:

1. Technology as a support tool: from the technological perspective, research must demonstrate that technology can be trusted and that proposed solutions are safe, inclusive, and fair.

2. People as active participants: from the human perspective, research should ensure that humans remain at the center of the decisions, with participatory and negotiation processes that promote co-creation and co-design of technology, services, and regulations. Such reliable decision processes increase trust and fairness of decisions.

GDN Streams

1. Conflict Resolution

Strategic conflict arises whenever humans interact, individually or in groups. Recently-developed methodologies and techniques that can help analysts understand strategic conflicts and provide strategic support to negotiators have been of great benefit to decision makers. New theoretical issues are now being explored, and at the same time new software systems are making modeling easier and analytical results clearer. Both theoretical and practical approaches have been used to study strategic conflicts in diverse areas including environmental management, global warming, energy, the food crisis, economic disparities, international trade and aging infrastructure.
The main objective of the Stream on Conflict Resolution is to provide a forum for discussion of recent advances in the development of formal conflict resolution techniques and their insightful application across a range of domains.

Stream organizers
Liping Fang, Toronto Metropolitan University, Toronto, Canada
Keith W. Hipel, University of Waterloo, Waterloo, Canada
Marc Kilgour, Wilfrid Laurier University, Waterloo, Canada

2. Negotiation Support Systems and Studies (NS3)

Electronic negotiations are nowadays common business practice. Dedicated electronic negotiation support systems (NSSs) enable complex, asynchronous, and dislocated negotiations.
NS3 Stream presents papers that (1) help bridging the gap between the vast amount of work on face-to-face negotiations and electronic negotiations as well as decision and negotiation aids embedded in negotiation processes; (2) that focus on the design and use of tools for decision support, communication support, document management, or conflict management for the negotiators and mediators in electronic negotiation processes

Stream organizers
Mareike Schoop, University of Hohenheim, Germany
Rudolf Vetschera, University of Vienna, Austria
Muhammed-Fatih Kaya, University of Hohenheim, Germany

3. Preference Modeling for Group Decision and Negotiation

A variety of methods, techniques, and normative models, mainly derived from multiple criteria decision making (MCDM) and game theory, may be used to support groups of negotiators and decision makers (DM) in defining their goals, eliciting preferences, and building the negotiation offers' scoring systems. The latter is fundamental for providing the groups with reliable decision support throughout the entire negotiation or group decision making process. However, many factors, such as cognitive issues, formal knowledge, and DM skills, may influence the actual use of scoring systems. Therefore, there is a constant need for redesigning the existing methods and designing new ones that allow for accurate preference modeling and elicitation for group decision and negotiation (GDN) process in a particular decision-making context, given the DMs' limitations regarding information processing and all formal and behavioral issues involved.
The main goal of this stream is to create a forum for scientists, researchers, and practitioners working on the topic of preference modeling for GDN that will allow them to exchange their experience and knowledge and discuss the recent developments and results of their research.

Stream organizers
Danielle Morais, Federal University of Pernambuco, Brazil
Tomasz Wachowicz, University of Economics in Katowice, Poland

4. Collaborative Decision Making

Making a decision for a group engaged in a common task is a difficult challenge. There are several kinds of group decision making processes. This stream addresses Collaborative Decision Making processes. By Collaborative Decision Making processes, we intend that involved participants must pool their efforts in order to define and work on the achievement of a common goal. They have to integrate multiple points of view which reveal to be difficult. They have to work together, although not necessarily in the same place or at the same time. Decisional processes are then complex and involve a non-closed set of actors. The difficult point for decision-makers is to make a balance between their own preferences and the building of common preferences within the group. One direct application in the daily life of such Collaborative Decision-Making processes can be implemented through the e_democraty which is defined as a form of government in which all adult citizens are presumed to be eligible to participate equally in the proposal, development, and creation of laws.
The purpose of this stream is to allow researchers to present methodologies, mathematical models, and software supporting Collaborative Decision-Making processes.

Stream organizers
Pascale Zaraté, Toulouse Capitole University, IRIT, France
Guy Camilleri, Toulouse 3 University, IRIT, France

5. Network Analysis of Decisions in Groups

We consider the topics related to the interaction among participants in groups, and try to reveal key players in such situations. The stream studies international conflicts, terrorist activity, decisions during pandemics, the problems of food security, migration of people because of tensions in the Middle East, etc., but not only these topics.
The session includes the papers that contain recent results obtained by research teams from academia and industry.

Stream organizers
Fuad Aleskerov, Professor, HSE University, Moscow, Russia
Alexey Myachin, Professor, HSE University, Moscow, Russia

6. Responsible NSS in the age of Generative AI

The ubiquitous growth of generative AI has generated highly intelligent and effective applications in virtually every aspect of human activities, including support for individual or institutional decision-making and negotiation. At the same time, AI-based applications have raised a global concern about its unintended consequences. Researchers and administrators in virtually all academic disciplines and practitioners' circles are debating on the future of generative AI in their respective field. This paper stream addresses the opportunities and limitations, concerns and risk of adopting generative AI – defined in its broadest sense – in the context of group decision and negotiation.

Stream organizers
Tung Bui, Matson Navigation Co. Professor
Shidler College of Business, University of Hawaii, Honolulu, USA

7. Risk evaluation and negotiation strategies

The complex process of determining the value of the identified hazards and estimated risks to those affected is often needed to model and analyze in practice. Negotiation strategies arises when conflicts have happened. The main goal of the Stream of Risk Evaluation and Negotiation Strategies is to discuss the current theories and applications and explore the future development related the stream.

Stream organizers
Haiyan Xu, Nanjing University of Aeronautics and Astronautics, China
Shawei He, Nanjing University of Aeronautics and Astronautics, China
Shinan Zhao, Jiangsu University of Science and Technology, China

Keynote speakers

Argumentation-based Deliberation: Foundations and Challenges, by Leila Amgoud

Recognized as vital in a group decision-making process, deliberation allows stakeholders discussing and reaching agreements on controversial issues before making ultimate decisions. It brings several benefits, one of which is ensuring well-informed and well-accepted decisions. Its backbone is argumentation, which consists of justifying claims by arguments, i.e., reasons behind claims. The greatest challenges facing deliberation systems are identifying, evaluating, and aggregating large sets of interacting arguments, generally of disparate types, and solving potential disagreements between stakeholders. In this talk, I will introduce abstract argumentation frameworks for deliberation, their formal foundations, and discuss their possible impacts on group decisions.

Leila Amgoud is a senior researcher at the French National Centre for Scientific Research (CNRS), a member of the IRIT Lab in Toulouse, and a deputy director of the same lab. She holds a M.Sc. (1996), a PhD (1999) and an Habilitation à Diriger des Recherches (2009) in Computer Science from the Paul Sabatier University, Toulouse, France. She is currently an Associate Editor of the Artificial Intelligence Journal and the Journal of Argument and Computation. She has been an EurAI Fellow since 2014. Her research interests are centered around knowledge representation and reasoning. She is specifically interested in argumentation-based approaches for reasoning and decision making under uncertainty, and more recently in explainability of AI models. She holds two chairs on "explainability" and "argumentation-based deliberation" at the interdisciplinary Artificial and Natural Intelligence Toulouse Institute (ANITI).

Supporting Advanced Analytics Practice with Gen AI, by Pedro Amorim

The generative AI hype, with all its glare, is overshadowing advanced analytics, which is a more straightforward approach to delivering value to businesses by improving decision-making (e.g., pricing, assortment, or distribution decisions). In this presentation, we'll discuss how companies should refocus on improving decision-making using advanced analytics, such as predictive and prescriptive approaches, while acknowledging the extra effectiveness and efficiency that generative AI may deliver to this endeavor. This path requires us to go beyond buzzwords and understand the technical complementariness of the different technologies.

Pedro Amorim is a seasoned professional bridging the realms of academia and analytics. Holding a Certified Analytics Professional and a Ph.D. in Industrial Engineering and Management, he boasts over a decade of hands-on experience leading analytics projects. Pedro is a Professor at the Faculty of Engineering of the University of Porto and Porto Business School, and he is a Co-founder and partner at LTPlabs. His multifaceted expertise underscores his commitment to advancing the field, both through practical applications and academic mentorship.

A GDN Odyssey, by Daniel Druckman

In this keynote address I take the audience on a journey through 27 years of publishing articles in GDN. My first article appeared in a 1997 special issue on "The Logic of Comparative Negotiation Analysis." This was an analysis in search of key dimensions of international negotiation. It was followed in 1999 by an article that honed in on Fred Ikle's taxonomy of negotiation, producing strong empirical support for his theory and a gracious thank you letter from Fred. The first of many articles on automated mediation began with a 2002 article that described the Negotiator Assistant system. This was followed in 2004 and then in 2014 by ambitious experimental evaluations of screen vs. human mediation. This stream culminated in our 2021awared winning article on robot meditation. The upshot is that robots produced more and better agreements than any of the other formats.

A slight detour from technical articles to the role of emotions resulted in a highly cited special issue that showed the many ways that emotions influence negotiation. Nine years later we produced another special issue on justice and fairness with a mix of quantitative and qualitative articles in a variety of settings. About that time I was working on positive affirmation as a technique for improving the chances of getting better agreements (2019) and on how negotiators react to turning points with matching or mismatching strategies. Most recently, in 2022, we answered the question: Does trust matter in negotiation? The answer is an emphatic yes.

This GDN journey spanned topics at the center of our field, from types of international negotiations to the importance of trust in any negotiation. The key insights that emerged reinforce our claims about the value of doing systematic research with practical implications. In summing up, I will try to push the field further in directions that broaden our understanding of these processes and contexts as well as strengthen GDN's contribution to the social and informational sciences.

Daniel Druckman is Professor Emeritus of Public and International Affairs at George Mason's Schar School of Policy and Government and an Honorary Professor at Macquarie University in Sydney and at the University of Queensland in Brisbane Australia. Two of his books, Doing Research: Methods of Inquiry for Conflict Analysis (Sage, 2005) and, with Paul F. Diehl, Evaluating Peace Operations (Lynne Reinner, 2010) received the outstanding book award from the International Association for Conflict Management (IACM). His co-authored article on robot mediation (2021) received a best article of the year award from GDN. He also received lifetime achievement and Rubin Theory to Practice awards from the IACM in 2003 and 2018 respectively. He also received lifetime achievement awards from the Novancia Business School in Paris in 2016 and from the Schar School of Policy and Government in 2023.

Round table

A round table was held under the conference's main theme *Human-Centric Decision and Negotiation Support for Societal Transitions.*
While technology allows us to make better and fast decisions, societal changes need the active participation from people. On the one hand, society is fast changing, on the other hand, the solution to a sustainable world seems to be a slow life style. Population growth demands for more resources, but there not enough resources. Policymakers must decide fast, still decisions must be based on real information.
There is no doubt we need technology. There is no doubt technology improves our lives.
There are some doubts on how much technology we need. There are some doubts on

The round table addresses the questions:
Are "human-centric" decisions really centered on humans? Are algorithms supporting our decisions or silently choosing on our behalf?

Moderated by Sérgio Pedro Duarte with the participation of:

- Fátima Dargam (EWG-DSS)
- Liping Fang (GDN Section)
- Isabel Paiva de Sousa (Porto Business School)

Doctoral Consortium

This event provides an opportunity for graduate students to explore their research interests in an interdisciplinary workshop under the guidance of a panel of distinguished experts in the field. During the DC, students present their research work in a relaxed and supportive environment, receive feedback and suggestions from experienced faculty members, discuss concerns about research, and network with faculty, peers, and future colleagues.

Doctoral Consortium Committee

Liping Fang (Chair), Toronto Metropolitan University, Canada
Masahide Horita, University of Tokyo, Japan
Marc Kilgour, Wilfrid Laurier University, Canada
Shaofeng Liu, University of Plymouth, UK
José María Moreno-Jiménez, University of Zaragoza, Spain
Leandro Rêgo, Federal University of Ceará, Brazil

Group Decision and Negotiation
- General stream -

Conference papers and extended abstracts

Buyer’s Choice of a Seller Using Smart Contracts

Elmira Mohammadhosseini Fadafan [1*] and Rudolf Vetschera [1]

[1] Department of Business Decisions and Analytics, University of Vienna, Vienna, Austria
[*] elmira.mohammadhosseini.fadafan@univie.ac.at

Abstract

Contractual relationships between buyers and sellers can be disrupted by unanticipated shocks to attributes of the exchanged good or service; in manufacturing, such relationships often involve one buyer of components or intermediate goods and many potential sellers. We study the buyer's selection of a seller given the option to initially agree on a smart contract which, in the advent of such unanticipated shocks, automatically adjusts the exchange price. Our benchmark analysis focuses on the case where a positive potential shock raises attribute values for both contracting parties, implying that the seller benefits more than the buyer from executing the original contract at the agreed exchange price. Taking the perspective of the buyer, we vary the shock and utility parameters to arrive at conclusions regarding the determinants of smart contract dominance in random buyer-seller matches. One of the key issues analysed in this paper is the possibility that after the potential shock, another seller might be better, and this argument encourages the buyer to switch the seller. For the case of the Nash bargaining-solution, we further investigate the impact of increasing the number of utility-generating attributes on the buyer's choice of a seller and the possible switch rates.

Keywords: Smart Contract Dominance; Seller Selection; Unexpected Shocks; Efficient Contracting.

1. Introduction

We assume the buyer, anticipating possible bargaining outcomes, will end up with an optimal contract from one of the sellers which gives her/him the highest utility value under the bargaining solution that typically results from negotiations with this seller. For example, if the buyer knows that bargaining with seller *A* typically results in the Nash bargaining solution and bargaining with seller *B* in the Kalai-Smorodinsky solution, the buyer will select *A* if the Nash outcome with that seller is better for the buyer than the Kalai-Smorodinsky outcome of bargaining with seller *B*. However, if a potential shock is realized, which has not been anticipated in the original contract, the outcomes may change, and the actual results might be different from those specified in the contract. In the scenarios that we analyse here, such changes create additional value, but are initially beneficial only for one side, while decreasing the utility of the other side.

A smart contract that automatically adjusts the transaction price can be used to redistribute value so that both parties benefit from exploiting the random effect. By specifying these adjustments ex ante (before the contract is executed or random shocks are known), a smart contract can avoid costly re-negotiations and conflicts between the contracting parties.

Still, it is possible that ex post, the best optimal bargaining solution from one of the sellers is no longer optimal and that contracting with another seller might lead to a higher utility of the buyer. Termination of existing contracts typically leads to transaction costs (e.g., penalties to be paid for dissolving a contract), but such a switch might be beneficial for the buyer even when considering such additional transaction costs. We therefore also study how often such changes could occur.

We study these questions using simulation of a smart contract model developed in previous research (Mohammadhosseini Fadafan, Utility improvements using smart contract, 2021); (Mohammadhosseini Fadafan & Vetschera, Dominant Smart Contracts Based on Major Bargaining Solutions, 2023). In the simulation, we use different preferences of the parties (represented by their utility functions) as well as magnitudes of shocks. We compare the outcomes of our proposed smart contract with those which would result if executing the original contract for a variety of problem settings.

Although, there are many studies concerning the profit-sharing and revenue-sharing in the supply-chain with disruptions (Cachon & Lariviere, 2005); (Krishnan & Winter, 2011); (Lei, Li, & Liu, 2012); (Cao, Wan, & Lai, 2013), the proposed model in this paper is based on different settings and addresses issues caused by unexpected shocks in the contract outcomes. The smart model in this paper proposes a sharing rule that distributes the additional total value created by the unforeseen shocks, which was not observable ex-ante without the discovery of this model.

2. Conclusions

In this work, we develop a model of smart contracts which by automatically adapting prices to unanticipated shocks to the value of attributes of the exchanged products or services avoids renegotiation, enforcement costs, and lasting legal conflicts with potentially valuable sellers. We construct a scenario in which potential shocks to the attribute values change the original contract outcomes for contracting parties. The average analysis from the numerical examples and our simulations show that there is a possibility that the buyer is better off by switching, ex-ante, from the original seller to some other seller, or ex-post, after realization of the shocks. Given these solutions, we perform multiple simulations using random parameters and shock magnitudes. Specifically, we run a very large number of simulations for different randomly chosen parameters to estimate the average buyers' switch rate to the other sellers and comparing the average ratios. These results show that based on the buyer's perspective, this average analysis may derive criteria to choose an ex ante optimal seller.

References

Cachon, G., & Lariviere, M. (2005). Supply Chain Coordination with Revenue-Sharing Contracts: Strengths and Limitations. *Management Science, 51*(1), 30-44. doi:http://www.jstor.org/stable/20110305

Cao, E., Wan, C., & Lai, M. (2013). Coordination of a supply chain with one manufacturer and multiple competing retailers under simultaneous demand and cost disruptions. *International Journal of Production Economics, 141*(1), 425-433. doi:https://doi.org/10.1016/j.ijpe.2012.09.009

Krishnan, H., & Winter, R. (2011). On the role of revenue-sharing contracts in supply chains. *39*(1), 28-31. doi:https://doi.org/10.1016/j.orl.2010.10.007

Lei, D., Li, J., & Liu, Z. (2012). Supply chain contracts under demand and cost disruptions with asymmetric information. *International Journal of Production Economics, 139*(1), 116-126. doi:https://doi.org/10.1016/j.ijpe.2011.11.031

Mohammadhosseini Fadafan, E. (2021). Utility improvements using smart contract. *SSRN*. doi:https://ssrn.com/abstract=4181621

Mohammadhosseini Fadafan, E., & Vetschera, R. (2023). Dominant Smart Contracts Based on Major Bargaining Solutions. *Group Decision and Negotiation, 33*, 327–350. doi:https://doi-org.uaccess.univie.ac.at/10.1007/s10726-023-09863-9

The Effect of Overconfidence on the Use of Recommender and Decision Support Systems

Ayşegül Engin [1*]

[1] Department of Business Decisions and Analytics, University of Vienna, Oskar Morgenstern Platz 1, Vienna, 1090, Austria
[*] ayseguel.engin@univie.ac.at

Abstract

To benefit from digitization, decision support systems and recommender systems find wide business-to-business applications with the increasing digitization. To ensure that organizations benefit from implementing such systems, how employees interact with them as users is crucial. Certainly, modifications to system elements, such as design elements, can improve usability by mitigating the gap between what the user can work with and what the system offers. However, cognitive biases such as hindsight bias and overconfidence bias affect the usability of the system and cannot be mitigated through a system element modification. The present study focuses on the effect of overconfidence as a consequence of hindsight bias on the use of personalized recommender systems and a more general decision support system. This study aims to contribute to the literature on recommendation systems by showing that users benefit from recommender systems depending on their cognitive biases differently. Using an experimental approach, the study looks into the use of recommender systems and general decision support in repeated decisions with higher order structures. Because of the higher-order structure, the designed decision situation for this study entails uncertainty. To disentangle decision behavior from outcome, which is affected by the uncertainty posed by decision situation, the study uses a reinforcement learning model to describe decision-making behavior of users.

Keywords: Behavioral OR; Recommender systems; Decision support systems; Reinforcement learning; Computational model

1. Introduction

Recommender systems are information filtering systems that provide users with suggestions based on their preferences or past behavior for a given decision situation (Ricci et al., 2021). One effective implementation strategy for the recommender systems is to overcome the online information overload (Zhang et al., 2019). Users who lack resources to gather information or simply do not want to spend time on a particular decision benefit from these systems daily.

Research on recommendation systems is primarily focuses on developing and using algorithms to improve recommendations. While it is essential to develop sophisticated recommender systems, it is equally important to consider users' perception of the system and their interaction with it. Recent research emphasizes the importance of understanding the interaction between users and sophisticated decision support systems (e.g. Käki et al., 2019; Nguyen et al., 2021; Khosrowabadi et al., 2022; Sroginis et al., 2023; Caro and de Tejada Cuenca, 2023).

Users do not only provide their inputs to the system and follow its recommendations (Sroginis et al., 2023); but also adjust their inputs (Käki et al., 2019; Khosrowabadi et al., 2022). Users may even choose to ignore the recommendations of the system, especially if they do not trust it (Caro and de Tejada Cuenca, 2023). This research is motivated by the literature that highlights how users' engagement with decision support is contingent on their individual characteristics and perceptions of the system. While adjustments of some users on the input may improve output (Nguyen et al., 2021), others may worsen it (Käki et al., 2019). Consequently, biased user input leads to unsatisfactory (Chen et al., 2023; Forouzandeh et al., 2021) and in some cases unfair (Gharahighehi et al., 2021) recommendations.

This study suggests that user characteristics can also play a significant role in their interaction with a decision support system. The study examines whether a personalized recommender system can help users learn higher-order structures of environments, such as interdependencies between choices, more quickly. The study categorizes users based on their level of overconfidence, arguing that overconfident users are less likely to trust a recommender system and therefore benefit less from its assistance.

2. References

Caro, F., de Tejada Cuenca, A.S., 2023. Believing in analytics: Managers' adherence to price recommendations from a DSS. Manufacturing & Service Operations Management 25, 524–542. doi:10.1287/msom.2022.1166.

Chen, J., Dong, H., Wang, X., Feng, F., Wang, M., He, X., 2023. Bias and debias in recommender system: A survey and future directions. ACM Transactions on Information Systems 41, 1–39. doi:10.1145/3564284.

Forouzandeh, S., Rostami, M., Berahmand, K., 2021. Presentation a trust walker for rating prediction in recommender system with biased random walk: Effects of h-index centrality, similarity in items and friends Engineering Applications of Artificial Intelligence 104, 104325. doi:10.1016/j.engappai. 2021.104325.

Gharahighehi, A., Vens, C., Pliakos, K., 2021. Fair multi-stakeholder news recommender system with hypergraph ranking. Information Processing & Management 58, 102663. doi:10.1016/j.ipm.2021.102663.

Käki, A., Kemppainen, K., Liesiö, J., 2019. What to do when decision-makers deviate from model recommendations? empirical evidence from hydropower industry. European Journal of Operational Research 278, 869–882. doi:10. 1016/j.ejor.2019.04.021.

Khosrowabadi, N., Hoberg, K., Imdahl, C., 2022. Evaluating human behaviour in response to AI recommendations for judgemental forecasting. European Journal of Operational Research 303, 1151–1167. doi:10.1016/j.ejor.2022.03.017.

Nguyen, D.T., Adulyasak, Y., Landry, S., 2021. Research manuscript: The bullwhip effect in rule-based supply chain planning systems– a case-based simulation at a hard goods retailer. Omega 98, 102121. doi:10.1016/j.omega.2019.102121.

Ricci, F., Rokach, L., Shapira, B., 2021. Recommender systems: Techniques, applications, and challenges, in: Recommender Systems Handbook. Springer US, pp. 1–35. doi:10.1007/978-1-0716-2197-4_1.

Sroginis, A., Fildes, R., Kourentzes, N., 2023. Use of contextual and modelbased information in adjusting promotional forecasts. European Journal of Operational Research 307, 1177–1191. doi:10.1016/j.ejor.2022.10.005.

Zhang, S., Yao, L., Sun, A., Tay, Y., 2019. Deep learning based recommender system. ACM Computing Surveys 52, 1–38. doi:10.1145/3285029.

Virtualising Soft OR interventions

Katharina Burger [1*], Leroy White [2]

[1] University College London, London, UK
[2] University of Exeter, Exeter, UK
[*] katharina.burger@ucl.ac.uk

Abstract

This paper addresses the need to appreciate opportunities that arise from the digital transformation of spaces for collaborative, creative group interactions. Specifically, it explores the experience of using virtual reality (VR) technology and immersive virtual environments, addressing the question: How do immersive VR spaces transform collaborative social interaction, which is central to group decision support activities? Empirically, we draw on in-headset recordings from a social virtual reality workshop to study participant interaction in a social immersive virtual environment in the metaverse. Theoretically, we investigate how affordances for action possibilities are realised. We contribute to ongoing debates about the organisation of online and offline spaces and conclude by reflecting on the implications for virtualising group decision processes more broadly. We suggest several areas that may be of interest to identify ways to reimagine the role of space for OR as a transformative tool.

Keywords: Group interaction, Virtual reality, Empirical study, Qualitative methods; Soft OR

1. Introduction

Group decision and negotiation (GDN) scholars have focused attention on the benefits yielded from technology for distributed work for quite some time (Qureshi & Vogel, 2001). Moreover, online and virtual communication tools have experienced a steady increase in interest from facilitators (Kersten & Lai, 2007). Relatedly, the role of online systems in facilitating group activities has attracted considerable attention (Brown et al., 2022; Wagner, 1995; Zimmermann et al., 2021). However, the suitability of novel social Immersive Virtual Environments (IVEs) for collaborative participant interactions to expand action possibilities is not well understood.

In this study, we add to considerations of how IVEs may be leveraged for virtualising group interaction. Thus, we wish to contribute to recent studies in GDN that have aspired to understand how people act in two-dimensional (2D) online spaces (Zimmermann et al., 2021). However, we go beyond the predominant focus on flat 2D digital systems as a form of online delivery and consider how immersion in three-dimensional (3D) space matters for group interaction. We are mindful of the importance of social interaction for Soft OR and group decision support (White et al., 2016). Therefore, social IVEs, i.e., 3D virtual spaces where multiple participants can engage with one another using VR head-mounted displays (HMDs), are of interest. They are claimed to change how people communicate, connect, and socialise with each other via full-body avatars, virtual objects in the scene and multimodal communication, rather than just by looking at a computer screen. However, IVEs are ambiguous technologies (Berente et al., 2011), posing challenges for those considering adopting such environments. For instance, how participants interact in IVEs together needs to be better understood. Therefore, our study focuses on the following research question: How do immersive VR spaces transform collaborative social interaction, which is central to group decision-support activities?

2. Background

2.1. The influence of space on group dynamics and decision-making

The importance of the physical setting in group work within operations research (OR) and systems methodologies has been a subject of discussion among scholars. White (2006) highlighted its critical role, while Cairns (2002) noted how the spatial arrangement impacts the design of processes, facilitating either standardisation or adaptability. Huxham (1990) and Bryant (1989) emphasised that the physical environment should be integrated with process-related issues and should support the intended activities. Innovations like the decision conferencing 'pod' by Phillips (described by Bryant, 1989) and an ideal decision-making environment designed by Hickling (1990), featuring a clear, well-lit area with flexible furnishings and ample technical support, demonstrate the potential for improvement.

Eden (1990) further argued that managing the physical space is crucial for handling complexity and enhancing consensus building, pointing out that a more dynamic and flexible interaction within a well-designed space could foster creative thinking and effective decision-making. Eden (1990) also touched upon the symbolic significance of the physical setting, particularly in small group work related to decision-making or strategic planning within organisations and cautioned against altering the architecture of traditional decision-making spaces like boardrooms to preserve their symbolic value.

This body of work collectively underscores the pivotal role of the physical environment in enhancing the efficacy of group processes and decision-making in OR and systems methodologies. Despite these insights, there is still a lack of diversity in the design of spaces for decision-making processes. Indeed, there are ongoing challenges involved in understanding how space matters for group interaction – often stemming from a focus on methods rather than situated social processes- and shows the need for further theorisation in GDN, particularly for studies seeking to understand the strengths and limitations of virtual spaces and virtualised participant experiences.

2.2. Space and virtual interaction

Online delivery methods for collaboration and learning have significantly evolved, leveraging technologies like Voice over Internet Protocol (VoIP) systems, including Zoom and MS Teams, alongside dedicated software packages such as Mural or Strategyfinder to facilitate modelling activities in a digital environment (Ackermann & Eden, 2020). However, transitioning to 2D online environments has also heightened participants' awareness of their physical separation and the limitations of interacting through screens and input devices. Luff et al. (2003) note that online non-immersive settings offer limited visibility of participants' physical gestures, thus diminishing the sense of copresence crucial for effective collaboration and learning. Challenges persist, such as inequalities in participant engagement due to the size and placement of video displays, volume levels, and camera angles (Saatçi et al., 2020). For instance, algorithms governing platforms like Zoom can disrupt traditional seating arrangements, affecting the social dynamics of online meetings without this issue being thoroughly addressed in the literature.

Given the critical role of social interaction in Soft OR interventions, the potential of immersive virtual environments warrants closer examination. Markowitz et al. (2018) argue that immersive VR experiences, which enable users to feel genuinely present in a simulated environment, differ markedly from conventional online meetings by providing a more engaging and interactive space. This immersive quality encourages participants to overlook the mediated nature of the experience, enhancing collaboration and learning outcomes. Dalgarno & Lee (2010) further support this view, suggesting that VR environments can facilitate richer collaborative learning experiences. Williamson et al. (2021) emphasise the importance of designing VR spaces that promote group formation, shared attention, and respect for personal space, underscoring the transformative potential of VR technologies in fostering effective and collaborative group interactions.

2.3. Affordances and social interaction in IVEs

Affordance theory provides a foundational perspective for understanding the interaction between environmental properties and human actions. It posits that the environment offers various possibilities for actions that are not inherently deterministic (Norman, 2002). This theory suggests that while certain actions are made possible by these affordances, their utilisation is not guaranteed and often requires learning and cultural exposure, highlighting a dynamic perception-action cycle (Costall, 1995; Linderoth, 2012).

Affordance theory has been applied to the study of IVEs, where it aids in exploring how users perceive and interact with virtual interfaces, objects, and spaces (Dincelli & Yayla, 2022; Lanamäki et al., 2016; Osiurak et al., 2017). For instance, affordance theory can be used to examine how VR environments enable various forms of interactions, collaborations, and decision-making processes (Bowman et al., 2012; Cleary et al., 2012; McVeigh-Schultz & Isbister, 2021).

The interaction with new technologies, including IVEs, has garnered attention across disciplines such as human-computer interaction and organisational studies, particularly focusing on the social presence experienced within IVEs. This sense of being with others in a shared space is significantly enhanced by virtually embodied interactions through avatars, leading to richer communication that includes spatial and non-verbal cues (Dalgarno & Lee, 2010; Mennecke et al., 2011). IVEs may be designed to mirror offline social expectations, providing a richly meaningful space for collaboration without predefined scripts, relying instead on cues and collaboration facilitators designed to promote cooperative behaviour (McVeigh-Schultz & Isbister, 2022).

Despite IVEs' rich potential for enhancing collaboration through affordances, there has been a relative lack of focus on the relational aspect of affordances—how they are perceived and emerge within these environments (Osmundsen et al., 2022). Addressing this gap requires adopting an emic perspective on affordance actualisation in IVEs, emphasising the subjective experience of interacting within IVEs and the role of perception in realising the potential actions and behaviours that these environments afford. It is this perspective that we focus on in our study.

3. Methodology

Our research is exploratory. We adopt a qualitative research design grounded in an interpretivist paradigm. Our analysis is based on a synchronous immersive social VR workshop with head-mounted displays (HMDs), and we rely on naturalistic observation as the research approach. This approach is particularly suited to exploring complex, context-dependent social dynamics that might not be readily observable or replicable in controlled experimental settings.

Table 1: Comparative Overview of Naturalistic Observation and Behavioural Experiment Methodologies

	Naturalistic Observation (with Videography)	**Behavioural Experiment**
Objective	To understand behaviour within its natural context.	To identify causal relationships between variables.
Methodological Orientation	Emic perspective: focuses on subjective experiences and interpretations within cultural and social contexts.	Etic perspective: emphasises objective measurement and analysis of behaviours.
Data Collection	Observations and recordings made in natural settings, without intervention. Videography captures the unfolding of social actions in real-time.	Controlled manipulation of variables in an experimental setting, often with predefined tasks for participants.
Analytical Framework	Interpretive and hermeneutical approaches aim to uncover meanings behind behaviours and interactions. Aims at depth of analysis.	Statistical analysis focused on hypothesis testing, measuring the effect of independent variables on dependent variables.
Focus	Richness and complexity of real-world interactions and social dynamics.	Isolation and manipulation of variables to establish cause-and-effect relationships.
Outcome	Generates in-depth insights and thick descriptions of social phenomena within their natural contexts.	Seeks to determine causality, predictability, and generalizability of findings beyond the experimental setting.

The qualitative paradigm of our study, emphasising naturalistic observation, diverges significantly from a comparative study design that might seek to directly contrast 2D and 3D environments. Our approach is rooted in interpretivism, seeking to understand the meaning-making processes of participants within VR environments from their perspectives.

3.1. The process of data collection and participant characterisation

Our participants were recruited via University-internal mailing lists and were all postgraduate taught or research students from different disciplines. We specifically targeted the recruitment for the workshop at novices who did not own HMDs. Before the workshop, participants were sent participant information sheets and ethical consent forms per the conditions of the favourable ethical review. Participants who consented were then sent an introductory tutorial about spatial.io, which the research assistant had prepared. We held the workshop synchronously in person to troubleshoot any possible technical issues and monitor participants' well-being during the VR HMD usage, which can cause motion sickness.

We started the workshop with a round of introductions. Next, participants were given a short questionnaire to establish their familiarity with VR. This confirmed that they were all novices. The workshop deployed six VR headsets. It had four novice participants using the HMDs, the participant facilitator/workshop organiser, and a cofacilitator/research assistant, also using HMDs. The workshop took place in May 2022 at a UK University, was conducted in English, although not all participants were English native speakers, and lasted half a day. A few activities in VR activities were completed as part of the workshop, with breaks in between, before the social VR session that forms the basis of this paper.

We used Oculus Quest headsets with handheld controllers and the spatial.io app platform for the workshop for the meeting space. In spatial.io, through virtual avatars, facilitators and participants can communicate using both verbal and nonverbal cues. At the start of the social VR activity, participants were helped to set up the HMDs, were asked to complete a basic tutorial individually, and then asked to join a group meeting space in spatial.io. Avatars were created through spatial.io's image translation algorithm, which enabled users to automatically generate avatars based on stock photos or personal photos, resulting in avatars that looked like themselves.

Participants were asked to press the record button within their virtual menus in their headsets so that their individual views would be recorded. They were not given specific instructions about what to do once they arrived in the virtual room. They were told to explore the space for roughly 20 minutes. We focus our analysis on what happened inside the VR space during this exploration time (Figure 1).

Figure 1: Screenshot from virtual board room in spatial.io

3D Videography as method

Through videography (Knoblauch et al., 2006), our research aims to understand the complexities and nuances of social interactions within immersive VR spaces as they occur naturally. In line with previous studies (Williamson et al., 2021), we captured multimodal interactions (speech, hand gestures, avatar body language and scene manipulations) through in-headset recordings, allowing us to review 3D content in an immersive, expressive, and asynchronous way (Chow et al., 2019). A multimodal recording encapsulates time-synced streams of interactions that can be captured and replayed, akin to streams of pixels and audio in a video recording, but in 3D. The data type we work with is thus 3D video data.

Videography (Knoblauch et al., 2006) typically focuses on actions, interactions, and practices within a setting or across a setting. This method is aimed at uncovering the spontaneous, unscripted, and context-rich interactions that define VR experiences, providing insights that could be overlooked in more structured comparative research. These approaches involve researchers either entering VR environments themselves or analysing recorded VR interactions to study social behaviour and user experience from an emic perspective, demonstrating the utility of immersion and intuition in virtual contexts (e.g., Boellstorff, 2008; Hine, 2000).

The process of analysing data

Qualitative analysis champions the development of theoretical insights inductively derived from empirical observations, as opposed to the imposition of pre-existing theoretical frameworks. This approach ensures that our interpretations remain firmly anchored in the empirical data, reflecting the intricate nature of social interactions within immersive VR settings.

Our analytical approach of immersion, intuition, and episode identification is based on a methodological and theoretical foundation for understanding complex phenomena, drawing from established methodologies in grounded theory (Glaser & Strauss, 2009). This approach emphasises generating insights directly from the data through a deep and iterative engagement that fosters an intuitive grasp of significant episodes. Such engagement ensures that the analysis remains firmly anchored in the empirical material, enhancing the validity of the findings (Saldana, 2014).

We further clarify the differences between the analytical approaches in Table 2.

Table 2: Contrasting Approaches to Data Analysis: Qualitative Immersion vs. Quantitative Positivism

	Qualitative Immersion, Intuition, Episode Analysis	**Quantitative Positivism**
Data Analysis	Qualitative, involving detailed interpretation of video data and contextual insights.	Quantitative, involving statistical tests to evaluate hypotheses.
Nature of Data Engagement	Deep engagement with descriptive data.	Detached, focusing on numerical data and statistical analysis.
Role of Researcher's Insight	Crucial, with intuition guiding interpretation. Reflexivity is emphasised.	Minimized in favour of objectivity. Personal insights are less emphasised.
Focus of Analysis	Identifying significant episodes or narratives to understand complexities.	Analysing variables to test hypotheses and predict outcomes.
Theories	Inductive approach: theories emerge from the data during analysis.	Deductive approach: starts with hypotheses based on existing theories to test.
Outcome of Analysis	Rich, contextualised understandings and narratives.	Statistical evidence supporting/refuting hypotheses; contributes to predictive models.
Examples of Research Questions	"How do individuals cope with interaction demands in immersive virtual environments?"	"Does the use of virtual environments increase the quality of decisions in small groups?"

The qualitative research tradition prioritises the generation of detailed, contextually nuanced understandings of specific occurrences (Stake, 1995) through a thorough exploration and interpretation of the intricate dynamics present within settings like immersive VR environments (Thanyadit et al., 2022; Williamson et al., 2021). Our analytical approach is thus designed to uncover and analyse the rich, multifaceted nature of social interactions. For the analysis, we adopted an emic perspective (Sacks, 1989) (see also Table 1) where the analyst investigates how participants form actions, develop mutual understanding, progress, and how their actions cohere (Linell, 2009).

In practical terms, this meant that, after the workshop, we reviewed the data, which consisted of an audio recording of the full workshop, two screencast recordings, and several in-headset recordings that captured parts of the social VR meeting.

3.2. Method for identifying affordances from naturalistic video recordings

Video recordings can provide insight into situated action and interaction, and this has been the case for some time in other social sciences, such as anthropology, communications, education, and sociology (Erickson, 1988; Goldman et al., 2014). More and more, scholars are highlighting the usefulness of video for studying organisational phenomena (Heath et al., 2000; Knoblauch et al., 2006). The analysis follows an inductive process.

Step 1: Immersion and familiarity

Immersion in data is considered instrumental for developing a nuanced comprehension of the subtleties inherent in social interactions within VR contexts, facilitating a process whereby researchers become profoundly attuned to the data (Emerson, 2007). The strategy of immersing oneself, relying on intuition, and identifying episodes during qualitative research is based on a solid theoretical and methodological foundation for understanding intricate phenomena, drawing from grounded theory (Glaser & Strauss, 2009). For instance, in organisational studies, immersion and intuition have been pivotal in uncovering episodes that reveal the underlying power structures and communication patterns that shape organisational culture and change (Vaara & Whittington, 2012).

We immersed ourselves in the data, reviewing the recordings, following where participants' attention was directed, how they talked, how they used nonverbal behaviours, and how they interacted with digital objects (Jordan & Henderson, 1995). This involves repeatedly listening to audio recordings, watching video footage, and reading through transcripts or field notes. Such deep immersion is essential for developing a holistic understanding of the data's context, content, and subtleties. Through this immersion, the researcher becomes increasingly familiar with and sensitised to the data. This familiarity allows them to recognise not just what is being said or done but how it is being expressed, including the emotional, cultural, or contextual nuances that accompany these expressions.

Step 2: Sensitization to subtleties and episode identification

Informed by the researcher's depth of engagement with the current data, intuition guides the recognition of significant segments or episodes. Episodes offer concrete examples or instances of how individuals interact, communicate, and engage with their environment, facilitating deeper insights into social processes and dynamics and thereby revealing complex configurations of affordance perception and actualisation (Mondada, 2011).

Episodes become recognisable to researchers immersed in data as they notice standout moments which signify episodes. As familiarity with the data increases, researchers become sensitised to its subtleties and nuances. This includes variations in tone, patterns of speech or behaviour, recurring themes, and anomalies (Mondada, 2009). It is during this phase that researchers start to notice moments that "stand out" from the rest of the data, whether due to their intensity or uniqueness. These episodes might manifest as moments of change, conflict, clarity, or any event that intuitively feels pivotal or illuminating (Mondada, 2019).

Step 3: Interpretation, reflection and theoretical integration with affordance theory

Once episodes emerged from the data, we continued to engage with it and considered our initial intuitive recognitions further. We followed prior research in which researchers used episodic analysis to examine how participants communicate and collaborate in VR settings, shedding light on the nuances of virtual interaction and the potential of VR for enhancing group processes (McVeigh-Schultz & Isbister, 2022).

As episodes contain specific instances or sequences of behaviour, situated within specific contexts, and representative of the temporal, contextual, and multifaceted nature of human behaviour, they provide the 'raw material' to identify how affordances are perceived and actualised (Goodwin, 2006; Mondada, 2019). We considered the sequential unfolding of action, e.g. bodily positions, movements, and conversations to study what the avatars were oriented towards and by what they were guided in their orientation in space (Mondada, 2003).

4. Findings

Our findings show how managing to collaborate in the IVE appears as a situationally emergent, contingent, and processual accomplishment. We present each episode in turn (Table 3).

Table 3: Overview of the identified episodes

	Shared visualisation	**Rapport-building**	**Multi-perspectivity**
Supported activity	Collaborative ideation	Creative self-expression	Collaborative discovery
Social presence	Moving closer together, mutual orientation to each other, observing each other's actions, and visual attention sharing.	Moving closer together; mutual orientation; observing each other (e.g., greetings); visual attention sharing	Orientation to each other; avatars commenting on each other's position
Objects focus attention	Sticky notes	Selfie-Stick	Boardroom table with chairs
Activities	Brainstorming activities can enable collaborative ideation as participants are encouraged to make their ideas visible to others using sticky notes on a wall or whiteboard that then constitute a mutual focus point around which discussion may spark new thoughts (Nielsen, 2012)	Taking a selfie is a visual means of creative self-expression (self-portraits) (Fernández-Castañón et al., 2022). Group selfies are engaging and enjoyable interactions that could strengthen collaboration (Kienzle, 2017) and enhance social interaction (Kumar et al., 2023).	Experiencing the partiality of one's field of vision, in most meetings, there is a strong 'sitting norm' (Mansfield et al., 2018).
Affordances	Shared visualisation allows participants to access a common visual space, enhancing mutual understanding, coordination, and the exchange of ideas during group activities. Interaction in the IVE	Interaction in the IVE enables participants to facilitate the establishment and maintenance of positive interpersonal relationships and mutual understanding among participants. In this way,	Interaction in the IVE enables participants to establish and maintain a sense of co-presence and facilitates mutual awareness, attention, and alignment of participants towards

	enables participants to collectively interpret and interact with visual information, fostering collaboration, communication, and decision-making processes.	it enables individuals to develop a sense of connection, expressing empathy, fostering a conducive atmosphere for effective communication and collaboration.	one another during collaborative activities.

Our case study highlights specific tools within these environments—sticky notes, a boardroom table, and a selfie stick—and their roles in enhancing the interaction. Their affordances support collaborative ideation, creative self-expression, and collaborative discovery, emerging interactively in the course of action. Collectively, they create conditions conducive to collaborative and creative exploration, which is important for collaborative interaction. this way, our study provides hints towards the opportunities that virtual spaces present for Soft OR interventions in both purely virtual and hybrid arrangements.

5. Discussion

Almost two decades ago, White (2006) concluded that space is 'rarely spoken about, or it is treated as if it were trivial, commonsense, or obvious' (White, 2006). Still today, in much contemporary work on group decision support, and in particular online facilitation, the role of spatial design is rarely considered beyond the spatial organisation of abstract elements in a model, as if participants were brains on sticks concerned solely with cognitive operations.

Our research investigates the nuances of participant interaction within IVEs, focusing particularly on how these spaces create the conditions for group decision-making processes Our focus extended beyond the examination of individual participant experiences in IVEs to explore the social interactions between multiple participants. This shift allowed us to consider how collective achievements emerge through relationships, advancing previous research that highlighted how online environments (Benbunan-Fich, 2002; Nunamaker Jr et al., 1991) and virtual worlds (Goh & Paradice, 2008; Schouten et al., 2010) transform interpersonal communication and collaboration dynamics.

Research on social VR, particularly regarding collaborative and creative decision support in workshops, is still nascent. There remains much to explore concerning coordinating movements, establishing behavioural norms, and designing for pro-social interactions within these immersive virtual environments. Informed by our exploratory study, we conclude by suggesting a number of research areas that may help advance our understanding of how VR technology offers an innovative avenue to transcend traditional spatial limitations and redefine the way collaboration is enabled in group decision support interventions.

VR time as a 'symbolic cut' from the everyday

Our exploratory study has shown that interacting in immersive virtual environments allows participants to experience a temporary separation from their everyday surroundings. For some, this may resonate with the idea that participants take time out of their everyday lives to attend and interact in workshops that are often used in Soft OR and GDN. Future research may consider whether interacting in VR allows for an even stronger separation from the everyday than traditional workshop settings—and whether this is positive or negative for looping back into the everyday.

VR customizability of virtual spaces

The flexibility of VR technology enables the customisation of virtual workspaces to fit the specific needs of group decision support sessions. Given this newfound abundance of environments for group decision interventions, how should facilitators leverage this flexibility? Should they aim to replicate real-world settings, e.g. adjusting the size of the space, the arrangement of virtual furniture, and even the ambient conditions (and thereby leverage the predictability of behavioural interactions based on norms), or should they seize the weirdness of VR (e.g. the ability to fly) (see also, McVeigh-Schultz & Isbister, 2021)?

VR and inclusivity

VR removes the physical and logistical constraints associated with GDN or Soft OR workshops, such as the need for physical space and the limitations it imposes on group interaction. Participants from different geographical locations can join a virtual space that mimics the advantages of face-to-face interactions without the associated costs and time of travel. However, ensuring participants have access to the necessary VR equipment is essential for the inclusive deployment of VR in group decision support.

6. Conclusion

Debates about the organisation of online and offline collaborative spaces are ongoing. There is still considerable uncertainty about the suitability of immersive VR for group decision support interventions and to provide spaces for group decision and negotiation. Our study has drawn attention to the exploratory movements and collaborative behaviours involved in developing familiarity with an IVE. Our perspective illuminates the relational, embodied ways in which participants refine their sensitivity to affordances in the IVE as they pursue collaborative activities. By leveraging VR, facilitators can rethink the role of space for GDN, exploring how virtual spaces influence group processes and outcomes. This may include experimenting with how different virtual environments affect group dynamics. As the technology becomes more accessible, its application in group decision support could redefine the traditional boundaries of workspace and collaboration, leading to more innovative and effective outcomes.

References

Ackermann, F., & Eden, C. (2020). Strategic Options Development and Analysis. In M. Reynolds & S. Holwell (Retired) (Eds.), *Systems Approaches to Making Change: A Practical Guide* (pp. 139–199). Springer London. https://doi.org/10.1007/978-1-4471-7472-1_4

Benbunan-Fich, R. (2002). Improving education and training with IT. *Communications of the ACM*, *45*(6), 94–99.

Berente, N., Hansen, S., Pike, J. C., & Bateman, P. J. (2011). Arguing the Value of Virtual Worlds: Patterns of Discursive Sensemaking of an Innovative Technology. *MIS Quarterly*, *35*(3), 685–709. https://doi.org/10.2307/23042804

Boellstorff, T. (2008). *Coming of age in second life: an anthropologist explores the virtually human*. Princeton University Press.

Bowman, D. A., McMahan, R. P., & Ragan, E. D. (2012). Questioning naturalism in 3D user interfaces. *Communications of the ACM*, *55*(9), 78–88.

Brown, K. K., Lemke, M. K., Fallah-Fini, S., Hall, A., & Obasanya, M. (2022). Planning, implementing, and evaluating an online group-model-building workshop during the COVID-19 pandemic: celebrating successes and learning from shortcomings. *System Dynamics Review*, *38*(1), 93–112. https://doi.org/10.1002/SDR.1704

Bryant, J. (1989). *Problem management: a guide for producers and players*. John Wiley & Sons Inc.

Cairns, G. (2002). Aesthetics, morality and power: Design as espoused freedom and implicit control. *Human Relations*, *55*(7), 799–820.

Chow, K., Coyiuto, C., Nguyen, C., & Yoon, D. (2019). Challenges and Design Considerations for Multimodal Asynchronous Collaboration in VR. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1–24. https://doi.org/10.1145/3359142

Cleary, A. M., Brown, A. S., Sawyer, B. D., Nomi, J. S., Ajoku, A. C., & Ryals, A. J. (2012). Familiarity from the configuration of objects in 3-dimensional space and its relation to déjà vu: A virtual reality investigation. *Consciousness and Cognition*, *21*(2), 969–975. https://doi.org/https://doi.org/10.1016/j.concog.2011.12.010

Costall, A. (1995). Socializing affordances. *Theory & Psychology*, *5*(4), 467–481.

Dalgarno, B., & Lee, M. J. W. (2010). What are the learning affordances of 3-D virtual environments? *British Journal of Educational Technology*, *41*(1), 10–32.

Dincelli, E., & Yayla, A. (2022). Immersive virtual reality in the age of the Metaverse: A hybrid-narrative review based on the technology affordance perspective. *The Journal of Strategic Information Systems*, *31*(2), 101717. https://doi.org/https://doi.org/10.1016/j.jsis.2022.101717

Eden, C. (1990). Managing the environment as a means to managing complexity. In C. Eden & J. Radford (Eds.), *Tackling Strategic Problems: The Role of Group Decision Support*. Sage Publications.

Emerson, R. (2007). Working with 'key incidents''.' In C. Seale, G. Gobo, J. F. Gubrium, & D. Silverman (Eds.), *Qualitative research practice* (pp. 427–442). SAGE. https://doi.org/http://dx.doi.org/10.4135/9781848608191.d35

Erickson, F. (1988). Ethnographic description. In U. Ammon (Ed.), *Sociolinguistics: An international handbook of the science of language and society* (pp. 1081–1095). Walter de Gruyter.

Fernández-Castañón, J., López-Manrique, I., & González, C. (2022). Developing creativity through portrait and selfie in elementary schools. *Journal of Higher Education Theory and Practice*, *22*(2), 95–106.

Glaser, B. G., & Strauss, A. L. (2009). *The discovery of grounded theory: Strategies for qualitative research*. Transaction Publishers.

Goh, S. H., & Paradice, D. B. (2008). Collaborative Worlds and Avatar-Based Communication: A Comparison of Virtual Worlds with Traditional Computer-Mediated Communications Media. *AMCIS 2008 Proceedings*, 23.

Goldman, R., Pea, R., Barron, B., & Derry, S. J. (2014). *Video research in the learning sciences*. Routledge.

Goodwin, C. (2006). Human sociality as mutual orientation in a rich interactive environment. Multimodal utterances and pointing in aphasia. In N. Enfield & S. C. Levinson (Eds.), *Roots of human sociality*. Berg.

Heath, C., Knoblauch, H., & Luff, P. (2000). Technology and social interaction: the emergence of "workplace studies." *Br. J. Sociol.*, *51*(2), 299–320. https://doi.org/10.1111/j.1468-4446.2000.00299.x

Hickling, A. (1990). Decision spaces: A scenario about designing appropriate rooms for group decision management. In C. Eden & J. Radford (Eds.), *Tackling Strategic Problems: The Role of Group Decision Support, pp*. The Role of Group Decision Support. Sage Publications, London.

Hine, C. M. (2000). *Virtual ethnography*. Sage. https://doi.org/10.4135/9780857020277

Huxham, C. (1990). On trivialities in process. In R. C. (Ed.), *Tackling Strategic Problems: The Role of Group Decision Support* (pp. 162–168). Tackling Strategic Problems: The Role of Group Decision Support. Sage.

Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *The Journal of the Learning Sciences*, *4*(1), 39–103.

Kersten, G. E., & Lai, H. (2007). Negotiation support and e-negotiation systems: An overview. *Group Decision and Negotiation*, *16*, 553–586.

Kienzle, J. (2017). The selfie (and ussie) as identity in the classroom. In S. Pixy Ferris & H. Wilder (Eds.), *Unplugging the Classroom* (pp. 93–102). Chandos Publishing. https://doi.org/https://doi.org/10.1016/B978-0-08-102035-7.00007-2

Knoblauch, H., Schnettler, B., Raab, J., & Soeffner, H.-G. (2006). Video analysis: methodology and methods. *Qualitative Audiovisual Data Analysis in Sociology. Frankfurt Am Main et al.:* Lang.

Kumar, J. A., Silva, P. A., Osman, S., & Bervell, B. (2023). Hey, let's take a selfie: insights of selfie defamiliarisation in the classroom. *Online Information Review*, 47(5), 820-837.

Lanamäki, A., Thapa, D., & Stendal, K. (2016). When is an affordance? Outlining four stances. *Beyond Interpretivism? New Encounters with Technology and Organization: IFIP WG 8.2 Working Conference on Information Systems and Organizations, IS&O 2016, Dublin, Ireland, December 9-10, 2016, Proceedings*, 125–139.

Linderoth, J. (2012). Why gamers don't learn more: An ecological approach to games as learning environments. *Journal of Gaming & Virtual Worlds*, *4*(1), 45–62.

Linell, P. (2009). *Rethinking Language, Mind, and World Dialogically: Interactional and Contextual Theories of Human Sense-Making*. Information Age Publishing.

Luff, P., Heath, C., Kuzuoka, H., Hindmarsh, J., Yamazaki, K., & Oyama, S. (2003). Fractured ecologies: creating environments for collaboration. *Human–Computer Interaction*, *18*(1–2), 51–84.

Mansfield, L., Hall, J., Smith, L., Rasch, M., Reeves, E., Dewitt, S., & Gardner, B. (2018). "Could you sit down please?" A qualitative analysis of employees' experiences of standing in normally-seated workplace meetings. *PLoS ONE*, *13*(6), 1–16. https://doi.org/10.1371/journal.pone.0198483

Markowitz, D. M., Laha, R., Perone, B. P., Pea, R. D., & Bailenson, J. N. (2018). Immersive Virtual Reality field trips facilitate learning about climate change. *Frontiers in Psychology*, *9*(NOV), 2364.

McVeigh-Schultz, J., & Isbister, K. (2021). The case for "weird social" in VR/XR: a vision of social superpowers beyond meatspace. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–10.

McVeigh-Schultz, J., & Isbister, K. (2022). A "beyond being there" for VR meetings: envisioning the future of remote work. *Human–Computer Interaction*, *37*(5), 433–453.

Mennecke, B. E., Triplett, J. L., Hassall, L. M., Conde, Z. J., & Heer, R. (2011). An examination of a theory of embodied social presence in virtual worlds. *Decision Sciences*, *42*(2), 413–450.

Mondada, L. (2003). Working with video: how surgeons produce video records of their actions. *Visual Studies*, *18*. https://doi.org/10.1080/1472586032000100083

Mondada, L. (2009). Emergent focused interactions in public places: A systematic analysis of the multimodal achievement of a common interactional space. *Journal of Pragmatics*, *41*(10), 1977–1997.

Mondada, L. (2011). Understanding as an embodied, situated and sequential achievement in interaction. *Journal of Pragmatics*, *43*(2), 542–552.

Mondada, L. (2019). Contemporary issues in conversation analysis: Embodiment and materiality, multimodality and multisensoriality in social interaction. *Journal of Pragmatics*, *145*, 47–62.

Nielsen, M. F. (2012). Using artifacts in brainstorming sessions to secure participation and decouple sequentiality. *Discourse Studies*, *14*(1), 87–109.

Norman, D. (2002). *The Design of Everyday Things*. Basic Books.

Nunamaker Jr, J. F., Dennis, A. R., Valacich, J. S., & Vogel, D. R. (1991). Information technology for negotiating groups: Generating options for mutual gain. *Management Science*, *37*(10), 1325–1346.

Osiurak, F., Rossetti, Y., & Badets, A. (2017). What is an affordance? 40 years later. *Neuroscience & Biobehavioral Reviews*, *77*, 403–417.

Osmundsen, K. S., Meske, C., & Thapa, D. (2022). Familiarity with digital twin totality: Exploring the relation and perception of affordances through a Heideggerian perspective. *Information Systems Journal*, *1*, 1–28. https://doi.org/10.1111/ISJ.12383

Qureshi, S., & Vogel, D. (2001). Adaptiveness in virtual teams: Organisational challenges and research directions. *Group Decision and Negotiation*, *10*, 27–46.

Saatçi, B., Akyüz, K., Rintel, S., & Klokmose, C. N. (2020). (Re) Configuring hybrid meetings: Moving from user-centered design to meeting-centered design. *Computer Supported Cooperative Work (CSCW)*, *29*, 769–794.

Sacks, H. (1989). *Harvey Sacks Lectures 1964–1965* (G. Jefferson, Ed.). Springer.

Saldana, J. (2014). *Thinking qualitatively: Methods of mind*. SAGE publications.

Schouten, A. P., van den Hooff, B., & Feldberg, F. (2010). Real decisions in virtual worlds: Team collaboration and decision making in 3D virtual worlds. *ICIS 2010 Proceedings*.

Stake, R. (1995). *The Art of Case Study Research*. SAGE Publications.

Thanyadit, S., Punpongsanon, P., Piumsomboon, T., & Pong, T. C. (2022). XR-LIVE: Enhancing Asynchronous Shared-Space Demonstrations with Spatial-temporal Assistive Toolsets for Effective Learning in Immersive Virtual Laboratories. *Proceedings of the ACM on Human-Computer Interaction*, *6*(CSCW1). https://doi.org/10.1145/3512983

Vaara, E., & Whittington, R. (2012). Strategy-as-Practice: Taking Social Practices Seriously. *Academy of Management Annals*, *6*(1), 285–336. https://doi.org/10.1080/19416520.2012.672039

Wagner, C. (1995). Facilitating space-time differences, group heterogeneity and multi-sensory task work through a multimedia supported group decision system. *Decision Support Systems*, *15*(3), 197–210. https://doi.org/10.1016/0167-9236(94)00061-V

White, L. (2006). Aesthetics in OR/systems practice: towards a concept of critical imagination as a challenge to systems thinking. *Systems Research and Behavioral Science: The Official Journal of the International Federation for Systems Research*, *23*(6), 779–791.

White, L., Burger, K., & Yearworth, M. (2016). Understanding behaviour in problem structuring methods interventions with activity theory. *European Journal of Operational Research*, *249*(3), 983–1004. https://doi.org/https://doi.org/10.1016/j.ejor.2015.07.044

Williamson, J. R., Li, J., & Vinayagamoorthy, V. (2021). Proxemics and social interactions in an instrumented virtual reality workshop. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3411764.3445729

Zimmermann, N., Pluchinotta, I., Salvia, G., Touchie, M., Stopps, H., Hamilton, I., Kesik, T., Dianati, K., & Chen, T. (2021). Moving online: reflections from conducting system dynamics workshops in virtual settings. *System Dynamics Review*, *37*(1), 59–71. https://doi.org/10.1002/SDR.1667

24th International Conference on Group Decision and Negotiation &
10th International Conference on Decision Support System Technology
Human-Centric Decision and Negotiation Support for Societal Transitions

Biform game analysis of incentive structures for circular food economy

Sinndy Dayana Rico Lugo[1*], Nariaki Nishino[1]

[1] Department of Technology Management for Innovation, Graduate School of Engineering, The University of Tokyo, Tokyo, Japan
[*] sinndydayana.rico@css.t.u-tokyo.ac.jp

Abstract

The agrifood supply chain, which encompasses food production and distribution, faces unique challenges in embracing circular economy practices, unlike non-perishable goods. Obstacles such as high food loss and waste, conflicting stakeholder interests, and limited collaborative incentives hinder the adoption of the circular food economy (CFE). Few studies have considered the related multi-stakeholder complex decision-making, resulting in a lack of literature on strategies that enable them to take part in CFE projects. To explore these challenges, this study used game theory, through the design of a biform model that combines noncooperative and cooperative behaviors to promote co-creation among diverse food system actors to reduce food loss and waste and boost profits. The model illustrates how stakeholders transact within and between chain echelons and how they can be incentivized to participate in CFE initiatives. Through various model simulations, three incentive structures were explored in several cost scenarios. Results offer insights to guide future CFE efforts and evidence-based decision-making, thus fostering collaboration and sustainable practices across the chain.

Keywords: Biform games; circular food economy; decision-making; agri-food supply chain; incentives.

1. Introduction

Supply chains have evolved into intricate networks that significantly influence business strategies, market competitiveness, and overall economic resilience, going beyond the movement of goods and impacting all aspects from production efficiency and cost management to environmental sustainability (Kalmykova et al., 2018). Particularly, the agri-food supply chain (AFSC), which encompasses food production and distribution, exhibits inefficiency in utilizing natural resources, leading to significant food loss and waste (FLW) with subsequent socio-environmental impacts due to distinctive barriers and a lack of decision-making aid.

The circular economy represents a step forward breaking the linear production model (Bauwens et al., 2020; Ellen MacArthur Foundation, 2015). However, various studies (Chiaraluce et al., 2021; Mesa Reyes et al., 2020) highlight that existing circulation models and policies find greater application in the technology sector but not in sectors linked to perishable products and food systems. Unique factors such as perishability, strict food quality standards, and the intricacies of the associated decision-making processes are deemed highly complex, posing challenges that the system may struggle to address.

The Circular Food Economy (CFE) aims to address these problems by promoting economically viable and ecologically friendly processes in the design, production, and distribution of food products (Rico Lugo, Kimita, et al., 2023b). It seeks to address the multi-objective nature of the AFSC and multiple stakeholder perspectives (Schröder et al., 2018) by analyzing realistic and robust alternatives to adopting sustainability.

Moreover, AFSC actors display cooperative and noncooperative behaviors, conflicting interests, and limited integration, resulting in interdependency. The lack of economic incentives hampers circularity adoption, which implies rationality (Miranda-Ackerman et al., 2017). This underscores the need for theoretical advancement and intelligent decision-making tools, implying the need for further research on CFE and decision-makers' interests.

Therefore, incentives to promote joining collaborative projects to reduce FLW disposal through circulation should be investigated. In this sense, the strategies that enable stakeholders to take part in such alternatives can be analyzed based on the game theory approach (Collins & Kumral, 2020; Rico Lugo et al., 2022). The biform approach has recently been used in supply chain studies because it combines elements from noncooperative and cooperative game theories (Feess & Thun, 2014).

The above factors lead to the following questions: How do collaborative initiatives among stakeholders at the AFSC impact the circulation of FLW, and what factors influence their cooperative and competitive behavior for the successful implementation of CFE?This study seeks to answer these questions by analyzing various CFE incentive structures using a biform game that represents the strategic decisions of stakeholders regarding circular alternatives in two stages: A noncooperative side related to the horizontal tendency to compete and a cooperative side associated with the distribution of CFE benefits through the formation of vertical alliances.

In this paper, the mathematical formulation of the biform game is summarized first, based on which three incentive structures are proposed. Then, the settings and results of a set of simulations are discussed, comparing the formation of coalitions in the Nash equilibrium, the potential of FLW reduction, the profits of players, and the total social welfare.

This paper presents the third phase of a research project that sought to explore CFE from the viewpoint of incentives for AFSC decision-makers. Results of previous phases can be found in Rico Lugo, Du, et al. (2023), Rico Lugo et al. (2022, 2024), Rico Lugo, Kimita, et al. (2023a, 2023b). The general project overview was presented in the Doctoral Consortium of GDN 2023.

2. Theoretical background

2.1. The agri-food supply chain and circulation challenges

The AFSC involves producing and distributing food products (including local crops) and processing and importing the products. This chain is generally consolidated into five echelons: farmer production, handling and storage, processing and packaging, distribution and marketing, and consumption (FAO, 2019; National Planning Department of Colombia, 2016; Teigiserova et al., 2020).

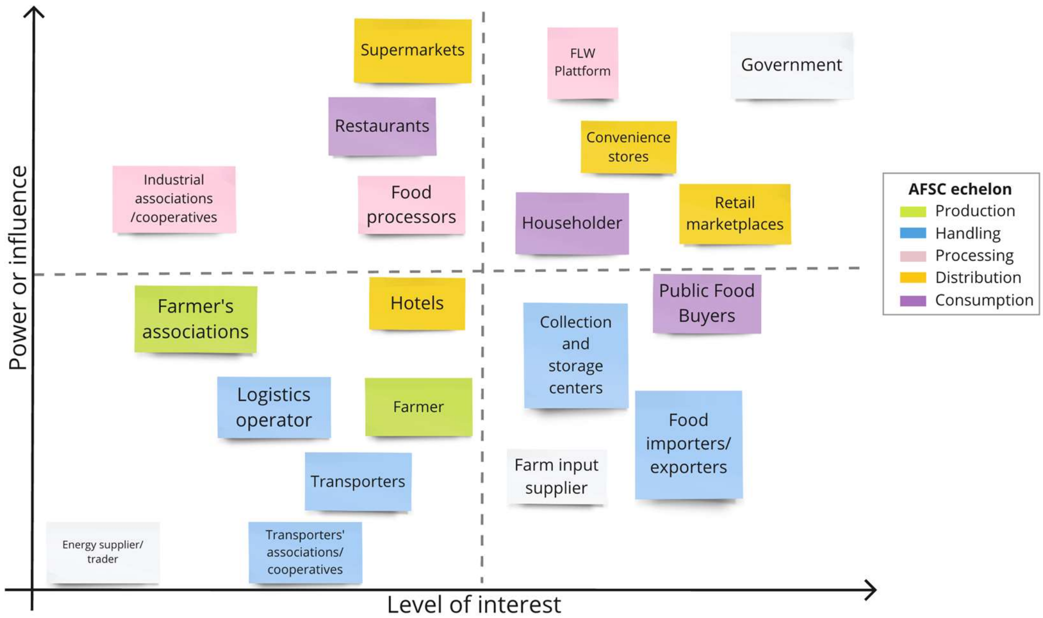


Figure 1: Power-interest map of AFSC stakeholders regarding CFE

Figure 1 is a power-interest map regarding CFE of the most common AFSC actors in those echelons, obtained through a workshop with multinational experts in food systems and circular economy in a previous research phase (Rico Lugo et al., 2024).

Moreover, the food ecosystem is traditionally a linear supply chain (Intergovernmental Panel on Climate Change, 2017; Teigiserova et al., 2020) that takes, makes, and disposes of resources (Rico Lugo, Du, et al., 2023). This results in the annual food loss and waste (FLW) representing 39% of the world's production, corresponding to 1.4 billion tons, €16,000 million, and 15.3 million equivalent CO_2 Tons (Ellen MacArthur Foundation, 2019; Intergovernmental Panel on Climate Change, 2017). Reducing FLW can provide a solution to food security challenges while preserving the environment, feeding more people, and making the food value chain more environmentally sustainable and resilient (Derqui et al., 2016). Therefore, considering the relevance of the effects of FLW from socioeconomic and environmental perspectives, based on quantitative studies (Betz et al., 2015), various ASFC stakeholders should be included to elucidate the real FLW situation and to suggest suitable mechanisms for incentivizing collaborative improvements (FAO, 2019).

2.2. Circular Food Economy

In recent years, the circular economy has attracted attention as a solution to chain sustainability issues (Bocken et al., 2016; Homrich et al., 2018; Wübbeke & Heroth, 2014). Particularly, the Circular Food Economy (CFE) concept has been adopted recently as a "co-creative food ecosystem that enhances food safety, food security, and biodiversity conservation, preventing food losses and waste, managing perishability, and using regenerative agriculture through reusing, recycling, recovering, and reprocessing edible food and inedible parts into circular loops and alliances" (Rico Lugo, Kimita, et al., 2023b). The CFE aims to co-build a sustainable socioeconomic environment based on food. It also adopts decision-making analysis and mechanism design to elucidate the rules, roles, and consequences that help realize a desirable food environment for all AFSC stakeholders.

2.3. Biform games and supply chains

Game theory analyzes the strategic interactions of the best outcomes from rational decisions, providing a systematic framework in which a participant's choice depends on others (Osborne & Rubinstein, 1994). The term "game" denotes situations where one participant's choices affect others' payoffs, emphasizing strategic thinking. Participants, called "players," are decision-makers with strategies that result in quantifiable outcomes called "payoffs" or "profits".

It distinguishes between noncooperative games, where players optimize individual gains without coalitions, and cooperative games, where decision-makers collaborate for joint benefits. Biform games (Brandenburger & Stuart, 2007) diverge by not directly treating strategic decisions as profits. Instead, they involve a second stage, a cooperative game where players create coalitions for enhanced benefits.

3. Biform game toward circular food economy

3.1. Agri-food supply chain structure

Based on the power-interest map of Figure 1 and the other characteristics of the CFE decision making process, it was decided to focus this biform model on the actors of the upper left side of the map. That was because increasing their interest in CFE projects through the incentive structure based on the biform game modeling will allow them to be relocated to the upper right side of the map, supporting the increment of actual implementation of CFE in the AFSC.

Therefore, this model considers the three AFSC stakeholders in the set of players $i \in N$, with $N = \{p, s, r\}$, where p represents "food processor," s represents "supermarket," and r represents "restaurant." The formulas of costs and incomes included are adapted to the reality of each stakeholder in the CFE decision-making context. The nomenclature of such formulas is shown in Figure 2, where product quantities are in blue (values close to each player), costs are in red, and incomes are in green. Model parameters are listed in Table 1.

In the project, p produces food products as supplies for s and r, which produce the final food products and services for external consumers. During the production processes of those stakeholders, some FLW are generated. Such FLW currently go out of the system as garbage, implying costs associated with garbage disposal, loss of investment, and environmental impact. There is also a new project of CFE focused on reprocessing such FLW, where p collects the FLW of s and r after acquiring new technology (machinery) used to transform it into new food-related products that it can sell.

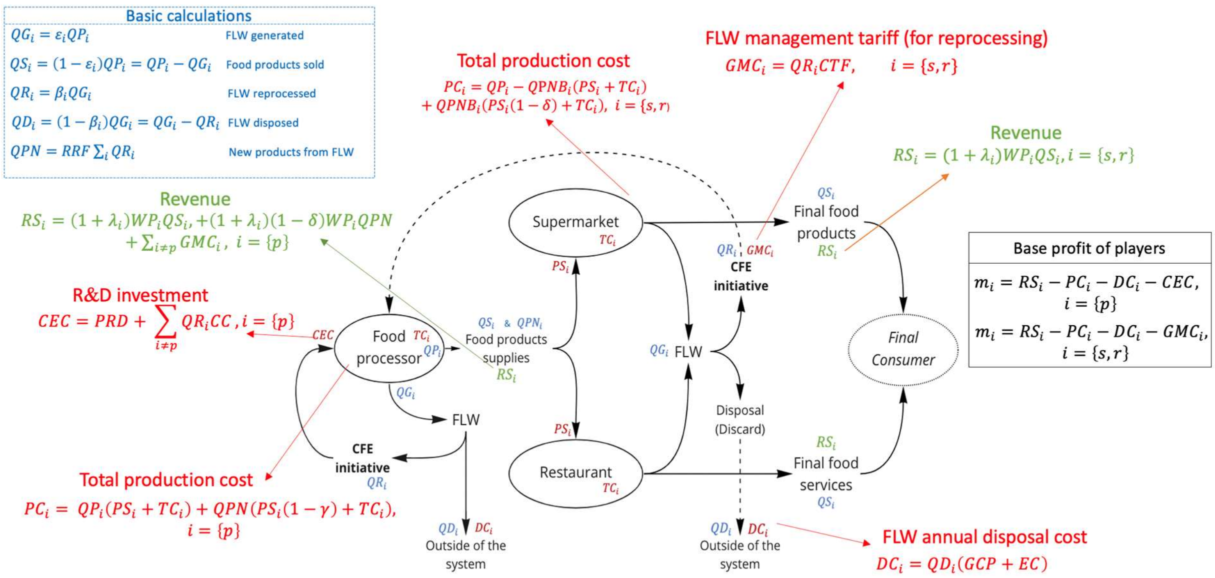


Figure 2: CFE decision-making situation with the main biform model formulas

Table 1. Model nomenclature

Nom.	Description	Nom.	Description
α_i	Confidence index	PS_i	Payment to suppliers to obtain one unit of food (\$/t)
ε_i	% of food products that become FLW	TC_i	Food preparation unit cost (\$/t)
β_i	% of FLW sent to the CFE initiative	DC_i	FLW disposal cost (\$/year)
δ	% of price reduction when purchasing FLW reprocessed	GCP	Organic garbage operator tariff (\$/t)
λ_i	% of potential sales increase	EC	Environmental and social cost (\$/t)
QP_i	Food products produced by player i (t/year)	GMC_i	FLW management tariff paid by the generator $\{s, r\}$ (\$/year)
QG_i	FLW generated by player i (t/year)	CP	Collection tariff of p for collecting FLW (\$/t)
QS_i	Food products sold by player i (t/year)	CEC	FLW cost paid by p for joining CFE (\$/year)
QR_i	FLW reprocessed by the CFE initiative (t/year)	PRD	R&D cost for reprocessing FLW (\$/year)
QD_i	FLW disposed of by player i (t/year)	CC	Cost for collecting FLW in s and r's locations (\$/t)
QPN_i	New products obtained from reprocessed food by player p (t/year)	RS_i	Player i's revenue (\$/year)
PC_i	Total production cost incurred by player i (\$/t)	WP_i	Wholesale price of food products/services (\$/t)

Because of the collaborative principle of CFE, it was assumed that the new project would require that at least the food processor and one of the other stakeholders agree to participate in it, then, FLW could decrease. The relationships among the effort, cost, and decrease in FLW disposal are assumed to be linear functions. More effort generates higher costs (Ma et al., 2019) (e.g., for technology acquisition, facility improvement, and training). As the effort becomes stronger, a greater reduction in FLW is achieved.

3.2. Biform model of the circulation of food loss and waste

A new biform model was designed for the particular context of the CFE in which AFSC players decide whether to join the CFE project for reducing FLW (increasing circulation) and how to distribute the surplus generated by forming coalitions. Figure 3 represents the extensive form of the game, based on the schematic of the biform games provided by (Brandenburger & Stuart, 2007). The ellipse connecting the nodes signifies the presence of information asymmetry and denotes that it is a simultaneous game (Fujiwara-Greve, 2015; McCain, 2009), implying that a player cannot differentiate between the nodes because of being unaware of the choices of other players. In the first stage, each player decides noncooperatively using the profit-maximizing premise of the Nash equilibrium concept. In the second stage, the overall reduction in FLW, consisting of the effort of all players, occurs, and the resulting surplus is divided based on the core concept.

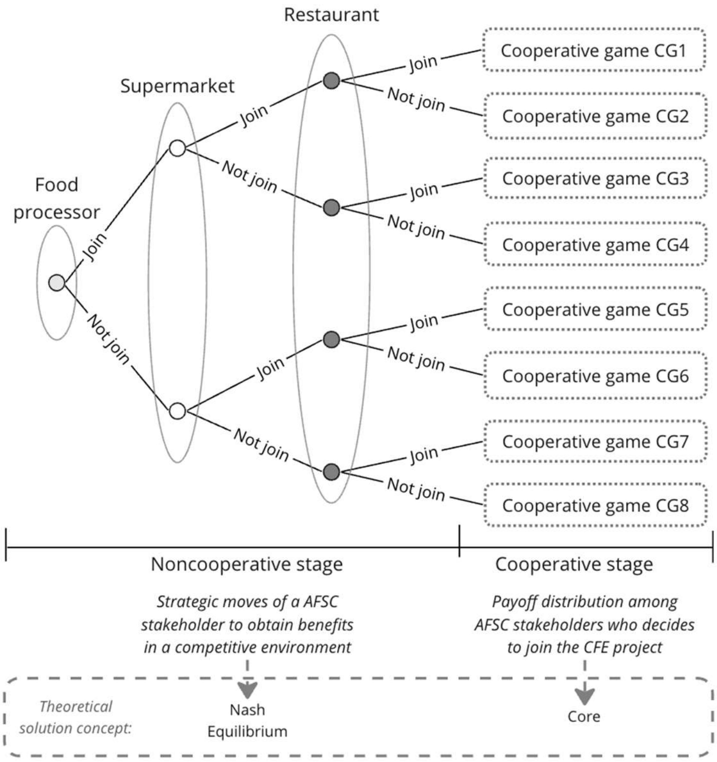


Figure 3: Extensive form of the CFE biform game

Based on the above, the proposed biform game is a collection

$$\left(SP_p, SP_s, SP_r; V; \alpha_p, \alpha_s, \alpha_r\right) \tag{1}$$

where

- The set of players is $N = \{p, s, r\}$;
- For each $i \in N$, SP_i is the finite set $\{J, NJ\}$, where *J* means to join the CFE project and *NJ* means not to join it;
- A coalition C is a subset of N;

- The resulting profile of strategies $(sp_p,\ sp_s, sp_r) \in SP_p \times SP_s \times SP_r$ defines a transferable utility cooperative game CG with characteristic function $V(sp_p,\ sp_s, sp_r) : P(N) \to \mathbb{R}$. In other words, for each $CG \subseteq N$, the value created by the subset C of players is $V(sp_p,\ sp_s, sp_r)$, given that the players chose the strategies $(sp_p,\ sp_s, sp_r)$;
- For each $i \in N$, $0 < \alpha_i < 1$, which represents the player i's confidence index;
- The final profit obtained for a player is denoted as $(x_p,\ x_s, x_r) \in \mathbb{R}^n$

Typically, biform games are analyzed using backward induction. The link between the first and second stages of the biform game is α_i. It can be understood as a representation of the players' beliefs about the fraction of the coalitional value they could capture in the cooperative part of the game (Brandenburger & Stuart, 2007; Fuentes González et al., 2020). α_i is used to calculate the weighted average of the upper and lower profits in the core of the cooperative stage, reducing the game to a noncooperative game to be solved using the Nash equilibrium concept. Also, the possible coalitions C considered in this study are singleton (no coalition) denoted $\{i\}$, vertical coalitions $\{p, s\}$ and $\{p, r\}$, and the grand coalition $\{p, s, r\}$.

3.3. Incentive structures for promoting CFE

Three incentive structures $k \in \{1,2,3\}$ were designed. $k = 1$ represents the current state of the decision-making situation, where no incentives are added, resulting in the core being a single point because there is no division of additional value between players in a coalition C. Thus, the competition fully determines the players' decisions based on the profit they can obtain from the strategies J and NJ, which is directly assessed by the Nash equilibrium.

In the other structures, it was assumed that players make agreements related to their coalitions that are reinforced by a third party such as a government entity or private investors through the provision of a subsidy to the coalition itself instead of giving a pre-defined partial subsidy to each player. This was considered suitable for better representing the societal desire to promote the participation of AFSC stakeholders who currently are not doing so because of the absence of incentives (see Section 2.1).

As more coalitions with more players are reached, higher the FLW reduction and greater the potential increment in sales thanks to improvement in the eco-friendly corporative image (Kalmykova et al., 2018; Rico Lugo, Kimita, et al., 2023b). That benefits not only the AFSC players in the game through profit increment but also the third party that provides the incentive in terms of the decrease in carbon emissions, blue water impact, and land impact associated with FLW, which finally materializes an action toward the SDGs.

Particularly, $k = 2$ focuses on providing a subsidy Δ for coalitions formed by stakeholders that do not directly compete for clients (vertical coalitions). In this structure, competition alone is not fully determinate because players in the coalition face a residual bargaining problem (Brandenburger & Stuart, 2007), searching for agreements on the distribution of the subsidy. Specifically, Δ is added to the characteristic functions $V(C)$ of the cooperative games CG associated with the strategy profile $(sp_p, sp_s, sp_r) \in \{(J, J, NJ), (J, NJ, J), \}$, which corresponds to a vertical coalition.

Moreover, $k = 3$ includes the provision of a subsidy Δ for the grand coalition (simultaneous vertical and horizontal collaboration) and was designed to particularly reflect the circumstances in which vertical coalitions may not be sufficient to promote the implementation of the CFE project and obtain the desired FLW reduction. Then, Δ is added to the characteristic function of the profile $(sp_p, sp_s, sp_r) = (J, J, J)$.

4. Simulation settings for obtaining the Nash equilibria

To assess the influence of the distinct incentive structures in the CFE decision-making, simulation analysis for obtaining the Nash equilibria of the proposed biform game was necessary. Python and MySQL were utilized for coding and data management.

Initial values for all parameters of the proposed biform model were set based on estimated data validated with AFSC experts, which corresponded to the input values under $k = 1$. Then, variances in the incentive Δ provided by a third party for both $k = 2$ and $k = 3$ were considered, represented as a percentage of increment of the original characteristic function $V(C)$ of the corresponding coalitions.

Also, due to the large number of parameters in the model, and the subsequent huge computational complexity, the most representative cost variables were selected considering the CFE barriers and drivers identified in the previous research phase (Rico Lugo, Kimita, et al., 2023b). Variations in the α_i were also considered to evaluate how the optimistic or pessimistic behavior of players can generate changes in the final formation of coalitions. Let us denote the possible α_i combinations as $\alpha_{comb} = (\alpha_p, \alpha_s, \alpha_r)$.

Table 2 shows a summary of simulation settings, which included changes in seven parameters as independent variables with five variations for each one, totaling 14,062,500 runs.

Table 2. Simulation settings

Independent variable	Variations description	Number of variations
Δ	0 for *k=1*, 20% and 30% for *k=2*, 20% and 30% for *k=3*	5
α_{comb}	(0.1,0.1,0.8), (0.1,0.2,0.7), …, (0.4,0.3,0.3), …, (0.8,0.1,0.1)	36
PC_p	-0.2, -0.1, 0.0, 0.1, 0.2	5
PC_s	-0.2, -0.1, 0.0, 0.1, 0.2	5
PC_r	-0.2, -0.1, 0.0, 0.1, 0.2	5
DC_p	-0.2, -0.1, 0.0, 0.1, 0.2	5
DC_s	-0.2, -0.1, 0.0, 0.1, 0.2	5
DC_r	-0.2, -0.1, 0.0, 0.1, 0.2	5
CEC	-0.2, -0.1, 0.0, 0.1, 0.2	5

5. Results: Effects on profits, social welfare, and coalitions

First, regarding *k=1,* results remain the same across all α_{comb} (see Section 3.3) and are textually summarized as follows.

From the economic aspect of the CFE toward supply chain circulation, it was found that when costs increase, social welfare decreases, with its maximum value ranging from 50 to 1,337 monetary units. The maximum final profits of the stakeholders were 299.0 for p, 501.0 for s, and 537.0 for r. It was found that p's cost highly influences social welfare. r and s's costs show a more gradual increment, resulting in low social welfare when both costs are high, especially for r.

From the collaborative side of the CFE, in this structure, the grand coalition was found in the Nash equilibrium 53,125 times (68.0%), whereas no coalitions appeared in 25,000 runs (32.0%). It was also found that the costs associated with p played a crucial role in the formation of a grand coalition but a high profit x_p did not always translate into a grand coalition. Moreover, it can be said that when the *CEC* cost took its highest value, no coalition appeared even with diminutions in the supermarket costs. Likewise, some high profits for player r were found without a coalition.

Nevertheless, three types of FLW reduction were attained. The first type (Type 1) corresponds to 17.82%, which resulted from the formation of the grand coalition. Type 2 refers to a reduction of 13.79% from the formation of a vertical coalition. Type 3 is the scenario in which no reduction (0%) was reached because no coalition was formed with other stakeholders. Those reduction types remained the same across all the incentive structures because no variations in the original FLW percentage were considered in this study; however, its attainment changed as the percentage of each coalition type changed in each structure.

Second, regarding the other incentive structures, different results were found depending on the α_{comb} because players in a coalition face a residual bargaining problem derived from the added incentive. The colormap in

Figure 4 shows a summary of results found in $k = 2$, while Figure 5 makes the same for $k = 3$. In both figures, the maximum values of the profits and social welfare reached across all the corresponding simulation runs are considered.

In the incentive structure $k = 2$, it was found that the tendency in social welfare is directly linked to the results of the profit of the food processor. This can be observed in the same tendency in the intensity of the green and gray cells in

Figure 4. When the p's profit took high value, the social welfare also took high value. This is also related to the increment in the formation of a vertical coalition because the presence of this type of coalition (dark blue cells) only appeared when the food processor had a notorious increment in its final profit.

α_{comb}	Attainment of a vertical coalition		Social welfare		Profit of food processor		Profit of supermarket		Profit of restaurant	
	k=2 (20%)	k=2 (30%)	k=2 (20%)	k=2 (30%)	k=2 (20%)	k=2 (30%)	k=2 (20%)	k=2 (30%)	k=2 (20%)	k=2 (30%)
(0.1,0.1,0.8)	22%	35%	1412	1494	349.5	387.2	501.0	502.7	565.4	604.6
(0.1,0.2,0.7)	13%	32%	1412	1494	349.5	387.2	507.6	517.4	555.6	589.9
(0.1,0.3,0.6)	5%	27%	1412	1494	349.5	387.2	517.4	532.1	545.8	575.2
(0.1,0.4,0.5)	0%	19%	1337	1494	299.0	387.2	501.0	546.8	537.0	560.5
(0.1,0.5,0.4)	0%	5%	1337	1494	299.0	387.2	501.0	561.5	537.0	545.8
(0.1,0.6,0.3)	0%	5%	1337	1489	299.0	408.6	501.0	511.0	537.0	570.1
(0.1,0.7,0.2)	0%	11%	1337	1486	299.0	408.6	501.0	524.0	537.0	554.4
(0.1,0.8,0.1)	3%	16%	1405	1484	369.3	408.6	502.6	537.0	537.0	538.7
(0.2,0.1,0.7)	13%	32%	1422	1508	369.0	416.4	501.0	502.7	555.6	589.9
(0.2,0.2,0.6)	5%	27%	1422	1508	369.0	416.4	507.6	517.4	545.8	575.2
(0.2,0.3,0.5)	0%	19%	1337	1508	299.0	416.4	501.0	532.1	537.0	560.5
(0.2,0.4,0.4)	0%	5%	1337	1508	299.0	416.4	501.0	546.8	537.0	545.8
(0.2,0.5,0.3)	0%	0%	1337	1337	299.0	299.0	501.0	501.0	537.0	537.0
(0.2,0.6,0.2)	0%	5%	1337	1502	299.0	437.2	501.0	511.0	537.0	554.4
(0.2,0.7,0.1)	0%	11%	1337	1499	299.0	437.2	501.0	524.0	537.0	538.7
(0.3,0.1,0.6)	5%	27%	1432	1523	388.5	445.6	501.0	502.7	545.8	575.2
(0.3,0.2,0.5)	0%	19%	1337	1523	299.0	445.6	501.0	517.4	537.0	560.5
(0.3,0.3,0.4)	0%	5%	1337	1523	299.0	445.6	501.0	532.1	537.0	545.8
(0.3,0.4,0.3)	0%	0%	1337	1337	299.0	299.0	501.0	501.0	537.0	537.0
(0.3,0.5,0.2)	0%	0%	1337	1337	299.0	299.0	501.0	501.0	537.0	537.0
(0.3,0.6,0.1)	0%	5%	1337	1515	299.0	465.8	501.0	511.0	537.0	538.7
(0.4,0.1,0.5)	0%	19%	1337	1537	299.0	474.8	501.0	502.7	537.0	560.5
(0.4,0.2,0.4)	0%	5%	1337	1537	299.0	474.8	501.0	517.4	537.0	545.8
(0.4,0.3,0.3)	0%	0%	1337	1337	299.0	299.0	501.0	501.0	537.0	537.0
(0.4,0.4,0.2)	0%	0%	1337	1337	299.0	299.0	501.0	501.0	537.0	537.0
(0.4,0.5,0.1)	0%	0%	1337	1337	299.0	299.0	501.0	501.0	537.0	537.0
(0.5,0.1,0.4)	0%	5%	1337	1552	299.0	504.0	501.0	502.7	537.0	545.8
(0.5,0.2,0.3)	0%	0%	1337	1337	299.0	299.0	501.0	501.0	537.0	537.0
(0.5,0.3,0.2)	0%	0%	1337	1337	299.0	299.0	501.0	501.0	537.0	537.0
(0.5,0.4,0.1)	0%	0%	1337	1337	299.0	299.0	501.0	501.0	537.0	537.0
(0.6,0.1,0.3)	0%	0%	1337	1337	299.0	299.0	501.0	501.0	537.0	537.0
(0.6,0.2,0.2)	0%	0%	1337	1337	299.0	299.0	501.0	501.0	537.0	537.0
(0.6,0.3,0.1)	0%	0%	1337	1337	299.0	299.0	501.0	501.0	537.0	537.0
(0.7,0.1,0.2)	0%	0%	1337	1337	299.0	299.0	501.0	501.0	537.0	537.0
(0.7,0.2,0.1)	0%	0%	1337	1337	299.0	299.0	501.0	501.0	537.0	537.0
(0.8,0.1,0.1)	0%	0%	1337	1337	299.0	299.0	501.0	501.0	537.0	537.0

Figure 4: Summarized results of structure $k = 2$

α_{comb}	Attainment of a grand coalition		Social welfare		Profit of food processor		Profit of supermarket		Profit of restaurant	
	k=3 (20%)	k=3 (30%)	k=3 (20%)	k=3 (30%)	k=3 (20%)	k=3 (30%)	k=3 (20%)	k=3 (30%)	k=3 (20%)	k=3 (30%)
(0.1,0.1,0.8)	90%	96%	1604	1738	325.7	339.1	527.7	541.1	750.6	857.8
(0.1,0.2,0.7)	90%	96%	1604	1738	325.7	339.1	554.4	581.2	723.9	817.7
(0.1,0.3,0.6)	90%	96%	1604	1738	325.7	339.1	581.1	621.3	697.2	777.6
(0.1,0.4,0.5)	90%	96%	1604	1738	325.7	339.1	607.8	661.4	670.5	737.5
(0.1,0.5,0.4)	90%	96%	1604	1738	325.7	339.1	634.5	701.5	643.8	697.4
(0.1,0.6,0.3)	90%	96%	1604	1738	325.7	339.1	661.2	741.6	617.1	657.3
(0.1,0.7,0.2)	90%	96%	1604	1738	325.7	339.1	687.9	781.7	590.4	617.2
(0.1,0.8,0.1)	90%	96%	1604	1738	325.7	339.1	714.6	821.8	563.7	577.1
(0.2,0.1,0.7)	98%	100%	1604	1738	352.4	379.2	527.7	541.1	723.9	817.7
(0.2,0.2,0.6)	98%	100%	1604	1738	352.4	379.2	554.4	581.2	697.2	777.6
(0.2,0.3,0.5)	98%	100%	1604	1738	352.4	379.2	581.1	621.3	670.5	737.5
(0.2,0.4,0.4)	98%	100%	1604	1738	352.4	379.2	607.8	661.4	643.8	697.4
(0.2,0.5,0.3)	98%	100%	1604	1738	352.4	379.2	634.5	701.5	617.1	657.3
(0.2,0.6,0.2)	98%	100%	1604	1738	352.4	379.2	661.2	741.6	590.4	617.2
(0.2,0.7,0.1)	98%	100%	1604	1738	352.4	379.2	687.9	781.7	563.7	577.1
(0.3,0.1,0.6)	100%	100%	1604	1738	379.1	419.3	527.7	541.1	697.2	777.6
(0.3,0.2,0.5)	100%	100%	1604	1738	379.1	419.3	554.4	581.2	670.5	737.5
(0.3,0.3,0.4)	100%	100%	1604	1738	379.1	419.3	581.1	621.3	643.8	697.4
(0.3,0.4,0.3)	100%	100%	1604	1738	379.1	419.3	607.8	661.4	617.1	657.3
(0.3,0.5,0.2)	100%	100%	1604	1738	379.1	419.3	634.5	701.5	590.4	617.2
(0.3,0.6,0.1)	100%	100%	1604	1738	379.1	419.3	661.2	741.6	563.7	577.1
(0.4,0.1,0.5)	100%	100%	1604	1738	405.8	459.4	527.7	541.1	670.5	737.5
(0.4,0.2,0.4)	100%	100%	1604	1738	405.8	459.4	554.4	581.2	643.8	697.4
(0.4,0.3,0.3)	100%	100%	1604	1738	405.8	459.4	581.1	621.3	617.1	657.3
(0.4,0.4,0.2)	100%	100%	1604	1738	405.8	459.4	607.8	661.4	590.4	617.2
(0.4,0.5,0.1)	100%	100%	1604	1738	405.8	459.4	634.5	701.5	563.7	577.1
(0.5,0.1,0.4)	100%	100%	1604	1738	432.5	499.5	527.7	541.1	643.8	697.4
(0.5,0.2,0.3)	100%	100%	1604	1738	432.5	499.5	554.4	581.2	617.1	657.3
(0.5,0.3,0.2)	100%	100%	1604	1738	432.5	499.5	581.1	621.3	590.4	617.2
(0.5,0.4,0.1)	100%	100%	1604	1738	432.5	499.5	607.8	661.4	563.7	577.1
(0.6,0.1,0.3)	100%	100%	1604	1738	459.2	539.6	527.7	541.1	617.1	657.3
(0.6,0.2,0.2)	100%	100%	1604	1738	459.2	539.6	554.4	581.2	590.4	617.2
(0.6,0.3,0.1)	100%	100%	1604	1738	459.2	539.6	581.1	621.3	563.7	577.1
(0.7,0.1,0.2)	100%	100%	1604	1738	485.9	579.7	527.7	541.1	590.4	617.2
(0.7,0.2,0.1)	100%	100%	1604	1738	485.9	579.7	554.4	581.2	563.7	577.1
(0.8,0.1,0.1)	100%	100%	1604	1738	512.6	619.8	527.7	541.1	563.7	577.1

Figure 5: Summarized results of structure $k = 3$

Overall, structure $k = 3$ had a better performance than $k = 2$ and, as one can anticipate, better than $k = 1$. This is observed by the darker colors of most of the cells in Figure 5. It can also be expected that even while maximum social welfare remains the same in each subsidy type in $k = 3$, its distribution across players changes, affecting the formation of the grand coalition. Particularly, when the p tends to be pessimistic in the bargain, its profits decrease and the percentage of achieving this type of coalition also decreases.

In addition, within the combinations of confidence indices in which $k = 3$ resulted in the formation of the grand coalition 100% times, the most homogeneous distributions of profits can be found when all players have an α_i between 0.3 and 0.4. This is directly related to the formulation of the biform model and the expected influence of the confidence index in the reduction of the cooperative stage into the non-cooperative stage of the game.

6. Discussion on the results toward circular food economy

The results showed that the differences in the distribution of welfare change across players as a result of variations in their confidence indices, ultimately affecting the formation of the grand coalition. Therefore, further analysis of the influence of those indices from a holistic viewpoint including all the incentive structures designed and the subsequent impact on the circulation of FLW in the chain is presented, first, from a general viewpoint of the current decision-making situation in $k = 1$, then, from the viewpoint of Nash equilibria and social welfare results.

First, the correlation matrix in Figure 6 provides a comprehensive overview of the relationships between the quantitative independent and dependent variables across all simulated data and incentive structures, allowing the identification of patterns and dependencies. The first seven rows/columns from the top correspond to the independent variables (production costs, discard costs, and technological costs), while the last four rows/columns are the dependent ones (profits and social welfare). This correlation does not include the coalition achievement in the Nash equilibrium of the games, which is discussed later.

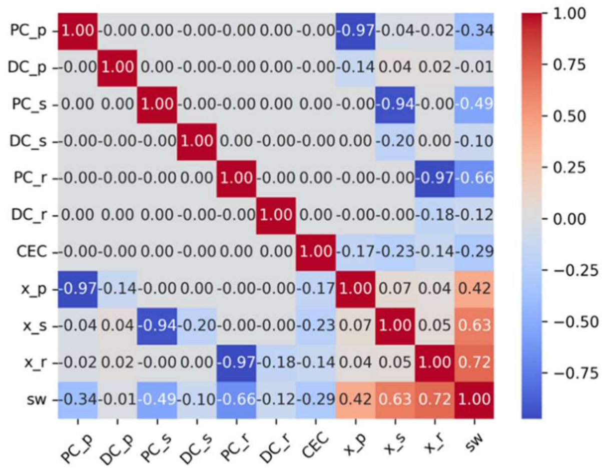


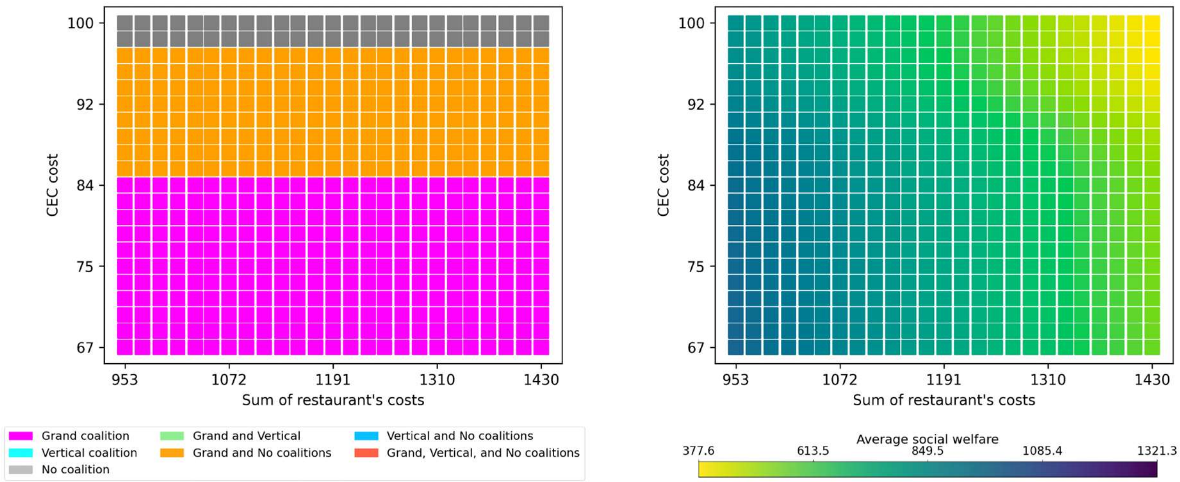
Figure 6: Correlation matrix of simulation data

Various insights are provided by the correlational values. For instance, no multicollinearity was found among the independent variables, demonstrating the reliability of the input. Moreover, the production costs of players, especially of r presented the highest negative impact on social welfare. The *CEC* cost, which was carried by p, was between the production costs correlation and the discard costs impact, which was also relevant to the increase or decrease of social welfare. In addition, the sum of the costs of the restaurant (PC_r and DC_r in the figure) had the most negative effect on social welfare. In terms of positive impacts, social welfare was especially correlated to r's profit, then by s, and finally by p's profit.

Based on the above, the variables for further analysis of the formation of coalitions were prioritized and selected to facilitate an overall discussion on the results, especially as a valuable summarized outcome of this study for supporting the decision-making process around CFE. Specifically, the sum of the restaurant's costs and the *CEC* cost were selected to analyze the changes in the formation of coalitions and the corresponding social welfare across all incentive structures.

6.1. General insights from incentive structure $k = 1$

Figure 7 presents the analysis for structure $k = 1$ as a reference for comparison with the incentive structures in which a subsidy is provided by a third party. Figure 7a shows the regions of formation of each type of coalition, whereas Figure 7b shows the average social welfare reached across the corresponding simulation runs.



(a) Coalition zones (b) Average social welfare zones

Figure 7: Coalitions and social welfare zones in $k = 1$

In both figures, the x-axis represents the sum of costs for the restaurant and the y-axis represents the *CEC* cost. For social welfare, the heatmap range covers the average values obtained across all the incentive structures to keep a standardized range in all graphs in this section.

In this case, three coalition zones are found. Only grand coalitions are formed when the *CEC* cost takes values from the original input value to the reduction of 20% (magenta zone in Figure 7a), whereas no coalitions appear when the *CEC* cost is 17.5% or more of the input value (gray zone). In addition, both grand coalitions and no coalitions appear for any positive increment of less than 17.5% in the *CEC* cost (yellow zone). Those zones of coalitions remain for any value of the sum of the costs of the restaurant within the simulation variations.

Moreover, a direct inverse relationship between the increment of costs and the average social welfare is observed in Figure 7b. This is represented by the green dots becoming lighter as the restaurant cost increases as well as when the *CEC* cost increases. This is an expected result that truly represents the actual situation of AFSC stakeholders.

However, an interesting insight from both graphs is that even in singleton coalitions, relatively high social welfare can be reached. This implies that when no external incentive is provided, the formation of the grand coalition and the subsequent increment of the circulation of FLW and reduction in the environmental impact of the supply chain activities can be ensured only when the current R&D cost (*CEC*) is reduced and higher social welfare is possible. In contrast, cases in which no coalition is reached (singleton coalition) appear even with a small increment in the *CEC* cost, putting pressure on the environmental system by discarding all FLW produced by the players. Then, the provision of incentives is necessary to increase the probability of obtaining coalitions and reaching the social target of the circulation of goods.

6.2. Relevant insights from the Nash equilibria in incentive structures $k = 2$ and $k = 3$

Figure 8 presents the coalition zones of all structures and incentives. The axes values and coalition colors are the same as the ones utilized in Figure 7. In this case, each structure and incentive percentage is in a column, starting from $k = 2$ with $\Delta = 1.2$ on the left until $k = 3$ with $\Delta = 1.3$ on the right. To visualize the effects of the confidence indices in those results, relevant α_{comb} were selected.

The best performance in terms of higher circulation of FLW is seen in $k = 3$ for all combinations of confidence indices, in line with the previous section. Particularly, the achievement of the grand coalition diminishes when p is highly pessimistic and when s is highly optimistic, whereas in other cases, this type of coalition is formed 100% of the time. This can be observed in the behavior of the small yellow zones in the figures of the third and fourth columns.

Regarding the formation of vertical coalitions, the effect of the increment in the subsidy value is evident in the first and second columns, especially when p's confidence is low (Figure 8a,b,e,f,i,j), a subsidy of 20% or

30% has no effect because no vertical coalition is reached (Figure 8m,n). Thus, the actual AFSC decision-makers need to further explore the provision of a higher incentive when the objective is promoting only vertical coalitions and the processor has strong market power.

Moreover, the occurrence of the dark orange zone in the middle of Figure 8j represents the rare case in which any type of coalition can be reached. This situation only happened 6% of the corresponding simulation runs. The combination of costs for players in this particular zone is consolidated as a critical scenario in which a small change in any variable can result in a different coalition type and the subsequent different levels of circulation and reduction of FLW. Hence, decision-makers and policymakers need to consider that when the supermarket has much higher power than others, variations between 0% and 10% in the CEC cost may result in a grand coalition, vertical coalitions, or even no coalitions, making it difficult to ensure the effectiveness of providing an incentive of $\Delta = 1.3$ for forming vertical coalitions within the second structure.

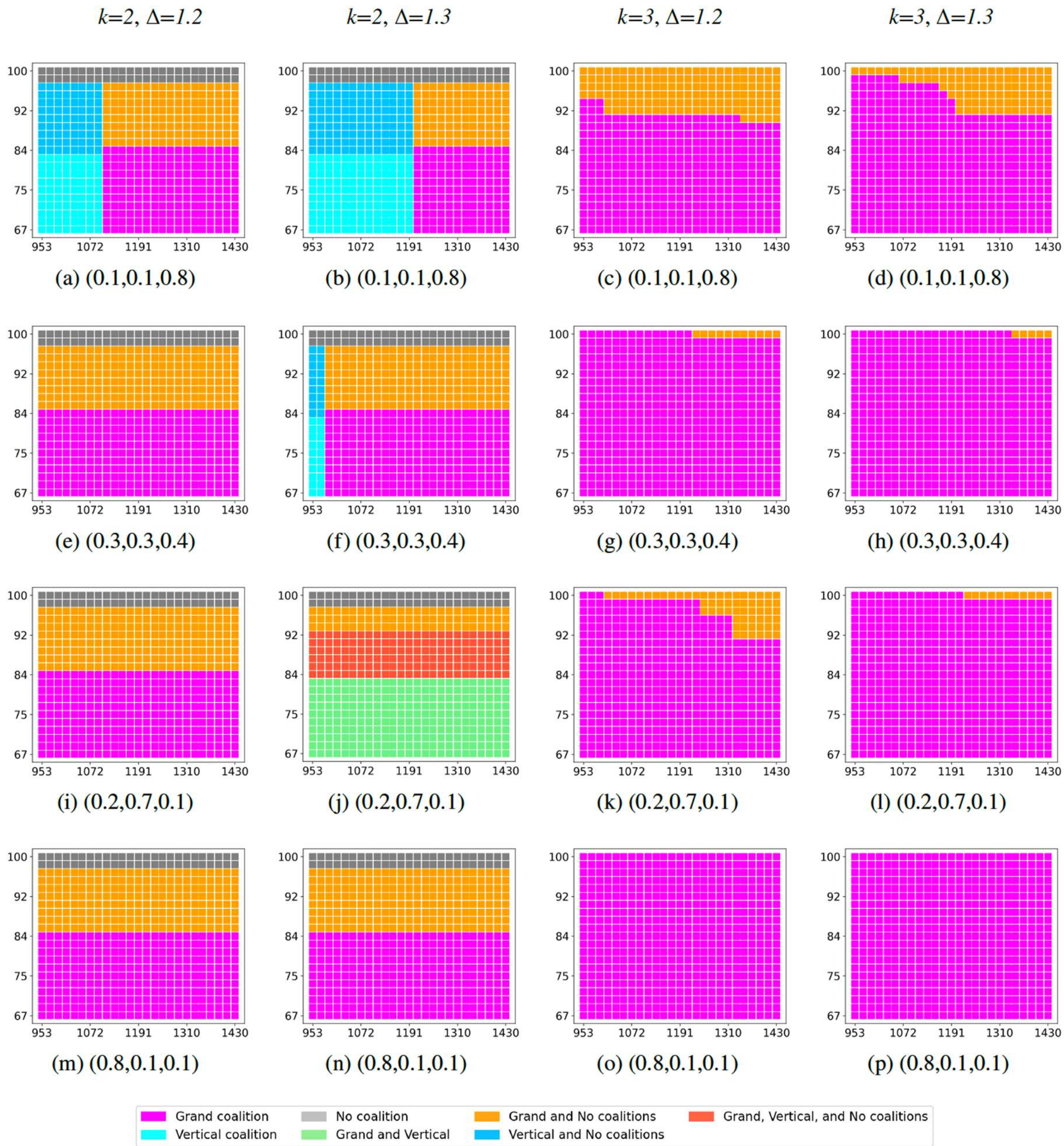


Figure 8: Coalition zones by structure, incentive, and combination of the confidence indices of players. X-axis: sum of r's costs. Y-axis: CEC cost

6.3. Relevant insights from the social welfare in incentive structures $k = 2$ and $k = 3$

Figure 9 shows the behavior of average social welfare that corresponds to the previously discussed Nash zones and combinations of the confidence indices of players. The axes values and heatmap colors are the same as the ones utilized in Figure 7.

One of the most relevant insights from this figure is that the coalition-type zones obtained in Figure 8 are almost not directly reflected in zones of higher or lower average social welfare when comparing the selected costs. This is evident in the overall increase in lighter green dots horizontally and vertically in each subgraph. It indicates that the lowest social welfare occurs when costs are higher, aligning with the AFSC stakeholders' rational tendency to prioritize strategies with better profit performance.

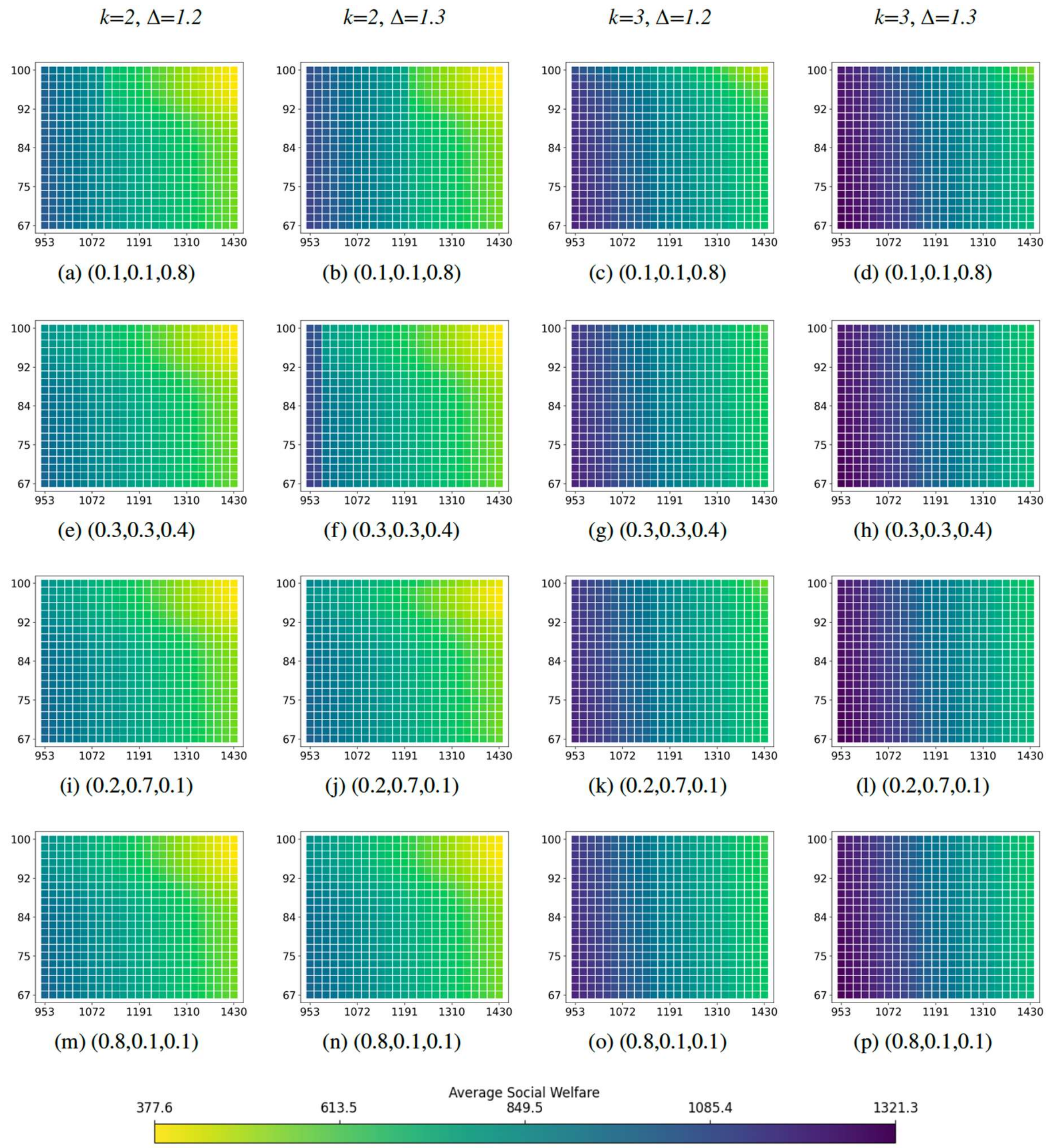


Figure 9: Social welfare zones by structure, incentive, and combination of the confidence indices of players. X-axis: sum of r's costs. Y-axis: CEC cost

Regarding particular zones of social welfare in $k = 3$ (third and fourth columns), the large dark region on the left means that a stable high social welfare can be reached regardless of the considered variations in the CEC cost when the sum of r's costs are equal or less than the original input value for all the confidence indices combinations. In addition, although the small light zone in the top-right of Figure 9c demonstrates the effect of the appearance of some singleton coalitions when the food processor is highly pessimistic, its impact is not as strong as in $k = 2$ or $k = 1$. Overall, within $k = 3$, AFSC stakeholders and the subsidy provider can expect both high social welfare and a high circulation of FLW.

Nevertheless, in relation to $k = 2$ (first and second columns) the effect of the exclusive formation of vertical coalitions is seen in the dark vertical zones on the left of Figure 9a,b,f. In contrast to the third incentive structure, in these AFSC scenarios, vertical collaboration notoriously improves social welfare when the costs of the restaurant are low and the food processor's confidence is relatively low (0.1-0.3). Such an improvement is better than the one offered by forming the grand coalition that occurs when the restaurant's costs increase. This understanding is key within the context of $k = 2$ since the provision of a subsidy of 30% can ensure an average high social welfare only in the particular case of low costs for the restaurant. This implies that although the grand coalition offers a high circulation of FLW, the social welfare is not expected to be higher than when vertical coalitions are present, ultimately affecting the circulation flow and impacting the environment by disposing of some of the FLW in the garbage.

7. Conclusions

This study proposes incentive structures of CFE to increase the profit of stakeholders while minimizing FLW and representing both cooperative and non-cooperative behavior through the formulation of a new biform model. The study demonstrated that there are differences, especially in terms of profits, between the conventional FLW disposal option and CFE projects for reprocessing it. Those differences were used to design the set of incentive structures to encourage collaboration among AFSC stakeholders.

To assess the influence of the distinct incentive structures in the CFE decision-making situation formulated, the analysis considered two different percentages of subsidies provided by a third party in the second and third structures. Within the various parameters included in the game model, changes in seven parameters were considered as independent variables in the simulation settings, with five variations for each variable. Even though additional types and percentages of incentives and more variations may be considered in more model parameters, the simulation performed was sufficient to analyze the effects of the changes in the various model variables within the CFE context, while keeping a feasible limit in this study. Future studies can explore further variations.

In addition, it is important to recognize that because of the sensitivity that providing cost-related data represents for actual AFSC businesses, no real data were used, representing a limitation of this study. However, various ranges of changes in the values of the simulation settings were included in an effort to deal with this constriction.

The simulation results analysis was meaningful because it indicates that the proposed model represents actual AFSC situations in which stakeholders negotiate to collaborate and in which a strong effect of the optimistic or pessimistic behavior of each actor exists, also finding special scenarios and situations in which decision-makers should be careful about the type of incentive structure and the percentage of subsidy to provide within each chain context and the confidence of stakeholders.

Overall, the incentive structure $k = 3$ presented better performance because it provided a higher percentage of attainment of grand coalitions, resulting in higher social welfare, higher profit for all players, and higher FLW reduction than incentive structures $k = 2$ and $k = 1$. Regarding the incentive structure $k = 2$, social welfare was particularly linked to the profit of the food processor.

Although the restaurant contributes with more monetary units, when the food processor's profits take high value, the social welfare also tends to take high value. This is related to the increment in the formation of a vertical coalition because the presence of this type of coalition only appeared when the food processor had a notorious increment in its final profit.

Also, if the interest of a certain AFSC chain context is to reach a vertical coalition for developing the CFE project, then, the best option would be to provide a subsidy of 30% under the incentive structure $k = 2$. In comparison, if the food processor is highly optimistic, then, no vertical coalition is expected under any of the proposed incentive structures. In such a case, the interest can be put in providing the incentive of $k = 3$ to reach 100% of the attainment of a grand coalition.

Finally, the proposed biform model offers a novel quantitative representation of the complexities of addressing the circular economy within the AFSC based on a robust approach, consolidating a guide for designing effective incentives and policies to encourage real-world collaborative efforts to reduce FLW. It provides valuable information for stakeholders, third parties, and the general public to understand the complexity of the system, recognize their role in CFE, and make informed decisions to increase the circularity in the chain in a profitable and eco-friendly manner while tackling various associated obstacles and drivers.

References

Bauwens, T., Hekkert, M., & Kirchherr, J. (2020). Circular futures: What Will They Look Like? *Ecological Economics*, *175*. https://doi.org/10.1016/j.ecolecon.2020.106703

Betz, A., Buchli, J., Göbel, C., & Müller, C. (2015). Food waste in the Swiss food service industry – Magnitude and potential for reduction. *Waste Management*, *35*, 218–226. https://doi.org/https://doi.org/10.1016/j.wasman.2014.09.015

Bocken, N. M. P., de Pauw, I., Bakker, C., & van der Grinten, B. (2016). Product design and business model strategies for a circular economy. *Journal of Industrial and Production Engineering*, *33*(5), 308–320. https://doi.org/10.1080/21681015.2016.1172124

Brandenburger, A., & Stuart, H. (2007). Biform games. *Management Science*, *53*(4), 537–549. https://doi.org/10.1287/mnsc.1060.0591

Chiaraluce, G., Bentivoglio, D., & Finco, A. (2021). Circular economy for a sustainable agri-food supply chain: A review for current trends and future pathways. In *Sustainability (Switzerland)* (Vol. 13, Issue 16). MDPI. https://doi.org/10.3390/su13169294

Collins, B. C., & Kumral, M. (2020). Game theory for analyzing and improving environmental management in the mining industry. *Resources Policy*, *69*. https://doi.org/10.1016/j.resourpol.2020.101860

Derqui, B., Fayos, T., & Fernandez, V. (2016). Towards a more sustainable food supply chain: Opening up invisibleWaste in food service. *Sustainability (Switzerland)*, *8*(7), 1–20. https://doi.org/10.3390/su8070693

Ellen MacArthur Foundation. (2015). Growth within: a circular economy vision for a competitive europe. *Ellen MacArthur Foundation*, 100.

Ellen MacArthur Foundation. (2019). Cities and Circular Economy for Food. *Ellen MacArthur Foundation*, 1–66. https://www.ellenmacarthurfoundation.org/assets/downloads/Cities-and-Circular-Economy-for-Food_280119.pdf

FAO. (2019). *The state of food and agriculture - Moving forward on food loss and waste reduction*. http://www.fao.org/3/ca6030en/ca6030en.pdf

Feess, E., & Thun, J. H. (2014). Surplus division and investment incentives in supply chains: A biform-game analysis. *European Journal of Operational Research*, *234*(3), 763–773. https://doi.org/10.1016/j.ejor.2013.09.039

Fuentes González, F., van der Weijde, A. H., & Sauma, E. (2020). The promotion of community energy projects in Chile and Scotland: An economic approach using biform games. *Energy Economics*, *86*. https://doi.org/10.1016/j.eneco.2020.104677

Fujiwara-Greve, T. (2015). *Monographs in Mathematical Economics, Non-Cooperative Game Theory* (Vol. 1). Springer Japan.

Homrich, A. S., Galvão, G., Abadia, L. G., & Carvalho, M. M. (2018). The circular economy umbrella: Trends and gaps on integrating pathways. *Journal of Cleaner Production*, *175*, 525–543. https://doi.org/10.1016/j.jclepro.2017.11.064

Intergovernmental Panel on Climate Change. (2017). IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse gas fluxes in Terrestrial Ecosystems. *Research Handbook on Climate Change and Agricultural Law*, 1–472. https://doi.org/10.4337/9781784710644

Kalmykova, Y., Sadagopan, M., & Rosado, L. (2018). Circular economy - From review of theories and practices to development of implementation tools. *Resources, Conservation and Recycling*, *135*(February 2017), 190–201. https://doi.org/10.1016/j.resconrec.2017.10.034

Ma, X., Bai, Q., Islam, S. M., Wang, S., & Liu, X. (2019). Coordinating a three-echelon fresh agricultural products supply chain considering freshness-keeping effort with asymmetric information. *Applied Mathematical Modelling*, *67*, 337–356. https://doi.org/10.1016/j.apm.2018.10.028

McCain, R. A. (2009). *Game Theory and Public Policy* (Issue 13096). Edward Elgar Publishing. https://ideas.repec.org/b/elg/eebook/13096.html

Mesa Reyes, J. A., Cardozo, N. J., Soler Becerra, N., Rico Lugo, Sinndy Dayana Estupiñan, A., Gutiérrez, E., Rios Saavedra, S., Navarrete, P., & Serrano, J. D. (2020). *Food Supply Plan of the Central Region of Colombia*.

Miranda-Ackerman, M. A., Azzaro-Pantel, C., & Aguilar-Lasserre, A. A. (2017). A green supply chain network design framework for the processed food industry: Application to the orange juice agrofood cluster. *Computers and Industrial Engineering*, *109*, 369–389. https://doi.org/10.1016/j.cie.2017.04.031

National Planning Department of Colombia. (2016). *Food Loss and Waste Quantification in Colombia*. https://es.slideshare.net/FAOoftheUN/food-loss-and-waste-quantification-in-colombia

Osborne, M. J., & Rubinstein, A. (1994). *A Course in Game Theory* (Vol. 1, Issue 0262650401). The MIT Press. https://ideas.repec.org/b/mtp/titles/0262650401.html

Rico Lugo, S. D., Du, B., Almeida, J. A. de, & Nishino, N. (2023). A circular food economy multicriteria decision problem based on the FITradeoff method. *Pesquisa Operacional*, *43*(spe1), e263528. https://doi.org/10.1590/0101-7438.2023.043spe1.00263528

Rico Lugo, S. D., Kimita, K., & Nishino, N. (2022). Circular Food Economy framework: Challenges and initiatives. *Procedia CIRP*, *112*, 28–33. https://doi.org/10.1016/j.procir.2022.09.019

Rico Lugo, S. D., Kimita, K., & Nishino, N. (2023a). A biform game in the context of the Circular Food Economy. *Procedia CIRP*, *118*, 9–14. https://doi.org/10.1016/j.procir.2023.06.003

Rico Lugo, S. D., Kimita, K., & Nishino, N. (2023b). Characteristics of decision process towards circular food economy: A review. *Cleaner Logistics and Supply Chain*, *7*. https://doi.org/10.1016/j.clscn.2023.100104

Rico Lugo, S. D., Kimita, K., & Nishino, N. (2024). Food Loss and Waste-Mixed analysis from the Circular Food Economy. In *EcoDesign 2021, EcoDesign for Sustainable Products, Services, and Social Systems* (II). Springer (In press). https://www.amazon.co.jp/-/en/Shinichi-Fukushige/dp/9819938961

Schröder, P., Anggraeni, K., & Weber, U. (2018). The Relevance of Circular Economy Practices to the Sustainable Development Goals. *Journal of Industrial Ecology*, *23*. https://doi.org/10.1111/jiec.12732

Teigiserova, D. A., Hamelin, L., & Thomsen, M. (2020). Towards transparent valorization of food surplus, waste and loss: Clarifying definitions, food waste hierarchy, and role in the circular economy. *Science of the Total Environment*, *706*. https://doi.org/10.1016/j.scitotenv.2019.136033

Wübbeke, J., & Heroth, T. (2014). Challenges and political solutions for steel recycling in China. *Resources, Conservation and Recycling*, *87*, 1–7. https://doi.org/10.1016/j.resconrec.2014.03.004

Multi-criteria decision analysis to frame the implementation of ecosystem accounting data in land use planning

Arron Wilde Tippett [1*]

[1] Norwegian University of Science and Technology, Ålesund, Norway
[*] arron.w.tippett@ntnu.no

Abstract

Biodiversity and ecosystem services are in decline with land use change one of the key drivers. Ecosystem accounting is hoped to address this issue by providing annual global data on the extent of ecosystems, their condition and the ecosystem services they provide. Spatial planning decisions in OECD countries have moved away from rational planning towards communicative planning, where participation is the mechanism for knowledge gathering and all knowledge is treated as socially constructed. The use of ecological data in planning decisions is thus of interest for future implementation of ecosystem accounting data in spatial planning and decision support. We focus on how ecological data is used in Norwegian municipality planning decisions. We used interviews to gather data on planning practice in Norway and to provide insight into how ecosystem accounting data may impact spatial planning. We make two proposals based on the theory of telecoupling and the relational theory of risk to improve participation through the use of ecosystem accounting data. We use MCDA as a framework to describe spatial planning in Norway and how participation results in the activation of criteria and weights for ecological data.

Keywords: MCDA, land-use planning, ecosystem accounting.

1. Introduction

Natural capital and biodiversity are under increasing pressure from pollution, overexploitation of resources and land-use change driven by industries, such as agriculture, energy and transport (IPBES, 2019). Humans are dependent on natural capital and the ecosystem services that flow from it, such as the removal of pollutants from air/water, protection from storms, food provision, clean water, and improved mental health (Barbier, 2017, Orth et al., 2020, Bratman et al., 2019). Ecosystem-based planning has long been hailed as the solution for managing complex socio-ecological systems and International targets have aimed at integrating the value of ecosystems into planning, such as the United Nations Sustainable Develop Goal 15 target 15.9, to "integrate ecosystem and biodiversity values into national and local planning, development processes, poverty reduction strategies and accounts" (UN, 2015). However, since the target was not met in 2020, as planned, there is even more urgency to act. Our market driven policies and consumerism have been blamed for the rampant consumption of natural capital and the inequitable distribution of gains and burdens of that consumption (Díaz et al., 2019). Even though private or government ownership of the land, is around 82 % globally (Rights and Resources Initiative, 2015), ecosystems and common pool resources are continually overexploited (IPBES, 2019).

Recent developments in the standardisation of Natural Capital accounting (United Nations, 2021) present a possibility to build knowledge in the ecosystem services integrated approach for land use planning. The European Commission and United Nations have been working over the past decade to develop natural capital accounting methodologies (European Commission. Directorate General for the Environment., 2013; European

Commission. Joint Research Centre., 2020) culminating in the publication in March 2021 of the United Nations System of Environmental-Economic Accounting – Ecosystem Accounting (United Nations, 2021). The UN's extensive accounting system, based on the System of National Accounts (SNA), provides a framework for the National mapping and tabulation of natural capital: ecosystem assets, their condition, and the related ecosystem service flows. The new accounts will provide cyclical data on the connections between natural and social systems which has thus far been unavailable. The repercussions for land use planning could be significant as planners will have access to data relating to key benefits that are received by communities and the ecosystems that provide them, therefore insights into the future risks associated with land use change. Ecosystem accounting may not be, however, the silver bullet to the ongoing biodiversity crisis (Bekessy et al., 2018) as it only values ecosystems that provide benefits to the socio-economic system. Ecosystem accounting does, however, provide an opportunity for the integration of socio-ecological risks in the planning system. Many ecosystem services are focussed on the protection from hazards that ecosystems offer us, such as flood attenuation and the reduction of air pollution. These multiple of services allow ecosystem services to be framed to match the values of different actors (Altinay, 2017; Opdam et al., 2015).

A key challenge to the implementation of new ecological data, including ecosystem accounting data, in land use planning comes from the current communicative rationality theory which underlies many planning systems. Instrumental rational planning was the post WW-II ideal, where technocrats designed and built towns and cities based on their ability to digest scientific knowledge and provide solutions that met the public interest now and into the future (Gunder et al., 2017). These ideals came under attack retrospectively after the move away from the welfare state and the demise of instrumental rational planning meant the rise of the communicative rational planning, incorporating local knowledge into the system through participatory approaches (Gunder et al., 2017). The communicative approach is now used in many parts of the world (OECD, 2017). The rational communicative theory is the backbone to planning in Norway, where participation is used to understand the preferences of affected actors and for the planner to come to a consensus (Nordahl, 2017). In this system, ecological and other data types are viewed as socially constructed alongside knowledge and data gathered during the participation process. The spatial planner is key to communicative planning but the assumptions that they can suppress their personal values, gather decision criteria and weights and efficiently calculate the objective function for planning projects have been questioned (Gunder et al., 2017). Critics of this approach believe that it ignores issues of power, assumes all actors to be altruistic (Gunder et al., 2017; Nordahl, 2017; Stjernström et al., 2023) and has made the planner a "designer of efficient and democratically acceptable market institutions" (Sager, 2009). Scholars note the importance of participation for good planning decisions but do not address how to activate participation through the use of ecological data. In this article we assess the zonal planning and participation practices of Norwegian municipality planners and we draw on two theories to propose how ecosystem accounting implementation can activate additional participants.

First, drawing on the theory of telecoupling from the ecosystem services literature (Martín-López et al., 2019) we suggest how an ecosystem accounting planning support system could activate additional stakeholders in the planning decision. Knowledge gathering phases of zonal planning decisions are based on the communicative theory of planning, whereby participation is used to "create" knowledge for use in planning decisions. In the Norwegian zonal planning decisions, the rules of participation limits the rights to be heard to those directly affected by plans. This is based on geographical proximity of households and businesses. Telecoupling of ecosystem services tells us that direct beneficiaries of ecosystem services (flood attenuation, air filtration, noise dampening, etc.) can be geographically removed from the ecosystem supply. For example, a flood plain in the uplands may have flood attenuation benefits for businesses and households based in a town in the lowlands. Telecoupling allows us to test the meaning of directly affected stakeholders within the planning system and provides an opportunity for activating additional participants in planning.

Second, we use the relational theory of risk to suggest how the framing of ecosystem accounting data can activate additional stakeholders within the planning decision. The relational view of risk informs us that an object has to have value in order to be at risk, moreover, the object at risk and the risk object (the object causing the damage) can reverse, depending on the perspective of the actor (Boholm and Corvellec, 2011). In planning, ecological data is often related to highlighting the location of species and ecosystems with some species and ecosystems protected by the law. In the cases of species or ecosystems without legal protection, it is up to the planner/developer and the participatory process to highlight the value of those aspects and which capabilities will be lost if they are damaged. If the participants in the planning decisions do not value the species or ecosystems, those ecological aspects are omitted from the process. Ecosystem services, or nature's contributions to people, use alternative value frames as opposed to relying on the intrinsic value of nature (Bekessy et al., 2018). Since the first global evaluation of ecosystem services in 1997 (Costanza et al., 1997) , the concept has been gradually developing and is now visible within multiple international frameworks, such as the SEEA-EA and the Montreal-Kunming agreement (Hughes and Grumbine, 2023). Critics of ecosystem services believe that it is crowding out people's intrinsic value of nature and risks decision makers using it simply as a part of monetary cost-benefit analysis (Bekessy et al., 2018). However, ecosystem services provides us with multiple potential frames for communicating the value of nature: provisioning, regulating, cultural services; local versus global services; ecosystem service providers and beneficiaries; and more (Altinay, 2017; Gómez-Baggethun and Barton, 2013; Opdam et al., 2015). Frameworks focussing on the intrinsic value of nature, such as the red lists of species, provide risk assessments of species and ecosystems and categorise their vulnerability (Hochkirch et al., 2023). These lists rely on either legislative protection of vulnerable species or a decision maker treating these species or as objects of value to include them in the decision-making process. Ecosystem services can be framed to match the values of different actors, broadly framing ecosystem services towards sustainability, socio-cultural and economic values can result in increased participation in decision-making (Opdam et al., 2015). It is also possible to engage actors by using local ecosystem services as a frame, as opposed to global benefits, such as global climate regulation (Altinay, 2017). We propose that the multiple frames that are provided through ecosystem accounting can help to improve participation in planning decisions by activating actors using their goals and values.

We used semi-structured interview with actors within the zonal planning system to understand the practical application of communicative planning and the use of ecological data in planning decisions. We used the theory of telecoupling and the relational theory of risk to propose how ecosystem accounting data may enhance participation in zonal planning decisions.

2. Methods

2.1. Decision context

In Norway, there is growing concern about the destruction of nature due to market driven developments. A recent report in the national news, and connected documentary, has highlighted the problem with the use of satellite images to highlight 44,000 cases of encroachment into nature in the past 5 years, with many examples of valuable ecosystems being completely removed to make way for grey infrastructure (NRK, 2024). The actor with the most authority for land use planning in Norway is the local municipality through the Planning and Building Act (2009). The PBA states that one of the primary objectives of planning is sustainable development, a balancing of economic, social and environmental needs. This is achieved through an open planning process, which can involve stakeholders representing economic, social and environmental interests. Land use planning at the municipality level is made in 3 broad ways. At the top-level municipality master plans are developed, if required (municipalities can chose to prolong existing plans) in each 4 year political cycle. These master plans provide an extensive plan of the municipality area, with existing land use types mapped and areas set aside for developments highlighted. The second broad level of plan is the area zoning plan. This type of plan is normally used in urban context to provide development rules for a neighbourhood sized area of a city. The third type is the detailed zoning plan which sets out specific rules for an area. Detailed zoning plans make up the majority of plans in Norway and, although each plan is owned by the municipality, around 70 % of detailed zoning plans are proposed by private interests (Stjernström et al., 2023). New zoning plans require planners to make decisions at multiple time points.

There are several points at which the planner is open to taking data/knowledge into account. The first is at the beginning of the process, when the planning proposal is made by a private interest, a notification letter is sent to directly affected stakeholders and interested parties (other municipality departments, the county governor, environmental non-governmental organisations) to give an opportunity for early participation (Figure 1) The second is late in the planning process where a public hearing (Figure 1) is organised for the plan, at this stage a significant change to the plan is unlikely due the large amount of resources already invested in the plan and one political review has been carried out.

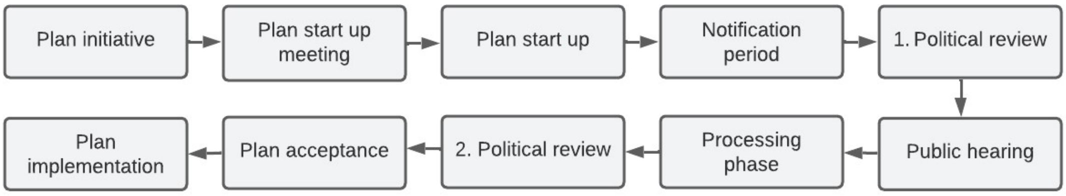


Figure 1: The phases of the Norwegian municipality zonal planning system

We used semi-structured interviews to gather data on the use of ecological data in municipality land use planning decisions in Norway. We stratified our sample into high performing, moderately performing and low performing municipalities using the Sabima ratings (Sabima 2024). Sabima, a Norwegian environmental non-governmental organisation, produces annual reports on municipality performance based on data mainly collected from the statistics bureau of Norway (SSB). Performance indicators based on these data cover a range of themes, such as planned developments, nature protection and dispensations (a tool used to bypass planning regulations). Our sample group was made up of municipality planners, county planners and county governors. The interview data was used to understand to understand the practical implementation of the Planning and Building Act, particularly how ecological data is used in planning decisions and how participation works in practice.

Multi-criteria decision analysis (MCDA) is used as a framework for describing the current use of ecological data in planning decisions, particularly how knowledge is gathered and how participation is used to add weights to decision criteria.

3. Results and Discussion

We interviewed 15 stakeholders from the planning system. 12 municipality planners, 1 county planner, 2 county governors,

Some of the key results from the interviews were:

- If the planner approves the plan start up, this signals their backing of the plan without any participation having taken place. The further the plan moves through the planning phases, the more difficult it is to significantly alter the plan.
- Participation is most effective at the notification period after this point, the decision is unlikely to significantly influenced as the developer will have invested significant resources into the plan.
- The notification period can activate participation from multiple stakeholders, including directly affected householders and businesses as well as the county governor, other municipality departments, the Norwegian water and energy authority and more.
- Municipality planners rely on a GIS tool containing a variety of data sets on the biodiversity, land use, energy, socio-economic aspects and more. This DOK data is the planner’s primary source of data for initial knowledge gathering at the start up phase.
- Only data that meets regulatory requirements will be included in final plans. This limits the data/knowledge gathering to approved data sets.
- DOK data and additional approved data sets are activated through participation at the notification and public hearing phases.

3.1. Implementation of ecosystem accounting into zonal planning

In our conception of the planning system the ecological data available to a planning decision is used as a proxy for the MCDA criteria and the stakeholders participating in the planning decision provide the weights to those criteria. **Erro! A origem da referência não foi encontrada.**2 provides a simplified visualisation of how data/knowledge moves through the detailed zonal planning. The first phase of planning, the start-up phase involves meetings between the planner and the developer. Each municipality uses a set of approved map layers (DOK) which are used for planning tasks. The contents of the DOK is agreed by the municipality and all data sets are updated daily. All the data sets are created by either governmental organisations or research organisations and cover planning themes from nature, societal security, and energy. The second phase in **Erro! A origem da referência não foi encontrada.** is the open for participation phases. this is a simplification of the planning process as we have two separate stages for participation (**Erro! A origem da referência não foi encontrada.**). For clarity, we visualise one. In this phase, the DOK data is also relevant as participating stakeholders may highlight specific aspects of these data to the planner, such as a protected ecosystem. Additional approved nature data, that has not been added to the municipality's DOK, may also be brought into the planning decision by the county governor, particularly aspects of national interest such as cultural landscapes or protected species.

We provide a model of how ecosystem accounting data may be implemented into zonal planning decisions in Norway. Drawing on the relational theory of risk, we propose that the multiple data frames provided by ecosystem accounting will allow additional participants to be activated in the planning process. The relational theory of risk tells us that in order for an object to be at risk, it has to be valued. Ecosystem accounting provides us with multiple frames for the instrumental value of nature, for example flood attenuation, air filtration, pollination services for crops. These multiple frames can be used to activate participants who value these services rather than participants who simply value the ecosystems themselves. Thus, we propose that ecosystem accounting will activate existing and additional actors within the planning system by bringing to light ecosystem services that overlap with their goals and values.

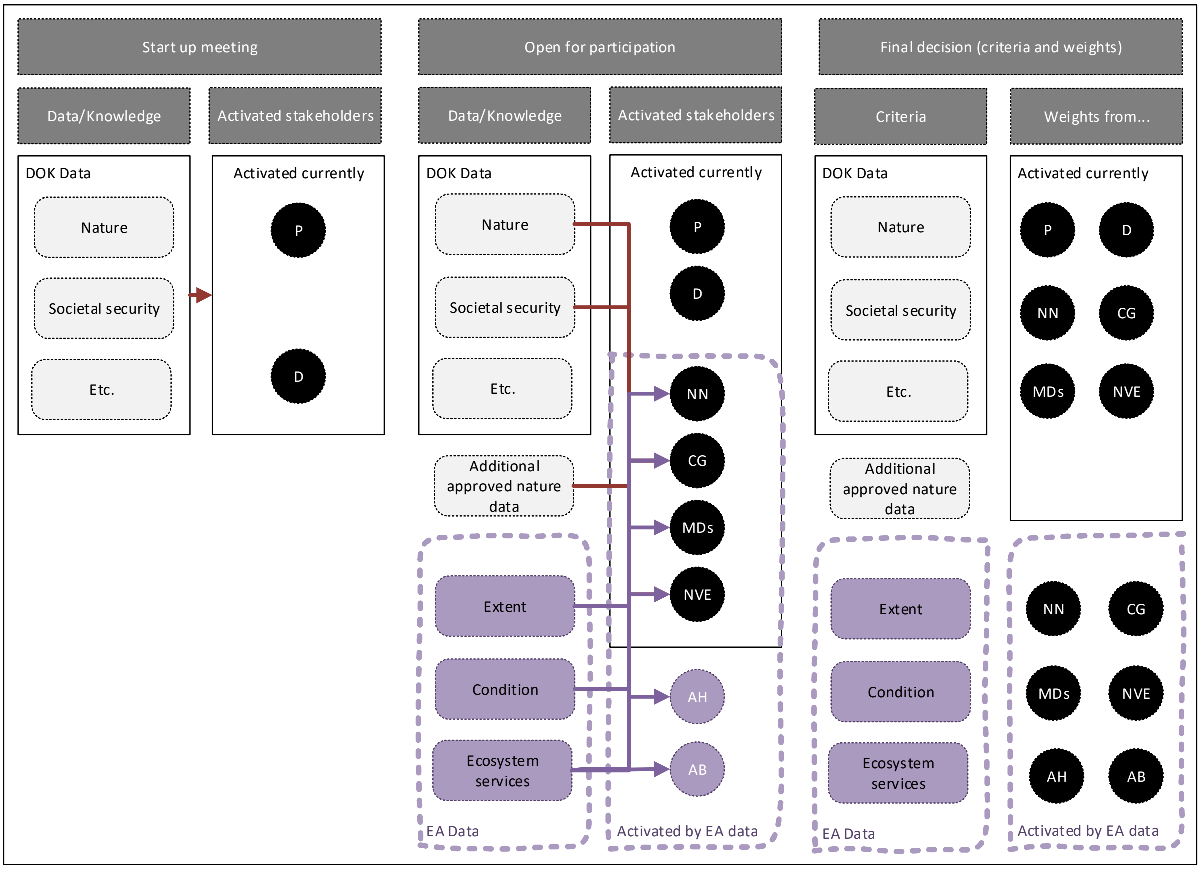


Figure 2: Broad zone planning processes highlighting the use of data/knowledge and the stakeholders activated by different data. (P=Planner, D=Developer, NN=Nearest Neighbour, CG=County Governor, MDs=Municipality Departments, NVE=Norwegian Water and Energy Authority, AH=Affected Households, Affected Businesses,

DOK=The Municipality Map Data, EA=Ecosystem Accounting)

The second phase of the planning process also involves the activation of those directly affected households and business to the project. In the current planning system, it is legally required to involve local, affected stakeholders. A buffer zone around the proposed development is created and those stakeholders within the zone are included. The theory of tele-coupling informs us that ecosystem service beneficiaries can be geographically removed from the ecosystems supplying the benefits. Ecosystem accounting, including data on the supply and demand of ecosystem services will provide municipalities with data on the location of beneficiaries and the ecosystems providing them. This will thus provide an opportunity to include additional affected stakeholders without changing making changes to the planning legislation. Households and businesses which currently benefit from ecosystem services, such as air filtration and flood protection, but are spatially removed from the ecosystems, see (Martín-López et al., 2019), should be included early in the planning process as affected stakeholders, similar to the nearest neighbour.

4. Conclusions

This article provides an insight into the use of ecological data in zonal planning decisions in Norwegian municipalities and suggests how ecosystem accounting data can be used to activate additional participation in planning decisions. MCDA provided a frame for conceptualising how ecosystem accounting can enhance participation and provide weights to ecological data (criteria) within zonal planning decisions.

References

Maes J, T. A., Erhard M, Liquete C, Braat L, Berry P, Egoh B, Puydarrieux P, Fiorina C, Santos F, Paracchini ML,, Keune H, W. H., Hauck J, Fiala I, Verburg PH, Condé S, Schägner JP, San Miguel J, Estreguil C, Ostermann O,, Barredo JI, P. H., Stott A, Laporte V, Meiner A, Olah B, Royo Gelabert E, Spyropoulou R, Petersen JE, Maguire, & C, Z. N., Achilleos E, Rubin A, Ledoux L, Brown C, Raes C, Jacobs S, Vandewalle M, Connor D, Bidoglio G. (2013). Mapping and Assessment of Ecosystems and their Services. An analytical framework for ecosystem assessments under action 5 of the EU biodiversity strategy to 2020. Luxembourg: Publications office of the European Union

Maes J, T. A., Erhard M,Grizzetti B,Barredo JI,Paracchini ML,Condé S,Somma F,Orgiazzi A,Jones A,Zulian A,Vallecilo S,Petersen JE,Marquardt D,Kovacevic V,Abdul Malak D, Marin AI,Czúcz B,Mauri A,Loffler,BastrupBirk A,Biala K,Christiansen T,Werner B. (2018). Mapping and Assessment of Ecosystems and their Services: An analytical framework for ecosystem condition. . Luxembourg: Publications office of the European Union

Altinay, Z., 2017. Visual Communication of Climate Change: Local Framing and Place Attachment. Coast. Manag. 45, 293–309. https://doi.org/10.1080/08920753.2017.1327344

Bekessy, S.A., Runge, M.C., Kusmanoff, A.M., Keith, D.A., Wintle, B.A., 2018. Ask not what nature can do for you: A critique of ecosystem services as a communication strategy. Biol. Conserv. 224, 71–74. https://doi.org/10.1016/j.biocon.2018.05.017

Boholm, Å., Corvellec, H., 2011. A relational theory of risk. J. Risk Res. 14, 175–190. https://doi.org/10.1080/13669877.2010.515313

Costanza, R., d'Arge, R., de Groot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'Neill, R.V., Paruelo, J., Raskin, R.G., Sutton, P., van den Belt, M., 1997. The value of the world's ecosystem services and natural capital. Nature 387, 253–260. https://doi.org/10.1038/387253a0

European Commission. Directorate General for the Environment., 2013. Mapping and assessment of ecosystems and their services: an analytical framework for ecosystem assessments under action 5 of the EU biodiversity strategy to 2020 : discussion paper – final, April 2013. Publications Office, LU.

European Commission. Joint Research Centre., 2020. Mapping and assessment of ecosystems and their services: an EU wide ecosystem assessment in support of the EU biodiversity strategy : supplement (indicator fact sheets). Publications Office, LU.

Gómez-Baggethun, E., Barton, D.N., 2013. Classifying and valuing ecosystem services for urban planning. Ecol. Econ., Sustainable Urbanisation: A resilient future 86, 235–245. https://doi.org/10.1016/j.ecolecon.2012.08.019

Gunder, M., Madanipour, A., Watson, V. (Eds.), 2017. The Routledge Handbook of Planning Theory, 1st ed. Routledge, New York, NY : Routledge, 2018. https://doi.org/10.4324/9781315696072

Hochkirch, A., Bilz, M., Ferreira, C.C., Danielczak, A., Allen, D., Nieto, A., Rondinini, C., Harding, K., Hilton-Taylor, C., Pollock, C.M., Seddon, M., Vié, J.-C., Alexander, K.N.A., Beech, E., Biscoito, M., Braud, Y., Burfield, I.J., Buzzetti, F.M., Cálix, M., Carpenter, K.E., Chao, N.L., Chobanov, D., Christenhusz, M.J.M., Collette, B.B., Comeros-Raynal, M.T., Cox, N., Craig, M., Cuttelod, A., Darwall, W.R.T., Dodelin, B., Dulvy, N.K., Englefield, E., Fay, M.F., Fettes, N., Freyhof, J., García, S., Criado, M.G., Harvey, M., Hodgetts, N., Ieronymidou, C., Kalkman, V.J., Kell, S.P., Kemp, J., Khela, S., Lansdown, R.V., Lawson, J.M., Leaman, D.J., Brehm, J.M., Maxted, N., Miller, R.M., Neubert, E., Odé, B., Pollard, D., Pollom, R., Pople, R., Presa Asensio, J.J., Ralph, G.M., Rankou, H., Rivers, M., Roberts, S.P.M., Russell, B., Sennikov, A., Soldati, F., Staneva, A., Stump, E., Symes, A., Telnov, D., Temple, H., Terry, A., Timoshyna, A., Swaay, C.V., Väre, H., Walls, R.H.L., Willemse, L., Wilson, B., Window, J., Wright, E.G.E., Zuna-Kratky, T., 2023. A multi-taxon analysis of European Red Lists reveals major threats to biodiversity. PLOS ONE 18, e0293083. https://doi.org/10.1371/journal.pone.0293083

Hughes, A.C., Grumbine, R.E., 2023. The Kunming-Montreal Global Biodiversity Framework: what it does and does not do, and how to improve it. Front. Environ. Sci. 11, 1281536. https://doi.org/10.3389/fenvs.2023.1281536

Martín-López, B., Felipe-Lucia, M.R., Bennett, E.M., Norström, A., Peterson, G., Plieninger, T., Hicks, C.C., Turkelboom, F., García-Llorente, M., Jacobs, S., Lavorel, S., Locatelli, B., 2019. A novel telecoupling framework to assess social relations across spatial scales for ecosystem services research. J. Environ. Manage. 241, 251–263. https://doi.org/10.1016/j.jenvman.2019.04.029

OECD, 2017. Land-use Planning Systems in the OECD: Country Fact Sheets. Organisation for Economic Co-operation and Development, Paris.

Opdam, P., Coninx, I., Dewulf, A., Steingröver, E., Vos, C., van der Wal, M., 2015. Framing ecosystem services: Affecting behaviour of actors in collaborative landscape planning? Land Use Policy 46, 223–231. https://doi.org/10.1016/j.landusepol.2015.02.008

Stjernström, O., Junker, E., Thorsen, H.W., 2023. The private in the public: The case of Norwegian private zoning plans. Land Use Policy 127, 106585. https://doi.org/10.1016/j.landusepol.2023.106585

United Nations, 2021. System of Environmental-Economic Accounting - Ecosystem Accounting (SEEA EA). United Nations.

Decision Support Systems for Local Politicians on Socio-Environmental Issues, Lessons Learnt from a Case Study in Ålesund, Norway

Christina Carrozzo Hellevik
Hadi Mohammed
Razak Seidu

Norwegian University of Science and Technology, Ålesund, Norway
* christina.hellevik@ntnu.no

Abstract

This case study explores a political decision-making process around the protection of a drinking water reservoir. It analyses information processing, including individual and group biases. A DSS is used to model future scenarios of pollution from road traffic, and reactions are studied. Mistrust and questioning of the model's accuracy were observed and generally no improvements in their understanding of information regarding risks. The DSS scientific community is urged to look closer into how it can support similarly important socio-environmental decisions at political level.

Keywords: DSS; Political decision; socio-environmental decision; case study

1. Introduction

Current and future environmental challenges make it more urgent than ever for policy-makers to improve decision-making processes by including evidence and expert advice, as well as increasing the transparency and accountability of their decisions. Human-induced environmental changes are characterised by high levels of complexity and uncertainty. The implications are that non-expert decision-makers applying a political logic alone to their decisions will face increasingly high risks of catastrophic effects which may lead to social and economic instability.

The rich literature from operations research on the use of rationality-based decision-support systems, and Multi-Criteria Decision Analysis approaches has potential to assist decision-makers in navigating the complex problems our societies face. However, too few empirical studies exist in the political domain to guide the community in this direction. Theories of policy cycle adapt Simon's phases (1960) of decision-making to the political context but apply the same logic of goal-directed planning, which is not often observed in practice, however desirable and practical it might be. Other theories from the political and social science field include incrementalism (Lindblom, 1959), whereby policy making is described as muddling through and making incremental changes. Implementability theory on the other hand, implies that the obstacles to a particular decision determine the choice of alternatives rather than the desirability of the outcomes (Bana e Costa et al., 2014).

In addition to challenges and conflicting views on how decision-making is structured in political contexts, another layer of complexity is added by the often muddled and opaque nature of the negotiation process and the decision-makers' seemingly unstable goals. In many contexts, researchers lack the materials to analyse the political decisions. We would add also a reluctance of politicians to engage in such processes, as clearly highlighted by Andresson et al. (2012) and a possible perceived superiority of the political processes over expert, technocratic ones.

Previous calls for more empirical research in applying DSS to political decisions, notably from Andersson et al. (2012), have remained largely unanswered. Andersson et al. (2012), as one of the very few articles that report on the full deployment of a DSS mapping stakeholder preferences and pointing to the optimal decision in a political context, found considerable resistance when it came to considering the proposed alternative by the politician. Their experience highlight mismatches between the assumptions and theory behind DSS and political processes. This makes exchanges between fields that analyse DSS use and fields looking at negotiation processes, as in this conference, even more necessary to try and decrypt the dynamics of political decisions, and attempt to better tailor both the decision support, and the timing of their introduction.

One particularity of environmental decision-making is that it often involves risks which may be more or less far removed in space and time. This means that the consequences of present decisions may not be felt by the community making the decision until several years in the future, or may never be felt by that particular community but may strike a neighbouring community or one geographically distant. Uncertainty levels around the models and the complexity of Earth systems make it difficult to predict consequences with high levels of accuracy. This makes human cognitive limitations more salient, and many cognitive biases have been identified and pointed out is studies of environmental decision-making (e.g., Richter et al., 2023; Zhao & Luo, 2021). Researchers on climate change communication have highlighted our general limited ability to project ourselves in the future and have therefore called for decision-support tools providing simulated future scenarios (Burch et al., 2009; Sheppard et al., 2011; Wong-Parodi et al., 2020).

Bearing all these elements into consideration, this study was carried out testing a DSS with a focus on simulated future scenarios based on the alternatives available to the politicians. We studied the socio-environmental issue of the main drinking water reservoir for over 60 000 people, namely Brusdalsvatn in Ålesund Norway. The case highlights issues of environmental and health risk linked to cumulative contamination mainly from a nearby motorway. Each alternative had an associated risk level which was both illustrated through visualisations and translated in terms of probabilities. The background of the decision in question was researched through document analysis and the political decision-making process at local municipality level was studied after the presentation of the DSS results.

2. Research Methods

The roadmap for building theories from case study research by Eisenhardt (1989) was followed closely. The first step was to define research questions, which are presented below.

R.Q. 1 How do decision-makers understand information on health and ecological risks?

R.Q.2 How do decision-makers include information on health and ecological risks in their decisions?

Then, the case was selected as explained in the introduction, as the adoption and implementation of a management strategy of the drinking water reservoir in Ålesund, Norway. Then, the protocols were created; i.e., the data collection was carried out by finding all the political meetings relating to the case using a municipality online tool to search “Brusdalsvatn” (https://alesund.kommunetv.no/search), downloading the supporting documents including official reports and consultation responses (n=14) from experts and actors with some level of authority in the case (County Governor, Senior municipal doctor, Public Health Institute, Food Safety Authority, Norwegian Water, see next section), and going through the recordings of the meetings on the channel KommuneTV, Ålesund. In addition, all the newspaper articles in the local newspaper (Sunnmørsposten, n=80) were also downloaded after searching for “Brusdalsvatn”. The material was analysed qualitatively to gain a thorough understanding of the background of the case. Then, hypotheses were formulated to answer the research questions and an intervention using a DSS was designed to test the hypotheses.

The first hypothesis that was formulated was that the decision-makers do not have enough understanding of the scientific information and of the health and ecological risks supporting the proposed management strategy. Elements such as cumulative effects of pollution, hydrodynamic models explaining water turnover, and the concrete health risks posed by various contaminants are not explicitly described in the supporting documents. Moreover, the format in which the scientific evidence is provided makes it difficult for non-experts to understand risk levels associated with the various contaminants and their sources. Indeed, there are gaps both in the risk formulation, including basic probability rates, and in the description of cause-effect dynamics between pollution and the deteriorating condition of the lake which may create miscommunication with the decision-makers, if these do not already possess the adequate mental models to understand.

It was further hypothesised that visualisations representing the lake's deteriorating condition could help understand and include the risk information into the decision. Finally, it was hypothesised that a short temporality is applied by the decision-makers in the current case (Hansson, 2005) and that presenting future scenarios including probability rates for alternative management strategies could help apply the long temporality required for a sustainable management of the catchment and include risk information in the decision. The researchers attended the political meeting in person, some presented the results from the DSS; while others took notes on the responses, questions and discussions that followed. The hypotheses were verified through a short literature review and tested through an intervention as described in section 4 and 5. But first, to better understand the decision context and stakes, we will now present the case.

3. Presentation of the Case of Brusdalsvatn, Ålesund Norway

The decision under analysis in the present case consists in adopting and implementing a proposed strategy to protect the catchment and drinking water reservoir of the Brusdalsvatn lake. The strategy addresses most of the pollution sources of the lake, both biological and chemical. As in other Scandinavian cases (e.g., Andersson et al. 2012), the decision-makers are the locally-elected politicians of the municipality, whereas the technical administration gathers the evidence and proposes the strategy, in this case, commissioned to a consultancy. The recommended strategy consists in:

- Diverting effluent water from surrounding houses away from the reservoir to limit biological contamination by parasites, bacteria and viruses
- Moving away the E39 motorway out of the catchment to limit contamination especially of heavy metals and microplastics
- Limiting activities such as forestry and farming to prevent runoff of pesticides, fertilisers and biological contaminants from husbandry and excessive soil runoff
- Limiting any other business activities in the catchment to limit other chemical contaminants such as brominated flame retardants (plastic producing company) and increased traffic
- Limiting access for recreation (e.g., dogwalking) and events, and ban swimming and ice skating in/on the lake to avoid biological contamination and increased traffic

For the purpose of this paper, the scope was narrowed down to the second point regarding pollution from the road, consisting mostly of heavy metals, as the DSS helps illustrate the dynamics behind that problem. However, it may be worth noting for the discussion part that the first measure was implemented at the cost of local inhabitants of the catchment, and that this produced a backlash for the politicians involving protests and outrage in the media. One of the reasons why this decision is particularly interesting for analysis is that the decision has been pending for three years, since the strategy was made official, and it has been reported at each meeting due to a failure to reach consensus on the course of action.

The proposed strategy is framed to answer two logics of decision-making, namely, a logic of consequences, and a logic of appropriateness. Firstly, using a cost-benefit approach, the strategy recommends the precautionary principle by protecting the catchment from activities, highlighting the higher costs of moving the water supply infrastructure to an alternative location, compared to the costs of implementing the recommended measures. From a logic of appropriateness or rule-following (March, 1994), the strategy document highlights the duties and laws binding municipalities. Indeed, Norwegian municipalities have, under the Water Framework Directive and the Drinking Water Directive, the responsibility for the protection of all water sources and drinking water sources under their jurisdiction. Moreover, under the Public Health Law, consideration needs to be paid to drinking water when designing the municipality's spatial plans. As a supplier of municipal services such as running water and health services, municipalities also have duties to ensure their quality and sustainability over time. As municipalities also have the legal duties of providing drinking water and well-functioning health services under the Public Health Law, it is part of their duties and in their own interest to ensure that the water is not contaminated. When it comes to the tools for enforcing this protection, these include fines for breaching the Pollution Act, the Plan and Building Act, the Food Act, the Soil Act, the Forestry Act and the Recreation Act (Lover - Lovdata).

So what can be the reason for a lack of decision on the implementation of this second point of the strategy if not a lack of understanding of the information?

As mentioned in the introduction, the time horizon applied to the decision (Hansson, 2005) may differ among decision-makers and be generally problematic in this case. A short-term horizon may favour inaction by saving costs and conflicts but it will in the long run lead to higher costs and conflicts and possible health effects among local populations. On the other hand, if one applies a long time horizon, the rational decision would be to follow the expert advice and the strategy for the catchment.

The main actors involved in this decision by responding to consultations and expressing their concerns either in open letters or to the press include:

- Food Safety Authority (Mattilsynet) has the responsibility for keeping the water quality and for the routines to be followed but they do not take any samples,
- Vannverk who take samples for monitoring purposes,
- Norwegian Road Authority (Vegvesen) who maintain the roads from snow and vegetation at local level but are also in charge of new road developments, such as the 4-lane plan within the valley,
- NTNU, the local university has collaboration with Ålesund municipality and has access to both the catchment and the lake to take samples for research, including the development of their simulation models. They are in possession of data, both published and unpublished on biological and chemical contamination,
- Senior municipal doctor (Overlege) oversees public health matters, such as the spread of diseases in the population and is therefore among the experts consulted on the drinking water reservoir decisions
- Public Health Institute (Folkehelseinstituttet) as advisors and researchers in matters of public health,
- County Governor has a legal status and represents the state. They are in theory able to punish the municipality for not respecting the law but rarely exercise that power. They tend to give expert opinion and warnings,
- Norwegian Water is an interest organization for water in Norway,
- Norwegian Road Authority have a very high status since the last government took over in 2008 and reformed the road system in Norway to provide more and safer access throughout the country. Some of these large projects are still ongoing such as the proposed enlargement of the E39 road within the catchment area. They also have responsibility for the maintenance of the road safety thereby clearing roads from snow, adding salt and removing excess vegetation,
- Local residents directly within the catchment and other users of the areas around the lake (leisure, schools, businesses etc.)
- Inhabitants of the 3 municipalities receiving and using water from the lake for drinking and other purposes.

4. Description of the DSS and Intervention

The Hydrodynamics and Water Quality Simulation and Visualization System (HWQSS, Figure 1) is a program for automation of simulation and visualization of hydrodynamics and water quality transport in drinking water reservoirs such as lakes and rivers. The system enables quick setup and running of different scenarios to investigate possible responses of a surface waterbody under changing contaminant levels and weather conditions. The simulation and visualization system in HWQSS relies on the Generalized Environmental

Modelling System for Surface waters (GEMSS) (http://gemss.com/gemss.html), which is an integrated system of 3D hydrodynamics and transport modules. GEMSS itself is composed of several modules for different application. However, for the purpose of simulating water quality in drinking water reservoirs, the following modules are used in HWQSS: Hydrodynamic and transport module – HDM; Water quality module – WQM; Generalized bacterial Module – GBM; Toxic module; and User definable module – UDM.

Using the GEMSS engine, the HWQSS tool automates various tasks required for model implementation. Such tasks include input preprocessing, generation of 3D grids for model execution, addition of boundary conditions (such as locations of inflows and outflows, raw water intake locations for treatment plants, associated discharge, and water quality), creation of model control file, running simulations, and swiftly generating necessary output figures for visualization. In addition, an important component of the HWQSS tool is that it makes it possible for users to edit model control files to test different scenarios of water quality and climate conditions. Presently, this tool has been developed mainly to support management of the drinking water source for the city of Ålesund in Norway. However, the tool can still be applied to similar water source reservoirs, provided necessary calibration data are available, as shown for instance in Lang et al. (2010).

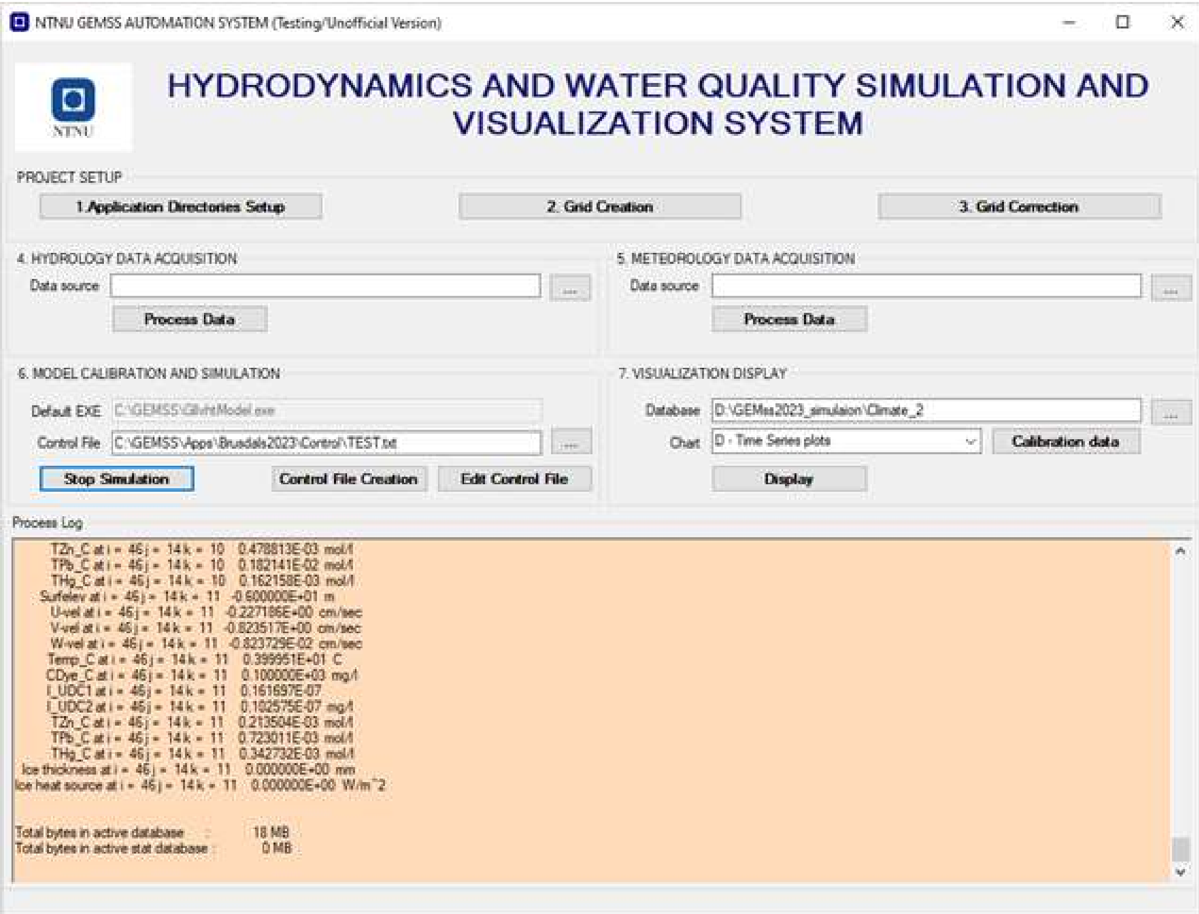


Figure 1. Screenshot of the HWQSS model for water quality scenario simulation of Brusdalsvatn, Ålesund

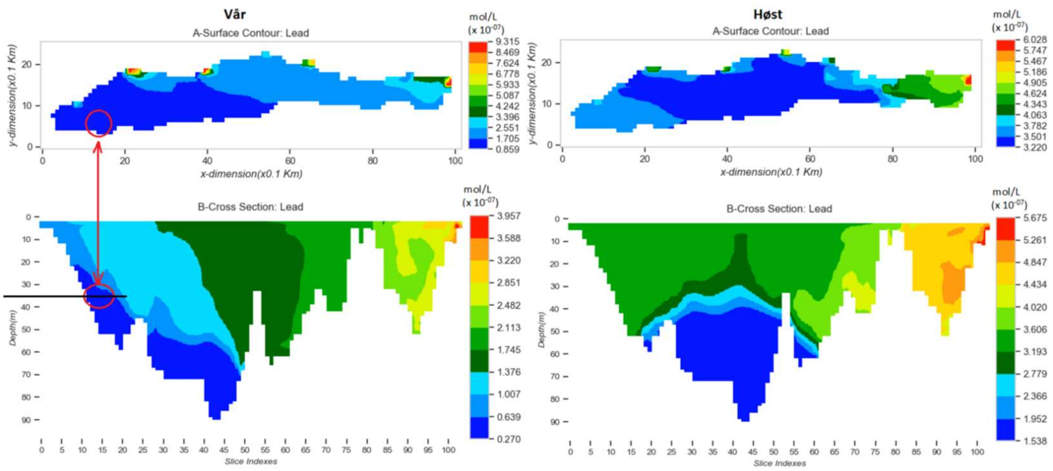


Figure 2. Simulation of the status quo scenario of Lead levels at different depths in the Brusdalsvatn lake for each of the 4 seasons

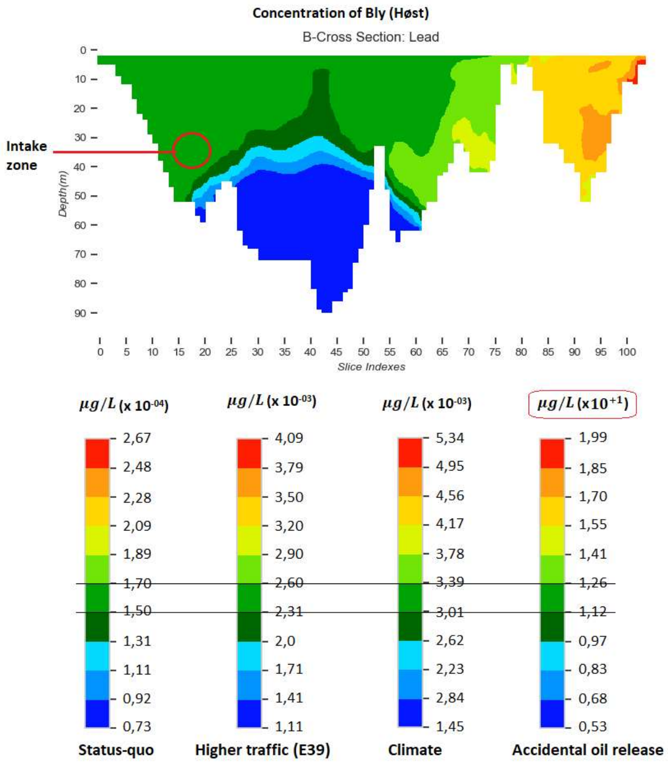


Figure 3. Simulation of all four scenarios for the riskiest season, Autumn, i.e., when turnover and mixing occur among the different thermoclines.

This decision-support tool was introduced in one of the local political meetings, where a decision had to be made about the implementation of the strategy, as information. Due to time constraints and the limits on the new elements of information that can be brought in to decision-making situations on the same day, the results from four simulated scenarios were shown in a presentation and explained with the help of visualisations as shown in figures 2 and 3. The scenarios in question comprised status quo, increased traffic on the road (as per the proposal to widen the road), extreme weather from climate change, and accidental oil release from a road accident. The alternative of protecting the catchment with consequent decreases in heavy metal influx was not included as it was deemed well-enough supported by the strategy document. In addition, the scenarios described were complemented by estimates of health effects on local inhabitants based on estimates provided by the Norwegian Health Authority.

The questions and discussion that followed were noted down and analysed to verify the three hypotheses formulated in section 2. To study whether the form of risk communication or temporality were hampering the consideration of risk information in the decision. The DSS addressed needs of transparency and clarity in predictions (including probability and uncertainty), as well as future scenario simulation, to enable the decision-makers to understand the risks and make informed decisions that truly reflect their preferences (Wong-Parodi et al., 2020; Sheppard et al., 2011).

5. Results and Discussion

The meeting was attended by the "Technical Committee for the Environment and Transport", which is made up of a selection of the same politicians that constitute the Municipality Board. After the tool was presented and the results from the four scenarios relating to the alternatives available to the decision-makers, these had the opportunity to ask questions and discuss.

The questions by the politicians were detail-oriented around the worst-case scenario only, and possible yet unlikely mitigation measures. The questions came from three of the politicians and all regarded the effects of the worst-case scenario. There appeared to be incredulity at first, as the question was posed of whether these results could truly reflect the impact of one single truck discharging into the lake. After the model was explained one more time, more questions were asked such as the effects on the model of replacing the vehicle used in the accidental oil release scenario by an electric or hydrogen vehicle. Another question addressed the effects of the addition of gutters between the road and the lake, to attempt to deviate some of the oil away from the water. Another question regarded the move of the water intake from 35m depth to 80m depth and the effects of such a measure on health risks. The results presented through the DSS were questioned again individually after the presentation was finished, and uncertainty around the numbers were raised by individual politicians and discussed with the researchers.

In summary, there was a feeling of denial or a refusal to take for granted that the results from the model were solid and the communication based on genuine concern. Furthermore, the responses showed a complete disregard of the information on the other cases reflecting very real alternatives (e.g., increased traffic through the extension of the road), overestimation of the effects of technical mitigation measures based on technological improvements, and generally no improvements in their understanding of information regarding risks.

Therefore, going back to the research questions, the researcher concludes that the reason for not including information about health and ecological risk in their decision-making is mixed. Indeed, it stems from a lack of understanding of the information, both when in text or visually displayed. Moreover, distrust is obvious in the questioning of the reliability of the model and proposing alternative, unrealistic mitigation measures, as opposed to the ones advised by the scientific experts is observed. This could relate to a lack of acceptance of the experts as having a higher position, based on their knowledge and experience. This could be explained by the tradition for egalitarianism in the country and "janteloven", according to which, no one is better than another. However, it is also very likely that the reluctance to carry out the decision to its full implementation due to the heavy costs, high level of complexity and negotiation required, unpopularity among the locals due to habit, trigger all the mechanisms observed. It is therefore a reverse process we observe, where the alternatives are known, and the reluctance to adopt them affect the way in which information is gathered (requesting more and more information, when all the necessary evidence is already on the table), and processed.

A similar tool has not been tested in a political context with decision-makers before, in spite of its potential in decision support. The closest study we have found to ours is that by and Bitterman and Webster (2024), collaborative policy modelling for Chesapeak Bay. Interestingly, they found similar results in that the expert system was regarded with distrust, and past experience of the decision-maker was considered more important than the model results.

6. Limitations and Implications and Future Research

Through this case study, we hope to stimulate more research around the field of political decision-making and the experimentation with DSS for socio-environmental issues. We hope to have highlighted how interesting it can be to intervene in real-world political processes and observe decision-making based after the intervention of a DSS. One of the limitations is that the tool could not be deployed as a full DSS mapping preferences and giving optimal choices. Instead, it was used as an expert information system. Another limitation is that the presentation of the results could not directly influence the decision made on that day. It would be interesting to see if such a tool could sway a decision for the better in a real decision case. Finally, there was very little insight into the thinking of the politicians. It would be interesting for a further study to engage with them individually and record their thought as they discover the tool and its results.

References

Andersson, A., Grönlund, Å., & Åström, J. (2012). “You can't make this a science!”—Analyzing decision support systems in political contexts. Government Information Quarterly, 29(4), 543-552.

Bana e Costa, C. A., Lourenço, J. C., Oliveira, M. D., & Bana e Costa, J. C. (2014). A socio-technical approach for group decision support in public strategic planning: the Pernambuco PPA case. Group Decision and Negotiation, 23(1), 5-29.

Bitterman, P., & Webster, D. G. (2024). The Collaborative Policy Modeling Paradox: Perceptions of water quality modeling in the Chesapeake Bay Watershed. Socio-Environmental Systems Modelling, 6, 18677-18677.

Burch, S., Shaw, A., Sheppard, S., & Flanders, D. (2009). Climate change visualization: using 3D imagery of local places to build capacity and inform policy. State of Climate Visualization, CSPR Report, 4, 65-73.

Eisenhardt, K. M. (1989). Building theories from case study research. Academy of management review, 14(4), 532-550.

Hansson, S. O. (2005) Decision Theory: A brief introduction. Stockholm: KTH Royal Institute of Technology.

Lang, U., Schick, R., & Schröder, G. (2010). The decision support system bodensee online for hydrodynamics and water quality in lake constance. Decision support systems, advances in, 81-98.

Lindblom, C.E. (1959). The science of “muddling through”. Public Administration Review. 19(2): 79–88.

March, J. (1994). A Primer on Decision Making: How Decisions Happen. Maxwell Macmillan

Richter, I., Gabe-Thomas, E., Queirós, A. M., Sheppard, S. R., & Pahl, S. (2023). Advancing the potential impact of future scenarios by integrating psychological principles. Environmental Science & Policy, 140, 68-79.

Sheppard, S. R., Shaw, A., Flanders, D., Burch, S., Wiek, A., Carmichael, J., ... & Cohen, S. (2011). Future visioning of local climate change: a framework for community engagement and planning with scenarios and visualisation. Futures, 43(4), 400-412.

Simon, H. A. (1960). The new science of management decision.

Spetzler, C., Winter, H., & Meyer, J. (2016). Decision quality: Value creation from better business decisions. John Wiley & Sons.

Wong-Parodi, G., Mach, K. J., Jagannathan, K., & Sjostrom, K. D. (2020). Insights for developing effective decision support tools for environmental sustainability. Current Opinion in Environmental Sustainability, 42, 52-59.

Zhao, J., & Luo, Y. (2021). A framework to address cognitive biases of climate change. Neuron, 109(22), 3548-3551.

A text-mining approach to understand the barriers and requirements for truck platooning deployment

Rhaydrick Sandokhan P. T. Tavares[1], Sérgio Pedro Duarte[1,2], Vera Miguéis[1,3] and António Lobo[1,2*]

[1] Faculdade de Engenharia da Universidade do Porto, Porto, Portugal
[2] CITTA – Centro de Investigação do Território, Transportes e Ambiente, Porto, Portugal
[3] INESC TEC – Institute for Systems and Computer Engineering, Technology and Science
[*] lobo@fe.up.pt

Abstract

The platooning technology facilitates the seamless coordination of two or more trucks as a connected convoy, using connectivity technology and automated driving support systems. While offering social-economic and environmental advantages, the transition to practical platooning systems necessitates advancements in infrastructure, in-vehicle technology, traffic legislation, and workforce capabilities. To comprehensively gauge the perspectives of key stakeholders influenced by this transformative shift, interviews were conducted with road regulators, truck drivers, and road operators. Leveraging Natural Language Processing (NLP) models, the study aims to analyse interview responses to ascertain if these diverse actors share similar views on critical aspects of truck platooning. Nevertheless, by employing NLP, the research seeks to uncover underlying patterns and commonalities in the opinions of road regulators, truck drivers, and road operators and the existing literature, shedding light on potential challenges and opportunities in the widespread adoption of platooning systems.

Keywords: truck platooning; participatory methods; natural language processing; systems approach

1. Introduction

Sustainability and greener solutions are big topics that are tackled every time we talk about innovative technology solutions to problems that we face as a society. Also, new technology is faced with the challenges of “dualism,” that is, functioning efficiently today while planning and innovating effectively for tomorrow (Katz & Allen, 1984). The main purpose of innovative technologies is to bring innovation and increase economic benefits while decreasing social-environmental impacts.

In recent years, the automotive industry has been trying to be at the vanguard of this process by developing hybrid and fully electric vehicles and automated systems for freight and passenger transport. Consequently, it is important to highlight that road transport is responsible for half of the freight transport (ERTRAC, 2019), so in that manner, it is vital that a viable solution is found, to answer the goals defined by the European Commission regarding safety and environmental concerns (European Commission, 2019; ERTRAC, 2022). Truck platooning is one of the proposed solutions since, at its core, it takes advantage of state-of-the-art communication and vehicle automation technologies while promising to reduce energy consumption and emissions.

The platooning systems enable two or more trucks to operate as a convoy, resembling a train. However, this is achieved through a virtual connection utilizing connectivity and automated driving support technology. The trucks maintain a distance from each other (PPMC, 2021), and the initial truck is designated as the leader, taking on the responsibility of the driving task with support from Cooperative Adaptive Cruise Control (CACC) and system monitoring (Willemsen et al., 2022). The subsequent trucks are referred to as followers, adjusting and responding to the leader's actions, with the potential for little to no human interaction (Castritius et al., 2020).

Although truck platooning can transform the operations of logistics companies (Ferrell et al., 2020), their workforce may perceive a potential loss of jobs and opportunities as well as could resist and see more risks than benefits regarding the safety of vehicle automation (Cunha et al., 2022). Therefore, it is important to understand users' requirements and acceptance, which is the key to ensure a successful deployment in the real world. Some tests in controlled environments have already been conducted, for instance in Germany (Castritius et al., 2020), but a full-scale implementation has yet to be carried out.

Developed under the project TRAIN, this research aims to extract knowledge from the transcribed speeches of the actors involved in the deployment of truck platooning, as well as from a literature review previously conducted under the same project. The main goal of the project is to map the risks and requirements from different actors in order to facilitate the design and deployment of high acceptance technology. The fact that all interested parties are involved in the design of the technology increases the potential for later adoption and reduces the chances for conflicts. Particularly, this work proposes to identify the main concerns from the actors, their sentiments and/or arguments towards the benefits and risks of the technology, and the relationships between actors. Moreover, the research should allow to contrast the insights from the transcribed speeches with those from the documents of reviewed international literature.

In the end, a comprehensive mapping of the main drivers and barriers to truck platooning technology acceptance and adoption at a large scale should obtained from the different groups. The remainder of this paper describes the data available, the methodology and how it is connected to the overall research project, and finalizes with some closing remarks, presenting the next steps of this work.

2. Data collection

This work is built on available data gathered in previous stages of the project. To gain a comprehensive understanding of the potential impact of platooning systems on various actors within the Portuguese context, a total 11 focus groups (FG) and semi-structured interviews were conducted with key stakeholders. The sessions involved drivers (4 FG), logistics operators (2 FG), road operators (2 FG and 1 interview), transport authorities and regulators (2 interviews). For the in-person session, there was video and audio recording on-site, while the online sessions were recorded directly with the videoconference software. The views expressed by the participants are their own and do not necessarily reflect the opinions of their respective companies or organizations. The primary objectives of these interviews were to assess:

a. The level of knowledge of the truck platooning technology, namely:
    a. Investigating the extent to which each stakeholder group is familiar with truck platooning technology;
    b. Obtaining insights into their understanding of the underlying mechanisms and functionalities of platooning systems;
    c. Identifying any knowledge gaps or misconceptions regarding the technology.
b. The usefulness of truck platooning
    a. Evaluating the perceived usefulness and potential benefits of implementing platooning systems according to each stakeholder's perspective;
    b. Examining any concerns or reservations related to the adoption of this technology;
    c. Exploring the stakeholders' expectations regarding improvements in efficiency, safety, and overall transport operations;

c. The interest and intention to use/implement truck platooning in the Portuguese context.

   a. Gauging the stakeholders' level of interest in adopting platooning systems in the Portuguese and broader European context;

   b. Identifying any regulatory or operational barriers that may hinder or facilitate the implementation of platooning technology;

   c. Assessing the stakeholders' intention to actively participate in or support the integration of platooning systems into existing transport infrastructures.

The FG sessions were approximately 1h30 long and composed by 6 to 8 participants. The interviews were shorter as there was only one participant to intervene. The recordings were transcribed starting with the audio recording, and supported by the video recording when two persons spoke at the same time.

3. Methodology

As mentioned, this research is part of a project focused on understanding the risks and requirements for the deployment of truck platooning. The project comprises three main stages: a stakeholder survey through qualitative and quantitative methods; driving simulation experiences to study driver behaviour and assess drivers' ability to use the technology; and designing guidelines for truck platooning deployment. This work is developed in the context of the first part of the project, which comprised the already mentioned literature review, FG and interviews.

The data collected in the literature review unveiled the representations from three groups of actors: drivers, decision-makers, and general public. Moreover, a preliminary analysis of the FG and interviews highlighted the main subjects addressed by the different groups. For a deeper assessment of the perceived acceptance, the first stage is composed by two sequential tasks: a qualitative analysis of the FG transcriptions and a quantitative analysis of a questionnaire targeting professional truck drivers. This paper describes the results from the qualitative analysis task.

Natural language processing (NLP) began in the 1950s as the intersection of artificial intelligence and linguistics. NLP was originally distinct from text information retrieval (IR), which employs highly scalable statistics-based techniques to index and search large volumes of text efficiently; Manning et al. (1999) provides an excellent introduction to IR. With time, however, NLP and IR have converged. Currently, NLP borrows from several remarkably diverse fields, requiring today's NLP researchers and developers to broaden their mental knowledge significantly.

In this research, we will first use NLP in the context of automated truck platooning for the extraction of sentiment from textual data. By employing sentiment analysis, one can discern the emotions and opinions expressed by different stakeholders, such as truck drivers, fleet managers, and regulatory bodies. Then, we will use topic segmentation to categorize discussions and opinions into distinct themes. Finally, text summarization will allow us to wrap up the main topics of discussion and the prevailing sentiments, streamlining the process of technology acceptance analysis. We are experimenting with different setups with well-known libraries and, in some cases, taking advantage of the transformers model already trained for other purposes (e.g., finance and medicine).

The focus groups and interviews have been transcribed in a proper format to apply different NLP algorithms to extract the sentiment of the stakeholders, analyse topic segmentation, and estimate the distance between them. By combining qualitative insights from interviews with advanced NLP analyses, this methodology aims to provide a nuanced understanding of stakeholder perspectives, sentiments, and interactions regarding the implementation of truck platooning systems in the Portuguese context. Moreover, similar techniques will be applied to the corpus of the literature reviewed under the TRAIN project, allowing the detection of evolving attitudes and the similarities and differences between the Portuguese and the international perspectives.

4. Closing remarks

The introduction of novel technology in a specific context calls for a deep understanding of its repercussions, given that its relevance and successful deployment are built upon the adopter's and users' preferences and the capability of the developers to address them. The dilemma for decision-makers lies in striking a harmonious balance between the advantages and drawbacks stemming from the integration of said technology. The utilization of a systemic approach affords a comprehensive outlook of the complete system, facilitating the examination of the factors exerting influence and the interplay among them. Understanding the sentiment surrounding automated truck platooning is crucial for gauging acceptance levels and addressing potential barriers and concerns regarding the technology's utility and impacts, which can be different from one context to another.

Leveraging text mining to comprehend human perceptions and behaviour is experiencing exponential growth, thanks to advancements in machine learning and natural language processing. This allows to speed up the grasping of the diverse perspectives of stakeholders, capturing convergent themes, discerning sentiments surrounding specific topics, and swiftly interpreting whether the requirements, objectives, and perspectives align cohesively.

In the case of truck platooning, there is a complex ecosystem in which several actors pose different requirements and have different expectations, making it a case for the application of text mining techniques. This work contributes with a systemic view of the perceptions and acceptance from different actors and points to the high-level requirements for the deployment of the technology. Sentiment analysis is performed in order to contrast the arguments from different groups, offering decision-makers the possibility to develop a user-centric technology.

The research counts also with a questionnaire to enlarge the sample and develop quantitative models for truck drivers. Moreover, simulated driving tests will evaluate the drivers' ability to perform necessary tasks, while providing a moment of experience and increasing the trust in the technology. In conclusion, the participatory and systemic approach of this study can generate valuable knowledge to inform policy makers, industry stakeholders, and researchers about potential areas of focus for further development and implementation of truck platooning systems.

Acknowledgement

This research was funded in whole by the Fundação para a Ciência e a Tecnologia, I.P. (FCT, Funder ID = 50110000187), under the project with DOI 10.54499/PTDC/ECI-TRA/4672/2020 and the grant with DOI 10.54499/CEECINST/00010/2021/CP1770/CT0003.

References

Castritius, S. M., Hecht, H., Möller, J., Dietz, C. J., Schubert, P., Bernhard, C., Morvilius, S., Haas, C. T., & Hammer, S. (2020). Acceptance of truck platooning by professional drivers on German highways: A mixed methods approach. *Applied Ergonomics, 85*, 103042. https://doi.org/10.1016/J.APERGO.2019.103042

Cunha, L., et al. (2022). Who Really Wants Automated Vehicles? Determinant Factors of Acceptability Profiles in Portugal. In *Proceedings of the 5th International Conference on Intelligent Human Systems Integration (IHSI 2022)* (pp. 22-24, Venice, Italy).

ERTRAC. (2019). Long Distance Freight Transport: A Roadmap for System Integration of Road Transport (pp. 1–44).

ERTRAC. (2022). Connected, Cooperative and Automated Mobility Roadmap Status: Final for Publication.

European Commission. (2019). EU Road Safety Policy Framework 2021-2030 - Next Steps towards 'Vision Zero.'

Ferrell, W., Ellis, K., Kaminsky, P., & Rainwater, C. (2020). Horizontal collaboration: Opportunities for improved logistics planning. *International Journal of Production Research, 58*(14), 4267–4284.

Janssen, R., Zwijnenberg, H., Blankers, I., & Kruijff, J. de. (2015). Truck Platooning: Driving the Future of Transportation (Issue February). http://cvt-project.ir/Admin/Files/eventAttachments/Truck Platooning by DSRC Feb 2015_541.pdf

Manning, C. D., & Schutze, H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.

Katz, R., & Allen, T. J. (1984, June). Organizational Issues in the Introduction of New Technologies.

PPMC. (2021). Truck Platooning Trial in the Netherlands with EcoTwin. PPMC Transport. http://www.ppmc-transport.org/truck-platooning-trial-in-the-netherlands-with-ecotwin

Willemsen, D., Schmeitz, A., Mascalchi, E., Huisman, R., Liga, V., Alieiev, R., Nordin, H., Xu, W., Luetzner, J., & Dhurjati, P. (2022). V2 Platooning Use Cases, Scenario Definition and Platooning Level (Issue 769115). https://platooningensemble.eu/library/deliverables

Quantum extensions of classical games and decision optimization

Marek Szopa [1*], Anna Gorczyca-Goraj [1,2] and Piotr Frąckiewicz [2]

[1] University of Economics, Katowice, Poland
[2] , Pomeranian University, Słupsk, Poland
[*] marek.szopa@uekat.pl

Abstract

The extension of classical games to the quantum domain, generated by the addition of unitary strategies to two classical strategies of each player is studied. The conditions that need to be met by unitary operations to ensure that the extended game is invariant with respect to the isomorphic transformations of the input game are determined. It has been shown that there are several types of these extensions, some of them being purely quantum and allowing for solutions that are unobtainable in classical games. It has been proven that the inclusion of a unitary operator in the extensions of two variants of a classical game, without adhering to the aforementioned conditions, can lead to the emergence of quantum games that are not equivalent, thereby exhibiting disparate Nash equilibria. Our utilization of the acquired results allows for the advancement of the classical Prisoner's Dilemma game into a quantum game, yielding a singular Nash equilibrium that is closer to Pareto-optimal solutions than the original game.

Keywords: Game theory; Quantum games; EWL scheme; Nash equilibria; Prisoner's dilemma

1. Introduction

The field of quantum game theory merges game theory and quantum computation theory to create a unique interdisciplinary research area. Through recent studies, researchers have begun to uncover the possibilities of utilizing quantum games to gain fresh perspectives on problem-solving tactics. This work focuses on expanding classical game strategies into the quantum domain (Eisert et al., 1999), specifically by adding a unitary strategy to the existing classical strategies of each player.

The objective of this research is to establish the crucial factors that enable the matrix of a quantum game extension to maintain its invariance when exposed to isomorphic transformations of the original classical game (Frąckiewicz, 2016). We demonstrate the existence of several clases of quantum games that preserve this invariance. By utilizing unitary operators, these novel tactics enable players to attain results in quantum games that were previously unattainable in traditional contexts. For example, the classical Prisoner's Dilemma frequently resulted in unsatisfactory outcomes for both players, but by utilizing quantum strategies, the equilibrium can be closer to a Pareto-optimal solution.

2. Key contributions

Extending classical games to quantum domain. The paper outlines how classical games can be extended to the quantum domain by incorporating unitary operations as player's strategies. This approach is important in exploring new dimensions of strategic interactions that are not possible in classical game theory.

Conditions for unitary operations. A major contribution of this work is identifying the necessary conditions for unitary operations, which play the role of player's strategies, to maintain the invariance of the extended game concerning isomorphic transformations of the input game. This ensures that the quantum extension remains consistent with the original classical game structure.

Classification of extensions. The study classifies the extensions into several types: three in case of the extension to quantum games combining three strategies and five in case of the games combining four unitary transformations. This classification helps in understanding the nature and impact of different quantum strategies on classical games.

Impact on game equilibria. It is shown that extending classical games with unitary operations that do not meet the identified conditions can result in quantum games with different properties, such as varied Nash equilibria, from their classical counterparts. This highlights the sensitivity of game dynamics to the introduction of quantum strategies.

Application to the Prisoner's dilemma (PD). A practical application of the study is demonstrated through the classical PD game. The paper extends this game to the quantum domain and shows that the quantum version has a unique Nash equilibrium closer to Pareto-optimal solutions than the classical version. The typical matrix representation of this game can be expressed as a bimatrix

$$\begin{pmatrix} (3,3) & (0,5) \\ (5,0) & (1,1) \end{pmatrix}.$$

In the classical form PD has a single Nash equilibrium, corresponding to the payoffs $(1,1)$. The quantum PD in one of the extension classes is of the form:

$$\begin{pmatrix} (3,3) & (0,5) & \left(\frac{5-t}{2},\frac{5-t}{2}\right) & \left(\frac{t+4}{2},\frac{t+4}{2}\right) \\ (5,0) & (1,1) & \left(\frac{t+4}{2},\frac{t+4}{2}\right) & \left(\frac{5-t}{2},\frac{5-t}{2}\right) \\ \left(\frac{5-t}{2},\frac{5-t}{2}\right) & \left(\frac{t+4}{2},\frac{t+4}{2}\right) & \left(-t^2-t+3,-t^2-t+3\right) & \left(t(t+4),t^2-6t+5\right) \\ \left(\frac{t+4}{2},\frac{t+4}{2}\right) & \left(\frac{5-t}{2},\frac{5-t}{2}\right) & \left(t^2-6t+5,t(t+4)\right) & \left(-t^2+3t+1,-t^2+3t+1\right) \end{pmatrix}.$$

Where $t\epsilon(0,1)$. If e.g. $t=3/4$ then the extended game has three Nash equilibria corresponding to payoffs $\left(\frac{19}{8},\frac{19}{8}\right)$ and $\left(\frac{23}{12},\frac{23}{12}\right)$, i.e. much closer to Pareto optimal results then the classical $(1,1)$. This implies that quantum strategies can potentially lead to more cooperative and beneficial outcomes in game-theoretic scenarios, especially when applied in negotiations.

Interdisciplinary nature. The research underscores the interdisciplinary nature of quantum game theory, which merges concepts from game theory and quantum computation. This fusion allows for the exploration of novel problem-solving strategies and outcomes.

Further research implications. The paper opens avenues for further research, particularly in exploring quantum correlations, unitary strategies in multiplayer games, and the classification of Nash equilibria in different quantum game extensions.

3. Conclusions

In summary, the paper makes a substantial contribution to the field of quantum game theory by extending classical game strategies to the quantum realm, identifying key conditions for unitary operations, and demonstrating the practical application of these concepts through the example of the Prisoner's Dilemma. The insights gained from this research have implications for the development of new strategies in the quantum domain and enhance the understanding of quantum game dynamics.

References

Eisert, J., Wilkens, M., & Lewenstein, M. (1999). Quantum Games and Quantum Strategies. *Physical Review Letters*, *83*(15), 3077–3080. https://doi.org/10.1103/PhysRevLett.83.3077

Frąckiewicz, P. (2016, August 14). *Strong Isomorphism in Eisert-Wilkens-Lewenstein Type Quantum Games* [Research Article]. Advances in Mathematical Physics; Hindawi. https://doi.org/10.1155/2016/4180864

Frąckiewicz, P., Szopa, M. (2024). Permissible extensions of classical to quantum games combining three strategies. *Quantum Inf Process* 23, 75 (2024). https://doi.org/10.1007/s11128-024-04283-3

Frąckiewicz, P., Gorczyca-Goraj A., Szopa, M. (2024). Permissible four-strategy quantum extensions of classical games. In preparation.

Group Decision and Negotiation
- Conflict resolution -

Conference papers and extended abstracts

24th International Conference on Group Decision and Negotiation &
10th International Conference on Decision Support System Technology
Human-Centric Decision and Negotiation Support for Societal Transitions

Uncertain Preferences to deal with the Framing Effect in the Graph Model for Conflict Resolution

Deoclécio P Castro [1,2*], Leandro Chaves Rêgo [2,3]
[1] PhD Program in Economics, Faculdade de Economia da Universidade do Porto, Porto, Portugal.
[2] Graduate Program in Modeling and Quantitative Methods, Universidade Federal do Ceará, Fortaleza/CE, Brazil.
[3] Statistics and Applied Math Department, Universidade Federal do Ceará, Fortaleza/CE, Brazil
[4] Statistics and Management Engineering Graduate Programs, Universidade Federal de Pernambuco, Recife/PE, Brazil
[*] deoclecio@gmail.com

Abstract

In this article our objective is to present an approach that allows us to carry out an analysis using the GMCR in conflicts where the decision makers' preferences present inconsistencies resulting from the Framing Effect. More specifically, we consider a GMCR with two decision makers who presented different preferences when, during the elicitation process, they were confronted with the same problem framed in two different ways. This behavior is inconsistent with Decision Theory and can be justified by the limitations of human rationality. Our approach uses uncertain preferences to deal with inconsistencies arising from the Framing Effect.

Keywords: behavioral decision making; cognitive bias; game theory; conflict analysis; graph model

1. Introduction

When faced with a conflict problem, a person naturally simplifies it to a level that can be more easily understood, allowing the decision makers (DMs) to anticipate the consequences of their different actions (Xu et al., 2018). Thus, a conflict model must be designed to reflect the way human beings perceive and react to reality. The Graph Model for Conflict Resolution (GMCR) is a mathematical model that uses concepts of Game Theory and graph theory to model and analyze strategic conflicts.

The main components to characterize a conflict are the DMs; the set of available actions for each DM throughout the conflict; the states where the conflict may be in and which are determined by a given combination of action choices by the DMs; and the preferences of each DM about the possible conflict's states (Fang et al., 1993).

Stability analysis is the systematic process of investigating the possible movements and reactions of DMs along the possible conflict evolution and calculating the most likely resolution to occur. This analysis depends primarily on the preferences of DMs about the possible scenarios that the conflict may take, as they are the ones that govern the decisions that the DMs may make. However, the decision-making process is conditioned by the limitations of human rationality (Simon, 1955). The Framing Effect was described by Tversky and Kahneman in 1981 (Tversky et al., 1981) as the change of preferences generated by the change of perspective of the same problem. This change makes preferences inconsistent and violates principles of rational decision-making.

In this work, we use uncertain preferences to deal with the inconsistencies that emerge from preference elicitation process through the Option Prioritization Method that presents a same problem in two different ways, that is, two different frames, for two DMs.

2. Background

2.1. GMCR

GMCR describes a conflict with a set of DMs $N=\{1, 2..., n\}$, $n = |N| \geq 2$; a set of possible states (scenarios) $S=\{s_1, s_2..., s_m\}$, which can arise according to the actions taken by the DMs involved in the conflict. For each DM, a directed graph models how they can change from a conflict state to another, where the nodes represent the states and the arcs represent the moves that DMs can make. Finally, each DM has some preferences about the set of states that can be expressed by a binary and asymmetric relation over S. In the stability analysis of this model, many stability concepts have been proposed to accommodate different behaviors that DMs may have in a conflict (Xu et al., 2018), (Fang et al., 1993). Among the most recent contributions, Hipel et al. (2020) refers to an approach that takes into account factors that influence the behavior of decision-makers as a future and desired advance for this field.

Preference Elicitation by Option Prioritization

According to Peng et al. (1997), a conflict can be represented in the form of options. Let $O_i = \{o_1^i, o_2^i, \dots, o_m^i\}$ be the set of options of DM $i \in N$, with $O_i \neq \emptyset$ and $O_i \cap O_j = \emptyset$, for $i \neq j$. Then, the set of all options in a conflict model is given by: $O = \bigcup_{i=1}^{n} O_i$. A state can therefore be defined in terms of options as being a combination of all the options in O that are, or are not, taken by the DMs in the conflict.

The Option Prioritization technique consists of specifying preferences by asking DMs to provide an ordered sequence of preference statements (from most to least preferred), which are Boolean combinations of the conflict's options. If a preference statement, Ω, is satisfied at state s, we denote by $\Omega(s)=True$, otherwise, we say that $\Omega(s)=False$.

A strict preference relationship based on an ordered sequence of preference statements is defined by Peng et al. (1997):

Definition 1: Let $s, s_1 \in S$ and $C_i = (\psi_{i_1}, \psi_{i_2}, \dots, \psi_{i_w})$ be an ordered sequence of all preference statements reported by the DM $i \in N$. State $s \in S$ is strictly preferred to $s_1 \in S$ for DM i if, and only if, there exists a positive integer $t \leq w$ such that $\psi_{i_1}(s) = \psi_{i_1}(s_1)$, $\psi_{i_2}(s) = \psi_{i_2}(s_1)$, ..., $\psi_{i_{t-1}}(s) = \psi_{i_{t-1}}(s_1)$, $\psi_{i_t}(s) = True$ and $\psi_{i_t}(s_1) = False$.

Uncertain Preferences

The uncertainty relation U was introduced by Li et al. (2002a), Li et al. (2004) and Li et al.(2002b) to express an uncertain relationship of preferences between two states of a conflict for a given DM. It can occur when information about DMs' preferences is incomplete, or the conflict is still ongoing, which limits the preferences to be expressed only with components of certainty involved. In our case, we use the uncertainty relationship to deal with inconsistent preferences that arise from the Framing Effect in the elicitation process.

2.2. Framing Effect

The study on the modeling of human behavior towards choices is a broad and challenging interdisciplinary area, whose pioneering works have been made by cognitive psychologists and behavioral economists (Tversky et al., 1981), (Güth, 2000). Arrow (1951) and Tversky et al. (1981) states that, although "explanations and predictions of people's choices, in everyday life and also in theory, are generally found in the assumption of human rationality", the theories based on rationality alone cannot do so because of “human imperfect perception and decisions”. These imperfections in human cognition are known as “cognitive biases”.

As an example, take the classic case of Asian Disease (Tversky et al., 1981). The answers are inconsistent and contradictory, since the expected returns are the same in all treatments. In the literature, this sort of phenomena is called Framing Effect. In the context of the GMCR, the inconsistencies that arise interferes in the conflict preference structure and, consequently, in the possible equilibria.

3. Demonstrative Case

To demonstrate the problem in question, let us consider a conflict situation with two DMs, *i and* j, with two and one options, respectively. There are 8 possible states, all feasible, in the conflict that can be presented through Table 1.

Table 1 – Feasible states.

		States							
DM	**Option**	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8
I	o_1	N	S	N	S	N	S	N	S
	o_2	N	N	S	S	N	N	S	S
J	o_3	N	N	N	N	S	S	S	S

We carry out the preference elicitation process in two stages with the intention of making the Framing Effect emerge. In Frame A, each of the DMs *i* and *j* was asked: "What would you most prefer to happen?". In Frame B the question was: "What would you most prefer NOT to happen?"

Consider that in response to these questions, the statements seen in Tables 2 and 3 below follow, where two different frames, $A, B \in \mathcal{E}$, are added in superscript to the notation of **Definition 1** to appoint under what frame the preference statement was made. These statements are given in terms of the Boolean operators: "&" operator AND, " | " operator OR, "if" operator IF, " ⇔ " operator IF AND ONLY IF, " - " operator NOT.

Table 2 – Preference statements under Frame A.

Statements	**DM *i***	**Statements**	**DM *j***
$\psi_{i_1}^{A}$	o_1 if o_3	$\psi_{j_1}^{A}$	o_1 & - o_2
$\psi_{i_2}^{A}$	- (o_1 | o_2)	$\psi_{j_2}^{A}$	o_2 ⇔ -o_3
$\psi_{i_3}^{A}$	- o_2 & o_3	$\psi_{j_3}^{A}$	- (o_1 & o_3)

Table 3 - Preference statements under Frame B.

Statements	**DM *i***	**Statements**	**DM *j***
$\psi_{i_1}^{B}$	- (o_1 | o_2)	$\psi_{j_1}^{B}$	- (o_1 & o_3)
$\psi_{i_2}^{B}$	- o_2 & o_3	$\psi_{j_2}^{B}$	o_2 ⇔ -o_3
$\psi_{i_3}^{B}$	o_1 if o_3	$\psi_{j_3}^{B}$	o_1 & - o_2

Note that under any frame, if state s_k first satisfies a statement than state s_w, then state s_w will satisfy the negation of this statement first than state s_k. Thus, if state s_k first satisfies a statement than state s_w under Frame B, then s_w will be preferable to s_k. In other words, s_w will be preferable to s_k if it satisfies the negation of a statement first then state s_k under Frame B. Thus, taking the negation of the statements under Frame B, one can use the traditional Option Prioritizing technique.

Table 4 presents the negation of the preference statements on options under Frame B and this will be the order of preference statements that we should use to obtain preferences using the GMCR+, for example.

Table 4 – Negation of preference statements under Frame B.

Statements	**DM *i***	**Statements**	**DM *j***
$\psi_{i_1}^{-B}$	$o_1 \mid o_2$	$\psi_{j_1}^{-B}$	o_1 & o_3
$\psi_{i_2}^{-B}$	$o_2 \mid - o_3$	$\psi_{j_2}^{-B}$	$o_2 \Leftrightarrow o_3$
$\psi_{i_3}^{-B}$	$- o_1$ & o_3	$\psi_{j_3}^{-B}$	$- o_1 \mid o_2$

If the DMs were completely rational, attending to the axiom of invariance of the description, the ordering of Tables 1 and 4 should be the same. The results differ from what is expected to be obtained from rational agents, since they should keep their preferences consistent when presented with the same problem. We can affirm, therefore, that inconsistencies in the preferences declared by the DMs arose as a result of the Framing Effect.

To exemplify how to obtain the preference relation under both frames. Consider states s_1 and s_6 for DM i. Both of them satisfy the first statement of DM i in Frame A: o_1 if o_3. However, as only s_1 satisfies the second statement of DM i in Frame A, - ($o_1 \mid o_2$), it follows that $s_1 \succ_i s_6$ under Frame A. To determine the preference relation under Frame B between these states, one must use the negation of preference statements as shown in Table 4. In that case, as only s_6 satisfies the negation of the first statement of DM i in Frame B, $o_1 \mid o_2$, it follows that $s_6 \succ_i s_1$ under Frame B. The preference relationships under both frames for all states are as follows:

For DM *i*

Frame A: $s_1 \succ_i s_6 \succ_i s_2 \sim_i s_3 \sim_i s_4 \sim_i s_8 \succ_i s_5 \succ_i s_7$

Frame B: $s_7 \succ_i s_2 \sim_i s_3 \sim_i s_4 \sim_i s_8 \succ_i s_6 \succ_i s_1 \succ_i s_5$

For DM *j*

Frame A: $s_6 \succ_j s_2 \succ_j s_3 \sim_j s_4 \sim_j s_5 \succ_j s_1 \sim_j s_7 \succ_j s_8$

Frame B: $s_8 \succ_j s_6 \succ_j s_1 \sim_j s_7 \succ_j s_2 \succ_j s_3 \sim_j s_4 \sim_j s_5$

Our contribution is made to identify these inaccuracies and present a method that can address similar problems allowing the Stability Analysis of the conflict to be carried out. To do that, we propose the following approach to obtain the complete representation of DMs' preferences.

4. GMCR with Framing Effect

The matrix mathematical approach has proven to be the most suitable for dealing with the GMCR with multiple states and DMs.

Notation: Let $m = |S|$ the number of feasible states, $i \in N$ and $s, q \in S$. For a graph model G, the positive strict preference, the negative strict preference, indifference, and uncertainty matrices for the DM $i,$ under

Frame $F \in \mathcal{E}$ are, respectively, the *m x m* matrices, $P_{i,F}^{+}$, $P_{i,F}^{-}$, $P_{i,F}^{=}$ and $P_{i,F}^{U}$, with inputs *(s,q)*, such that:

$$P_{i,F}^{+}(s,q) = \begin{cases} 1 & if\ q \succ_i s, \\ 0 & \text{otherwise,} \end{cases}$$

$$P_{i,F}^{-}(s,q) = \begin{cases} 1 & if\ s \succ_i q, \\ 0 & \text{otherwise,} \end{cases}$$

$$P_{i,F}^{=}(s,q) = \begin{cases} 1 & if\ q \sim_i s, \\ 0 & \text{otherwise,} \end{cases}$$

$$P_{i,F}^{U}(s,q) = \begin{cases} 1 & if\ q\ U_i\ s, \\ 0 & \text{otherwise.} \end{cases}$$

If $P_{i,F}^{+}(s,q) = 1$, then DM i prefers state q to state s under Frame F. On the other hand, if $P_{i,F}^{+}(s,q) = 0$, then DM i either prefers s to q, or is indifferent between s and q, or the preference between states s and q is uncertain (by the property of the completeness of preferences). The matrix operations $\circ$ and + correspond to the Hadamard Product and the Matrix Sum, respectively.

Definition 2 For a graph model G, the positive strict preference matrix for DM i, $i \in N$, obtained through two different frames, $A, B \in \mathcal{E}$, is a matrix *m x m*, P_i^{+} such that:

$$P_i^{+} = \left[\left(P_{i,A}^{+} \circ P_{i,B}^{+}\right) + \left(P_{i,A}^{=} \circ P_{i,B}^{+}\right) + \left(P_{i,A}^{+} \circ P_{i,B}^{=}\right)\right]$$

Definition 3 For a graph model G, the negative strict preference matrix for DM i, $i \in N$, obtained through two different frames, $A, B \in \mathcal{E}$, is a matrix *m x m*, P_i^{-} such that:

$$P_i^{-} = \left[\left(P_{i,A}^{-} \circ P_{i,B}^{-}\right) + \left(P_{i,A}^{=} \circ P_{i,B}^{-}\right) + \left(P_{i,A}^{-} \circ P_{i,B}^{=}\right)\right]$$

Definition 4 For a graph model G, the indifference matrix for DM i, $i \in N$, obtained through two different frames, $A, B \in \mathcal{E}$, is a matrix *m x m*, $P_i^{=}$ such that:

$$P_i^{=} = \left(P_{i,A}^{=} \circ P_{i,B}^{=}\right)$$

Definition 5 For a graph model G, the matrix of uncertain preferences for DM i, $i \in N$, obtained through two different frames, $A, B \in \mathcal{E}$, is a matrix *m x m*, P_i^{U} such that:

$$P_i^{U} = \left[\left(P_{i,A}^{+} \circ P_{i,B}^{-}\right) + \left(P_{i,A}^{-} \circ P_{i,B}^{+}\right)\right]$$

Let E be a matrix *m x m* with all entries equal to 1. Considering GMCR with Uncertain Preferences as proposed in Li et al. (2002a), Li et al. (2004), Li et al. (2002b), and knowing that the matrices defined above configure a partition of preferences over all states of a conflict, it is possible to verify their consistency by verifying the following identity:

$$P_i^+ + P_i^- + P_i^= + P_i^U = E.$$

4.1. Applying to the Demonstrative Case

Now, we can obtain the preference matrices for the Demonstrative Case presented above, for DMi and DMj:

i) Positive strict preference matrices:

$$P_i^+ = \begin{bmatrix} 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 1&1&1&1&0&1&0&1\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0 \end{bmatrix} \qquad P_j^+ = \begin{bmatrix} 0&0&0&0&0&1&0&0\\ 0&0&0&0&0&1&0&0\\ 0&1&0&0&0&1&0&0\\ 0&1&0&0&0&1&0&0\\ 0&1&0&0&0&1&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&0&0 \end{bmatrix}$$

ii) Negative strict preference matrices:

$$P_i^- = \begin{bmatrix} 0&0&0&0&1&0&0&0\\ 0&0&0&0&1&0&0&0\\ 0&0&0&0&1&0&0&0\\ 0&0&0&0&1&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&1&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&1&0&0&0 \end{bmatrix} \qquad P_j^- = \begin{bmatrix} 0&0&0&0&0&0&0&0\\ 0&0&1&1&1&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 1&1&1&1&1&0&1&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0 \end{bmatrix}$$

iii) Indifference preference matrices:

$$P_i^= = \begin{bmatrix} 1&0&0&0&0&0&0&0\\ 0&1&1&1&0&0&0&1\\ 0&1&1&1&0&0&0&1\\ 0&1&1&1&0&0&0&1\\ 0&0&0&0&1&0&0&0\\ 0&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&1&0\\ 0&1&1&1&0&0&0&1 \end{bmatrix} \qquad P_j^= = \begin{bmatrix} 1&0&0&0&0&0&1&0\\ 0&1&0&0&0&0&0&0\\ 0&0&1&1&1&0&0&0\\ 0&0&1&1&1&0&0&0\\ 0&0&1&1&1&0&0&0\\ 0&0&0&0&0&1&0&0\\ 1&0&0&0&0&0&1&0\\ 0&0&0&0&0&0&0&1 \end{bmatrix}$$

iv) Uncertain preference matrices:

$$P_i^U = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix} \qquad P_j^U = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

In order to exemplify what we propose, consider state s_5 of the conflict as the *status quo*. For DM i under frame A, $s_5 \succ_i s_7$, so $P_{i,A}^{+}(s_7, s_5) = 1$. For DM i under frame B, $s_7 \succ_i s_5$, so $P_{i,B}^{-}(s_7, s_5) = 1$. Using **Definition 5** we can compute the observed in cell (5,7) of conflict's uncertain preference matrix:

$$P_i^U(s_7, s_5) = [(1 \circ 1) + (0 \circ 0)] = 1$$

As a second example, consider state s_2 of the conflict as the *status quo*. For DM j under frame A, $s_2 \succ_j s_3$. For DM j under frame B, $s_2 \succ_j s_3$. By **Definition 2** we can compute the strict preference matrix for DM j, P_j^{+} above, where we observe 1 in cell (2,3), i.e., the state s_2 is strictly preferred to s_3 for DM j in conflict considering both frames.

Similar analyzes are valid for all states, DMs and frames.

Knowing the preference matrices, especially the uncertainty matrices, P_i^U, one is able to perform the stability analysis using the GMCR with Uncertain Preferences. It is important to highlight that different equilibria may be obtained when compared to the case in which the conflict is analyzed by considering each frame separately, but our method takes into account the Framing Effect in the elicitation process.

5. Final Considerations

In this work we propose an approach that deals with inconsistent preferences that may arise from failures in the perception of the rational decision process resulting from the Framing Effect.

During the modeling process, different steps can be impacted by cognitive biases, since human judgment is present in many of them, such as the generation of alternatives, the description of scenarios or the method chosen to elicit preferences. We proposed a method to obtain the preferences of two decision-makers who, in the elicitation process, are faced with a problem presented in two different ways, frames. As a result, we obtained matrix representations of decision makers' preferences, which allows us to analyze conflict stability that considers behavioral aspects of the decision-making process.

As a development of the present work, we intend to generalize the analysis for conflicts with n-DMs and n-Frames, as well as to present a set of interrelationships between the solution concepts under different frames and the solution concepts for conflicts with uncertain preferences.

Acknowledgments.

This work was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), financing code 001, Conselho Nacional de Desenvolvimento Científico e Tecnológico do Brasil (CNPQ) 406697/2023-0 and 308980/2021-2.

This research has been financed by Portuguese public funds through FCT - Fundação para a Ciência e a Tecnologia, I.P., in the framework of the project with reference UIDB/04105/2020.

References

1. Arrow, Kenneth J.: "Alternative Approaches to the Theory of Choice in Risk -Taking Situations," Econometrica 19, 404–437, 1951.
2. Fang, L., Hipel K.W., Kilgour, D.M.: Interactive Decision Making: The Graph Model for Conflict Resolution. Wiley. New York, 1993.
3. Güth, W.: Boundedly rational decision emergence: a general perspective and some selective illustrations. J. Econ. Psychol. 21(4):433–458, 2000.
4. Hipel, K.W., Fang, L. & Kilgour, D.M.: The Graph Model for Conflict Resolution: Reflections on Three Decades of Development. Group Decis Negot 29, 11–60, 2020.
5. LI, K. W.; HIPEL, K. W.; KILGOUR, D. M.; FANG, L. 2 x 2 games with uncertain ordinal preferences. *In*: THE 2002 International Conference on Control and Automation: final program and book of abstracts: Xiamen, Fujian Province, China, June 16-19, 2002. [Los Alamitos, Calif.?]: IEEE; Xiamen: Xiamen University, 2002a. p. 1471–1475.
6. LI, K. W; HIPEL, K. W.; KILGOUR, D. M.; FANG, L. Preference uncertainty in the graph model for conflict resolution. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, United States, v. 34, n. 4, July, 2004.
7. LI, K. W.; HIPEL, K. W.; KILGOUR, D. M.; FANG, L. Stability definitions for 2-player conflict models with uncertain preferences. *In*: IEEE INTERNATIONAL CONFERENCE ON SYSTEMS, MAN AND CYBERNETICS, 2002, Hammamet, Tunisia. Proceedings [...]. [United States]: IEEE, 2002b. v. 7, p. 13–18.
8. PENG, X.; HIPEL, K. W.; KILGORU, D. M.; FANG, L. Representing ordinal preferences in the decision support system GMCR II. *In*: IEEE INTERNATIONAL CONFERENCE ON SYSTEMS, MAN AND CYBERNETICS, 1997, Orlando, FL. Proceedings [...]. [United States]: IEEE, 1997. v. 1, p. 809–814.
9. Simon, H.A: A behavioral model of rational choice. Q J Econ 69(1):99–118, 1955.
10. Tversky, A., Kahneman, D.: The Framing of Decisions and the Psychology of Choice, Science 211, 453–458, 1981.
11. Xu H., Hipel K.W., Kilgour D.M., Fang L.: Conflict resolution using the graph model: strategic interactions in competition and cooperation. Springer, Cham, 2018.

Matrix Representation of Maximinh Stability in the Graph Model for Conflict Resolution for Bilateral Conflicts

Giannini I. A. Vieira [1], France E. G. de Oliveira [2] and Leandro C. Rêgo [3, 4*]

[1] Crateús Campus, Universidade Federal do Ceará, Crateús 63700-000, Brazil
[2] Graduate Program in Statistics, Universidade Federal de Pernambuco, Recife, PE, 50670-901, Brazil
[3] Statistics and Applied Math Department, Universidade Federal do Ceará, Fortaleza, CE, 60455-760, Brazil
[4] Graduate Programs in Statistics and Management Engineering, Universidade Federal de Pernambuco, Recife 50670-901, Brazil
[*]Corresponding author, e-mail: leandro@dema.ufc.br

Abstract

Recently, the maximin decision rule was introduced within the graph model for conflict resolution (GMCR). Through this introduction, a new stability concept was derived, called Maximin_h stability. This concept can be useful to deal with conflicts in which decision makers (DMs) have cautious profiles, that is, in situations in which no prior information about the preferences of other DMs is necessary and DMs want to prevent themselves from the worst scenario. In this paper, our objective is to propose matrix representations in order to determine more efficiently Maximin_h stable states in the GMCR for bilateral conflicts.

Keywords: maximin stability; matrix method; conflict resolutions

1. Introduction

A strategic conflict is a disagreement between individuals, groups, organizations or nations, often arising from differences in interests, values, objectives or beliefs. To obtain a deeper understanding of conflicts and find possible ways to resolve them, mathematical models have been used to represent and analyse conflicts. These models provide a systematic way to manage conflicts and reach mutually acceptable solutions.

Among the various models existing in the literature, the Graph Model for Conflict Resolution (GMCR) (Kilgour et al., 1987) has attracted significant attention from prominent researchers, mainly because it is flexible and enables the study and resolution of strategic conflicts in various domains of human interactions.

An important and fundamental part of modelling a strategic conflict through GMCR is the stability analysis stage. In the stability analysis of this model, there are several stability concepts, such as: Nash stability (Nash, 1950), general metarationality (GMR) (Howard, 1971), symmetric metarationality (SMR) (Howard, 1971), sequential stability (Fraser & Hipel, 1979), symmetric sequential stability (Rêgo & Vieira, 2016), limited-move stability of horizon *h* (Kilgour, 1985) and $Maximin_h$ stability (Rêgo & Vieira, 2020) that try to capture the way DMs behave when they are in conflict situations. Some of these concepts, for example: limited-move stability of horizon *h* and $Maximin_h$ stability, have variable horizon, i.e., allow for various reaction and counteraction movements by the DMs involved in the conflict.

The $Maximin_h$ stability has some important properties. For example, it includes Nash, GMR and SMR stabilities as special cases. In addition, it can be useful in conflict situations in which a DM has no knowledge about the opponents' preferences, as it assumes that the focal DM acts in the conflict in such a way as to anticipate that the worst-case scenario will happen after his initial move.

To obtain *Maximin$_h$* stability, a procedure similar to backward induction can be used. However, this can be difficult when there are many states in the conflict or when analysing stability over a long horizon. Many existing works in the literature on GMCR use matrix representations to capture stabilities according to different stability notions (Xu et al., 2007a), (Xu et al., 2007b), (Xu et al., 2009), (Wu et al., 2020) and (Rêgo & Vieira, 2021). Motivated by the ideas of these works, in this paper the main objective is to provide matrix representations to more easily determine which states are stable according to the notion of *Maximin$_h$* stability.

This work is organized as follows. In Section 2, a brief reminder about the GMCR and about the stability concept *Maximin$_h$* is made. In Section 3, we present the matrix representation of *Maximin$_h$* stability for conflicts with two DMs. Finally, we finish in Section 4 with main conclusions and directions for future work.

2. Background

Based on concepts of game theory and on the conflict analysis of Fraser and Hipel (Fraser & Hipel, 1984), the GMCR describes a conflict by a set of DMs, $N = \{1, 2, \ldots, n\}$, a set of possible states, S= $\{s, s_1, s_2, \ldots, s_w\}$, that can arise in a conflict according to actions that can be taken by individuals involved in a conflict. For each DM, the way in which he can change the conflict state is modelled by a directed graph, where nodes represent the states and the arcs represent the moves that DMs can make to change the conflict from one state to another. Let $R_i(s)$ denote the set of states to where DM i can move to in a single step from state s. Finally, each DM has some preferences over the set of states that can be expressed by an asymmetric binary relation on S, in which for each pair of states s, $s_1 \in S$ the notation s $\succ_i s_1$ expresses DM i 's strict preference for state s over state s_1. For *Maximin$_h$* stability, $\succ_i$ is assumed to be also transitive and complete generating a complete ranking of the states for DM i. Let the subset of states in $R_i(s)$ that are preferred to s by DM i, be denoted by $R_i^+(s)$.

The differences in the stability concepts commonly used in the GMCR are related to the credibility of the reactions of the DMs and to the number of steps ahead that the conflict is analysed, which is known as conflict horizon. In this paper, our objective is to provide matrix representation to obtain stable states according to the concept of *Maximin$_h$* stability. Thus, next, we will briefly recall this concept and its main components, considering conflicts with two DMs to make the understanding of matrix representation given in the next section easier.

Consider a GMCR with two DMs, DM i and DM j, and we admit that the focal DM while analyzing the conflict of horizon h believes that the other DM will act moving to the worst possible scenario for the focal DM knowing that the focal DM always moves to the best possible scenario considering a horizon h' less than h.

Let $K_i(s)$ be the cardinality of the set of states that are worse than the state s for DM i. A DM who foresees a sequence of length at most h is said to be a DM with horizon h, where h is a positive integer number. Let $G_h^i(j, s) \in S, i, j \in N$, be the state that DM i believes that will be the final state of conflict with horizon h, in that DM j moves first from state s. Then, in Rêgo & Vieira (2020) $G_h^i(\bullet, s)$ is defined inductively as follows:

$$G_h^i(i,s) = \begin{cases} s, \text{ if } h = 0 \text{ or } R_i(s) = \emptyset \text{ or } K_i(s) \geqslant A_h^i(i,s), \\ \quad G_{h-1}^i\left(j, M_h^i(i,s)\right), \text{ otherwise,} \end{cases}$$

where $M_h^i(i,s)$ is some state $s_1^* \in R_i(s)$ that satisfies $K_i\left(G_{h-1}^i(j, s_1^*)\right) = \max\left\{K_i\left(G_{h-1}^i(j, s_1)\right) : s_1 \in R_i(s)\right\}, j \neq i$ and $A_h^i(i,s) = K_i\left(G_{h-1}^i\left(j, M_h^i(i,s)\right)\right)$. The intuition is that, when it is DM i's turn to move and the horizon is h, he anticipates the state that gives the best result for him among the states that he anticipates when it is DM j's turn to move and the horizon is $h-1$.

In turn, $G_h^i(j,s)$ is given by

$$G_h^i(j,s) = \begin{cases} s, \text{ if } h = 0 \text{ or } R_j(s) = \emptyset \text{ or } K_i(s) \leq A_h^i(j,s), \\ G_{h-1}^i\left(i, M_h^i(j,s)\right), \text{ otherwise,} \end{cases}$$

where $M_h^i(j,s)$ is some state $s_1^* \in R_j(s)$ that satisfies $K_i\left(G_{h-1}^i(i,s_1^*)\right) = \min\left\{K_i\left(G_{h-1}^i(i,s_1)\right) : s_1 \in R_j(s)\right\}, j \neq i$ and $A_h^i(j,s) = K_i\left(G_{h-1}^i\left(i, M_h^i(j,s)\right)\right)$. The intuition is that, when it is DM j's turn to move and the horizon is h, DM i anticipates the state that gives the worst result for him among the states that he anticipates when it is his turn to move and the horizon is $h-1$. Thus, a maximin stable state is defined as follows (Rêgo & Vieira, (2020).

Definition 1: A state s $\in S$ is maximin stable with horizon h for DM $i \in N$ iff $G_h^i(i,s) = s$.

3. Matrix Representations of *Maximin$_h$* Stability

In this section, matrix representation for determining *Maximin$_h$* stabilities is presented. To present the main result of this section, it is necessary to remember some matrices that will be important for a good understanding of this paper. The first three matrices below can be found with more details in work of Xu et al., (2009).

Consider the $|S| \times |S|$, 0-1 matrix, J_i, defined as

$$J_i(s,s_1) = \begin{cases} 1, \text{ if } s_1 \in R_i(s), \\ 0, \quad \text{otherwise.} \end{cases}$$

In other words, the J_i matrix entry (s,s_1) receives the value 1 if state s_1 is unilaterally reachable from state s by DM i. Otherwise, the (s,s_1) entry of J_i matrix receives the value 0.

Also consider the $|S| \times |S|$ strict preference matrix given by

$$P_i^+(s,s_1) = \begin{cases} 1, & \text{if } s_1 \succ_i s, \\ 0, & \text{otherwise.} \end{cases}$$

Let the signal matrix, denoted by $sign(\cdot)$, such that if W is a matrix of order $|S|$, then $sign(W)$ is a matrix of order $|S|$ with (s,s_1) entry given by:

$$sign[W(s,s_1)] = \begin{cases} 1, if\ W(s,s_1) > 0, \\ 0, if\ W(s,s_1) = 0, \\ -1, if\ W(s,s_1) < 0. \end{cases}$$

To achieve our main result we need to introduce some new matrices. First, consider D to be the identity matrix of order $|S|$ and let $E' = [\,1\ 1\ 1\ \cdots\ 1\,]$ be a row matrix with dimension $|S|$. Also consider that if F is a 0-1 matrix then F^c represents the complement of the binary matrix F, i.e., Fc(s,s1)= 1- F(s,s1). Finally, let $\boldsymbol{K}_i$ be a column matrix with dimension $|S|$ given by $\boldsymbol{K}_i = [E' \cdot P_i^+]^T$, where $[\bullet]^T$ represents transpose of a matrix. The matrix $\boldsymbol{K}_i$ defined above and the function $K_i(s)$ are related, i.e., it is not difficult to conclude that if s and $s_1 \in S$, then $\boldsymbol{K}_i(s_1, 1) = K_i(s_1)$.

Now we will define two column matrices with dimension $|S|$, called of $Min(A)$ and $Max(A)$. Let $A = [a_{tk}]$ a matrix of order $|S|$, then $Min(A) = [b_k]$ and $Max(A) = [c_k]$ are such that $b_k = \min(a_{k1}, a_{k2}, \ldots, a_{ks})$ and $c_k = \max(a_{k1}, a_{k2}, \ldots, a_{ks})$, respectively. Note that the number in row k of matrix $Min(A)$ is the smallest number that appears in row k of A. Similarly, the number in row k of matrix $Max(A)$ is the largest number that appears in the row k of matrix A.

We now define two important matrices, denoted by B_h^i and B_h^j, which will be useful to capture the anticipated payoff that DM i expects to receive after a conflict horizon h if he/she or the opponent, respectively, is the first one to move. Each row of these matrices contains what DM i anticipates to receive at the end of the conflict if the conflict state changes from the row state to the column state. If the column state is not reachable from and not equal to the row state, then the corresponding entry will be zero in B_h^i and will be equal to |S| in B_h^j, since in the first case, where the focal DM moves, a maximization move will follow and in the second one a minimization move will follow. Formally, B_h^i and B_h^j are defined, respectively, as follows:

$$B_h^i = J_i \circ \left(C_{h-1}^j \cdot E'\right)^T + \mathrm{D} \circ (\boldsymbol{K}_i \cdot E')^T$$

and

$$B_h^j = |S| \times \left(J_j + D\right)^c + (J_j \circ \left(C_{h-1}^i \cdot E'\right)^T + \mathrm{D} \circ (\boldsymbol{K}_i \cdot E')^T)$$

in which the notation "$\circ$" represents the Hadamard product, $C_0^i = C_0^j = \boldsymbol{K}_i$, $C_h^i = \max(B_h^i)$ and $C_h^j = \min(B_h^j)$, for all integer horizon $h \geq 1$.

Note that obtaining the matrices B_h^i and B_h^j, for a certain fixed horizon, depends on a recursion involving the matrices C_h^i and C_h^j. In the Lemma below, the proof of which will be omitted in this short paper, it is established that the entry $(s_1, 1)$ of the column matrix C_h^i is equal to the maximum gain that DM i can obtain by choosing to stay in s or moving away from s and what he anticipates when DM j moves in the sequence considering a horizon of size $h-1$. Similarly, we have that the entry of $(s_1, 1)$ of the column matrix C_h^j is equal to the minimum gain that DM i can obtain when DM j choose to stay in s or move away from s and what DM i anticipates when he moves next considering a horizon of size $h-1$.

Lemma1: For all s $\in S$ and integer horizon $h \geq 1$, we have:

$$C_h^i(s,1) = Max\{A_h^i(i,s),\ K_i(\mathrm{s})\}$$

and

$$C_h^j(s,1) = Min\{A_h^i(j,s),\ K_i(\mathrm{s})\}$$

Finally, in Theorem 1 below, whose proof will be omitted in this short paper, we provide a result that allows obtaining *Maximin$_h$* stable states using only matrix operations.

Theorem 1: A state s $\in S$ is *Maximin$_h$* stable for DM i iff $\left(\mathrm{sign}\left(\boldsymbol{K}_i - C_h^i\right)\right)(\mathrm{s},1) = 0$.

Example: In order to illustrate the matrix representation obtained in this work, consider a hypothetical conflict containing two DMs, DM i and DM j, and four states, s_1, s_2, s_3 and s_4. Consider that the accessibility sets of this conflict are: $R_i(s_1) = s_3$, $R_i(s_2) = s_4$, $R_i(s_3) = s_1$, $R_i(s_4) = s_2$, $R_j(s_1) = s_2$, $R_j(s_2) = s_1$, $R_j(s_3) = s_4$ and $R_j(s_4) = s_3$. Suppose further that the preferences of DM i and DM j are, respectively, $s_1 \succ_i s_3 \succ_i s_2 \succ_i s_4$ and $s_3 \succ_i s_1 \succ_i s_4 \succ_i s_2$. In this conflict, we have that the matrices J_i, J_j and P_i^+ are, respectively, given by:

$$J_i = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \ J_j = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \text{ and } P_i^+ = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{pmatrix}.$$

Next, note that $\boldsymbol{K}_i = [E'.P_i^+]^T = \left[(1\ 1\ 1\ 1).\begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{pmatrix}\right]^T = \begin{bmatrix} 3 \\ 1 \\ 2 \\ 0 \end{bmatrix}$ and, consequently, $\left(\boldsymbol{K}_i \cdot E'\right)^T = \begin{pmatrix} 3 & 1 & 2 & 0 \\ 3 & 1 & 2 & 0 \\ 3 & 1 & 2 & 0 \\ 3 & 1 & 2 & 0 \end{pmatrix}$.

For simplicity, we will use the matrix representation obtained in Theorem 1 to analyze stability of the conflict states for DM 1, considering horizons $h = 1, 2$ and 3.

First note that, for $h = 1$, to obtain the column vector C_1^i, we first need to obtain the matrix B_1^i, which is given by:

$$B_1^i = J_i \circ (\boldsymbol{K}_i \cdot E')^T + \mathrm{D} \circ (\boldsymbol{K}_i \cdot E')^T$$
$$= \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \circ \begin{pmatrix} 3 & 1 & 2 & 0 \\ 3 & 1 & 2 & 0 \\ 3 & 1 & 2 & 0 \\ 3 & 1 & 2 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \circ \begin{pmatrix} 3 & 1 & 2 & 0 \\ 3 & 1 & 2 & 0 \\ 3 & 1 & 2 & 0 \\ 3 & 1 & 2 & 0 \end{pmatrix} = \begin{pmatrix} 3 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 \\ 3 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

As $C_1^i = \max\left(B_1^i\right) = \begin{bmatrix} 3 \\ 1 \\ 3 \\ 1 \end{bmatrix}$, we have that $\boldsymbol{K}_i - C_1^i = \begin{bmatrix} 0 \\ 0 \\ -1 \\ -1 \end{bmatrix}$ and consequently $\mathrm{sign}\left(\boldsymbol{K}_i - C_1^i\right) = \mathrm{sign}\left(\begin{bmatrix} 0 \\ 0 \\ -1 \\ -1 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \\ -1 \\ -1 \end{bmatrix}$. Therefore, by Theorem 1, we have that states s_1 and s_2 are *Maximin₁* stable for DM i.

For $h = 2$, note that $B_2^i = J_i \circ \left(C_1^j \cdot E'\right)^T + \mathrm{D} \circ (\boldsymbol{K}_i \cdot E')^T$. Thus, to obtain this matrix, and consequently C_2^i, we first need to obtain the matrix C_1^j.

Note that

$$B_1^j = |S| \times \left(J_j + D\right)^c + (J_j \circ \left(C_0^i \cdot E'\right)^T + \mathrm{D} \circ (\boldsymbol{K}_i \cdot E')^T)$$
$$= \begin{pmatrix} 0 & 0 & 4 & 4 \\ 0 & 0 & 4 & 4 \\ 4 & 4 & 0 & 0 \\ 4 & 4 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 1 & 0 & 0 \\ 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 \end{pmatrix} + \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 3 & 1 & 4 & 4 \\ 3 & 1 & 4 & 4 \\ 4 & 4 & 2 & 0 \\ 4 & 4 & 2 & 0 \end{pmatrix}.$$

That way, we have $C_1^j = \min\left(B_1^j\right) = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$. Now note that

$$B_2^i = J_i \circ \left(C_1^j \cdot E'\right)^T + \mathrm{D} \circ (\boldsymbol{K}_i \cdot E')^T$$
$$= \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \circ \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \circ \begin{pmatrix} 3 & 1 & 2 & 0 \\ 3 & 1 & 2 & 0 \\ 3 & 1 & 2 & 0 \\ 3 & 1 & 2 & 0 \end{pmatrix} = \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

As $C_2^i = \max\left(B_2^i\right) = \begin{bmatrix} 3 \\ 1 \\ 2 \\ 1 \end{bmatrix}$, we have that $\boldsymbol{K}_i - C_2^i = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -1 \end{bmatrix}$ and consequently, $\operatorname{sign}\left(\boldsymbol{K}_i - C_2^i\right) = \operatorname{sign}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ -1 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -1 \end{bmatrix}$, which implies, by Theorem 1, that states $s_1, s_2,$ and s_3 are *Maximin₂* stable for DM i.

Analogously to the previous cases, note that for $h = 3$ we have $B_3^i = J_i \circ \left(C_2^j \cdot E'\right)^T + \mathrm{D} \circ (\boldsymbol{K}_i \cdot E')^T$. Thus, to obtain this matrix, and consequently C_3^i, we first need to obtain the matrix $C_2^j = \min\left(B_2^j\right)$.

Note that

$$B_2^j = |S| \times \left(J_j + D\right)^c + (J_j \circ \left(C_1^i \cdot E'\right)^T + \mathrm{D} \circ (\boldsymbol{K}_i \cdot E')^T)$$
$$= \begin{pmatrix} 0 & 0 & 4 & 4 \\ 0 & 0 & 4 & 4 \\ 4 & 4 & 0 & 0 \\ 4 & 4 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 1 & 0 & 0 \\ 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 3 & 0 \end{pmatrix} + \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 3 & 1 & 4 & 4 \\ 3 & 1 & 4 & 4 \\ 4 & 4 & 2 & 1 \\ 4 & 4 & 3 & 0 \end{pmatrix}.$$

Thus, we conclude that $C_2^j = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}$. Now notice that

$$B_3^i = J_i \circ \left(C_2^j \cdot E'\right)^T + \mathrm{D} \circ (\boldsymbol{K}_i \cdot E')^T = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 3 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

As $C_3^i = \max\left(B_3^i\right) = \begin{bmatrix} 3 \\ 1 \\ 2 \\ 1 \end{bmatrix}$, we have that $\boldsymbol{K}_i - C_3^i = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -1 \end{bmatrix}$ and consequently, $\operatorname{sign}\left(\boldsymbol{K}_i - C_2^i\right) = \operatorname{sign}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ -1 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -1 \end{bmatrix}$, which implies that $s_1, s_2,$ and s_3 are *Maximin₃* stable for DM i.

Rêgo & Vieira (2020) presented a result that establishes that *Maximin₂ₙ₊₁* stability implies *Maximinₕ* for all $h \geq 2n + 1$ and that *Maximin₂ₙ* implies *Maximin₂ₘ* for all m such that $1 \leq m \leq n$. Thus, if a state satisfies *Maximin₃* stability for DM i, it also satisfies *Maximinh* stability for each $h \geq 3$. And if a state does not satisfy *Maximin₂* stability for DM i, it does not satisfy *Maximinₕ* stability for every h. Using this result, we have that in the hypothetical conflict above, states $s_1, s_2,$ and s_3 are *Maximinₕ* stable, for DM i, for every $h \geq 3$, while state s_4 is not *Maximinₕ* stable for any value of h.

4. Conclusions

The $Maximin_h$ stability concept has shown itself to be an important notion of stability, as it can be useful to model the behavior of a cautious decision maker, that is, the decision maker acts in the conflict in order to defend himself against the worst possible scenario that may arise in course of the conflict.

In this paper, we provide matrix representation to obtain $Maximin_h$ stability in bilateral conflicts. The matrix method proposed here can make the stability analysis stage, according to this concept, more efficient, especially in conflicts with many feasible states or when one wants to analyse a conflict a large number of steps ahead. In future work, we intend to extend this result to multilateral conflicts and also expand this research by considering other preference structures.

Acknowledgments. The authors would like to acknowledge the financial support of Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ) under grants 406697/2023-0 and 308980/2021-2.

References

N. M. Fraser, K. W. Hipel, Conflict analysis: models and resolutions, North-Holland, 1984.

N. M. Fraser, K. W, Hipel, Solving Complex Conflicts, IEEE Transactions on Systems, Man, and Cybern, SMC-9, pp. 805-817, 1979.

N. Howard, Paradoxes of Rationality, MIT Press, Cambridge, MA, 1971.

D. M. Kilgour, Anticipation and stability in two-person noncooperative games, Dynamic models of international conflict, pp. 26-51, 1985.

D. Kilgour, K. Hipel, L. Fang, The graph model for conflicts, Automatica, 23, no. 1, pp. 41-55, 1987.

J. F. Nash, Equilibrium Points in n - Person Games, Proceedings of National Academy of Sciences of the U.S.A., 36, pp. 48-49, 1950.

L. C. Rêgo, G. I. A. Vieira, Symmetric Sequential Stability in the Graph Model for Conflict Resolution with Multiple Decision Makers, Group Decision and Negotiation, p. 1-18, 2016.

L. C. Rêgo, G. I. A. Vieira, Matrix representation of solution concepts in the graph model for conflict resolution with probabilistic preferences and multiple decision makers, Group Decision and Negotiation, 30:697–717, 2021.

L. C. Rêgo, G. I. A. Vieira, Maximinh stability in the graph model for conflict resolution for bilateral conflicts, IEEE Transactions on Systems Man Cybernetics-Systems 50, pp. 3760-3769, 2020.

N. Wu, D.M. Kilgour, K.W. Hipel, Y. Xu, Matrix representation of stability defnitions for the graph model for confict resolution with reciprocal preference relations, Fuzzy Sets Syst, 2020.

H. Xu, K.W. Hipel, D.M. Kilgour, Matrix representation of conficts with two decision-makers, In: IEEE international conference on systems, man and cybernetics, ISIC, IEEE, pp 1764–1769, 2007a.

H. Xu, D.M. Kilgour, K.W. Hipel, Matrix representation of solution concepts in graph models for two decision-makers with preference uncertainty, Dyn Contin Discret Impul Syst 14(S1):703–737, 2007b.

H. Xu, K.W. Hipel, D.M. Kilgour, Matrix representation of solution concepts in multiple decision maker graph models, IEEE Transactions on Systems, Man, and Cybernetics, Part A Systems and Humans, 39, (1), 96-108, 2009.

Third-party Conflict Intervention based on a Two-stage Optimization Model within the Inverse Graph Model Framework

Nannan Wu [1,2], Ruyi Huang [1], Yejun Xu [2] and Liping Fang [3*]

[1] School of Management Science and Engineering, Nanjing University of Information Science and Technology, Nanjing, China
[2] College of Management and Economics, Tianjin University, Tianjin, China
[3] Department of Mechanical and Industrial Engineering, Toronto Metropolitan University, Toronto, Ontario, Canada
[*] lfang@torontomu.ca

Abstract

In the process of resolving real-world conflicts, third-party intervention often plays an important role. The graph model for conflict resolution (GMCR) is a useful methodology in resolving conflict, in which the forward GMCR can be used for the prediction and analysis of conflict evolution, and the inverse GMCR can help decision makers (DMs) adjust preferences to achieve a given desired state. This study proposes an improved 0-1 mixed linear approach to define fuzzy Nash stability first. Subsequently, a two-stage optimization model is developed to modify DMs' preferences with the minimum adjustment cost first and the minimum adjustment amount second such that a desired state is an equilibrium.

Keywords: Graph model for conflict resolution; Third-party intervention; 0-1 mixed linear approach; Preference adjustments; Optimization

1. Introduction

The graph model for conflict resolution (GMCR) (Kilgour, Hipel, & Fang, 1987), a game-theoretic approach, can provide strategic advice for decision makers (DMs) involved in a conflict. In the determination and prediction of conflict outcomes, the forward GMCR can comprehensively consider the main aspects of conflicts, use the minimal information of DMs' preferences to model conflicts, dynamically simulate interactive processes, predict equilibria, and provide reliable suggestions for conflict resolution (Xu, Hipel, & Kilgour, 2009). The inverse GMCR can explore the required preferences of DMs if a third-party wants to make a specific state stable (Kinsara, Kilgour, & Hipel, 2015). The core idea of the inverse GMCR is producing the DMs' possible preferences over states (preference relation profiles) that will make a desired state stable for a DM or an equilibrium for a conflict. However, it is difficult for a third-party to choose an appropriate preference relation profile to design a mediation strategy since there are usually numerous possible preference relation profiles (Han, Xu, Fang, & Hipel, 2022; Wang, Hipel, & Fang, 2018).

Considering the cost of conflict intervention, some optimization models have been proposed. Zhang, Dong, and Chiclana (2023) proposed a linear minimum cost optimization model to consider conflict intervention problem under Nash stability and general metarationality (GMR). Based on the advances on GMCR research, particularly in the minimum cost inverse graph model (Wu, Xu, & Ke, 2019; Rêgo, Silva, & Rodrigues, 2021; Tao, Su, & Javed, 2021) and 0-1 linear methods in Zhang et al. (2023), this study proposes a two-stage optimization model to support conflict resolution with the minimum preference adjustment cost first and then the minimum amount of preference adjustments under fuzzy Nash (FNash) stability.

The contributions of this short paper are twofold: 1) An improved 0-1 mixed linear approach is proposed to define FNash; and 2) A two-stage optimization model is developed to modify DMs' preferences with the minimum adjustment cost first and then the minimum amount of preference adjustments for determining the required preferences with inverse GMCR. The remainder of this short paper is organized as follows: Section 2 introduces the forward and inverse GMCR. Sections 3 and 4 are dedicated to the improved 0-1 mixed linear approach and the two-stage optimization model, respectively. Section 5 concludes the short paper and provides prospects for future work.

2. Preliminaries

2.1 Forward graph model for conflict resolution (GMCR)

Within a graph model, $N=\{1, 2, \ldots, n\}$ is the set of DMs. DMs' possible courses of action are called options, which are indicated by $O = \{o_1, o_2, \ldots, o_q\}$. The set of feasible states consisting of option selections is indicated by $S = \{s_1, s_2, \ldots, s_m\}$. The collection consisting of each DM's set of directed arcs is denoted by $A = \{A_1, A_2, \ldots, A_n\}$, where, for any $s_i, s_j \in S$, $(s_i, s_j) \in A_k$ if and only if (*iff*) DM k can cause the conflict to move from state s_i to state s_j. If $B = \{B^1, B^2, \ldots, B^n\}$ denotes the collection consisting of each DM's matrix of preference relations, then a graph model can be described as $\langle N, O, S, A, B\rangle$.

The following is a brief introduction to stability analysis, when DMs have fuzzy preference relations (FPRs).

Definition 1. FPR (Tanino, 1984): Let $S = \{s_1, s_2, \ldots, s_m\}$ and $k \in N$. DM k's FPR over S can be represented by a matrix $B^k = (b_{ij}^k)_{m\times m} \subset S \times S$, where $b_{ij}^k = \mu^k(s_i, s_j): S \times S \to [0,1]$ represents DM k's degree of preference for s_i over s_j, and satisfies $b_{ij}^k + b_{ji}^k = 1$, $b_{ii}^k = 0.5$, and $0 \le b_{ij}^k \le 1$, for $k \in N$, and $i, j = 1, 2, \ldots, m$.

In Definition 1, the element b_{ij}^k represents the preference degree of DM k for state s_i over state s_j. For an FPR $B^k = (b_{ij}^k)_{m\times m} \subset S \times S$, a commonly used consistency index (CI) (Herrera-Viedma, Alonso, Chiclana, & Herrera, 2007) is calculated as

$$CI(B^k) = 1 - \frac{4}{m(m-1)(m-2)} \sum_{1\le i<j<z\le m}^{m} \left|b_{ij}^k + b_{jz}^k - b_{iz}^k - 0.5\right|$$
$$= 1 - \frac{4}{m(m-1)(m-2)} \sum_{i=1}^{m-2} \sum_{j=i+1}^{m-1} \sum_{z=j+1}^{m} \left|b_{ij}^k + b_{jz}^k - b_{iz}^k - 0.5\right| \quad (1)$$

Usually, a threshold $\overline{CI}$ is set in advance. If the current consistency level is lower than the threshold, $CI(B^k) < \overline{CI}$, the DM's preferences need to be revised. Otherwise, the consistency level of B^k is to be acceptable, i.e. $CI(B^k) \ge \overline{CI}$.

If $B^k = (b_{ij}^k)_{m\times m}$ is DM k's FPR over S, DM k's *fuzzy relative strength of preference* (FRSP) of state s_i over s_j is denoted as α_{ij}^k, where $\alpha_{ij}^k = \alpha^k(s_i, s_j) = b_{ij}^k - b_{ji}^k$, $k \in N$, $i, j = 1, 2, \ldots, m$. If DM k is willing to move from state $s_t \in S$ to $s_i \in S$ iff $\alpha_{it}^k = \alpha^k(s_i, s_t) \ge \lambda^k$ or $\alpha_{ti}^k = \alpha^k(s_t, s_i) \le -\lambda^k$, then DM k's *fuzzy satisficing threshold* (FST) is called λ^k $(0 < \lambda^k \le 1)$.

For each DM $k \in N$, its *reachable list* from state $s_t \in S$ is $R_k(s_t) = \{s_i \in S: (s_t, s_i) \in A_k\}$. DM k's *fuzzy unilateral improvement list* from state s_t is $R_k^+(s_t) = \{s_i \in R_k(s_t): \alpha_{it}^k \ge \lambda^k\}$. The members of $R_k(s_t)$ and $R_k^+(s_t)$ are called *unilateral moves* (UMs) and *fuzzy unilateral improvements* (FUIs) from s_t by DM k, respectively.

Definition 2. (Hipel, Kilgour, & Bashar, 2011) A state $s_t \in S$ is FNash stable for DM $k \in N$, denoted by $s_t \in S_k^{FNash}$, iff $R_k^+(s_t) = \emptyset$.

Based on the UMs and FUIs for DMs, different stability concepts can be proposed to model diverse behaviors of DMs. In this study, due to space limitations, we only consider FNash stability.

2.2 Inverse graph model for conflict resolution (GMCR)

The difference between the forward and inverse GMCR is that, the forward GMCR starts with the input of DMs' preferences and ends with equilibria, while the order of the inverse GMCR is opposite, by inputting a desired equilibrium to inversely infer the required preferences of each DM (Kinsara et al., 2015). Therefore, the forward and inverse GMCR are applied to different scenarios accordingly. The inverse GMCR can help third-party obtain information about DMs' required preferences, or adjust DMs' preferences at minimum cost to achieve a desired equilibrium. Our research aims to obtain the required preferences for a given desired solution, while considering both the minimum cost and amount of preference adjustments based on a linear optimization approach. The optimal solutions help third-party develop good mediation strategies.

3. An improved 0-1 linear approach to fuzzy Nash (FNash) stability

Based on the study of Zhang et al. (2023) and further considering DMs' FSTs, an improved approach for representing FNash stability via a series of 0-1 mixed linear constraints is developed below.

Lemma 1: Let $B^k = (b_{ij}^k)_{m\times m}$ be DM k's FPR over S, and its FST is λ^k. Define

$$x_{it}^k = \begin{cases} 0, \alpha_{it}^k \geq \lambda^k \\ 1, \alpha_{it}^k < \lambda^k \end{cases}, s_i \in R_k(s_t). \tag{2}$$

It follows that

1) A state s_t is FNash stable for DM k iff $\sum_{s_i\in R_k(s_t)} x_{it}^k = \|R_k(s_t)\|$.

2) A state s_t is an FNash equilibrium iff $\sum_{k=1}^n \sum_{s_i\in R_k(s_t)} x_{it}^k = \sum_{k=1}^n \|R_k(s_t)\|$.

Lemma 2: The 0-1 variable x_{it}^k $(s_i \in R_k(s_t))$ used in Lemma 1 can be determined using the following constraints:

$$\begin{cases} \alpha_{it}^k - \lambda^k < Z \cdot x_{it}^k \\ \alpha_{it}^k - \lambda^k \geq Z \cdot (x_{it}^k - 1) \\ x_{it}^k \in \{0,1\} \end{cases} \tag{3}$$

where Z is an adequately large number.

Lemmas 1 and 2 provide a 0-1 mixed linear approach to determine the FNash stable state within the graph model framework. Due to space limitations, we omit their proof.

4. Two-stage optimization model for fuzzy Nash (FNash) stability

Assuming that the minimum adjustment cost of DMs' preferences is regard as the primary goal, the optimization model is constructed as follows:

Stage 1: Minimizing the costs of preference adjustments

Let $\bar{\alpha}_{it}^k$ be DM k's FRSP of state s_i over s_t after preference adjustments. In order for s_t to be FNash stable, $R_k^+(s_t) = \emptyset$ ($\bar{\alpha}_{it}^k < \lambda^k$ or $\bar{\alpha}_{ti}^k > -\lambda^k$) for $k \in N$ and $s_i \in R_k(s_t)$, which means that

$$\begin{cases} \bar{\alpha}_{it}^k - \lambda^k < Z \cdot x_{it}^k \\ \bar{\alpha}_{it}^k - \lambda^k \geq Z \cdot (x_{it}^k - 1) \\ x_{it}^k \in \{0, 1\} \\ \sum_{s_i\in R_k(s_t)} x_{it}^k = \|R_k(s_t)\| \end{cases} \tag{4}$$

where $\bar{\alpha}_{it}^k = \bar{b}_{it}^k - \bar{b}_{ti}^k$.

Let c_{ij}^k be the unit cost of preference adjustments for changing b_{ij}^k into $\bar{b}_{ij}^k$ by DM k. Obviously, the primary objective is to optimize the adjustment costs of FPRs between $\{B^1, B^2, \ldots, B^n\}$ and $\{\bar{B}^1, \bar{B}^2, \ldots, \bar{B}^n\}$:

$$min\, J_1^{FNash} = \sum_{k=1}^n \sum_{i=1}^{m-1} \sum_{j=i+1}^m c_{ij}^k \left|b_{ij}^k - \bar{b}_{ij}^k\right|$$

Thus, the inverse graph model with minimum adjustment costs of FPRs for state s_t to be an FNash equilibrium is constructed as Model (1):

$$\begin{cases} \min J_1^{FNash} = \sum_{k=1}^{n}\sum_{i=1}^{m-1}\sum_{j=i+1}^{m} c_{ij}^k \left|b_{ij}^k - \bar{b}_{ij}^k\right| \\ s.t. \begin{cases} \bar{\alpha}_{ij}^k = \bar{b}_{ij}^k - \bar{b}_{ji}^k, k \in N;\ i,j = 1,2,\ \ldots,\ m \\ \bar{\alpha}_{it}^k - \lambda^k < Z \cdot x_{it}^k, k \in N;\ s_i \in R_k(s_t) \\ \bar{\alpha}_{it}^k - \lambda^k \geq Z \cdot (x_{it}^k - 1), k \in N;\ s_i \in R_k(s_t) \\ \sum_{k \in N}\sum_{s_i \in R_k(s_t)} x_{it}^k = \sum_{k \in N} \|R_k(s_t)\| \\ CI(\bar{B}^k) = \\ 1 - \frac{4}{m(m-1)(m-2)}\sum_{i=1}^{m-2}\sum_{j=i+1}^{m-1}\sum_{z=j+1}^{m}\left|\bar{b}_{ij}^k + \bar{b}_{jz}^k - \bar{b}_{iz}^k - 0.5\right| \geq \overline{CI}, k \in N \\ \bar{b}_{ij}^k + \bar{b}_{ji}^k = 1, k \in N;\ i,j = 1,2,\ \ldots,\ m \\ \bar{b}_{ij}^k \in [0,1], k \in N;\ i,j = 1,2,\ \ldots,\ m \\ x_{it}^k \in \{0,1\}, k \in N;\ t,i = 1,2,\ \ldots,\ m \end{cases} \end{cases} \tag{5}$$

In Model (1), $\bar{b}_{ij}^k \in [0,1], k \in N, i,j = 1,2,\ \ldots,\ m$, are the decision variables.

Stage 2: Minimizing the amounts of preference adjustments

After the minimum cost of preference adjustments is determined, the second stage goal is to minimize the amounts of modifications in DMs' preference matrices. Assuming that $\bar{J}_1^{FNash}$ is the optimal value for Model (1), the second stage model is given as Model (2):

$$\begin{cases} \min J_2^{FNash} = \sum_{k=1}^{n}\sum_{i=1}^{m-1}\sum_{j=i+1}^{m} \delta_{ij}^k \\ s.t. \begin{cases} \sum_{k=1}^{n}\sum_{i=1}^{m-1}\sum_{j=i+1}^{m} c_{ij}^k \left|b_{ij}^k - \bar{b}_{ij}^k\right| = \bar{J}_1^{FNash} \\ \left|b_{ij}^k - \bar{b}_{ij}^k\right| \leq Z\delta_{ij}^k, k \in N;\ i,j = 1,2,\ \ldots,\ m \\ \delta_{ij}^k \in \{0,1\}, k \in N;\ i,j = 1,2,\ \ldots,\ m \\ \text{The other constraints are the same as in Model (1)} \end{cases} \end{cases} \tag{6}$$

In Model (2), $\bar{b}_{ij}^k \in [0,1], k \in N, i,j = 1,2,\ \ldots,\ m$, are the decision variables, Z is a sufficiently large number, and $\delta_{ij}^k, k \in N, i,j = 1,2,\ \ldots,\ m$, are 0-1 variables, which indicate whether b_{ij}^k is adjusted.

Note that in the two-stage optimization model, Stage 1 ensures that an equilibrium is achieved by minimizing the cost of preference adjustments. Based on the results (the minimum cost of preference adjustments) in Stage 1, Stage 2 makes sure that there is a minimum amount of changes in DMs' preference matrices to maximize the preservation of DMs' original preference information. Both Model (1) and Model (2) can be converted into a 0-1 mixed linear optimization model and have at least one optimal solution. When Model (1) has multiple optimal solutions, Model (2) may have multiple solutions. When Model (1) has a unique optimal solution, Model (2) also has the same unique solution. The unique solution problem is not the focus of this study and we do not provide detailed discussions because of space constraints.

5. Conclusions

This study presents an improved 0-1 mixed linear approach to define FNash stability and develops a two-stage optimization model to adjust DMs' preferences with the minimum cost first and amount second. In the future, we will present an inverse graph model with other stability concepts, such as fuzzy general metarationality (FGMR), fuzzy symmetric metarationality (FSMR), fuzzy sequential stability (FSEQ), and fuzzy symmetric sequential stability (FSSEQ). In addition, it is also an interesting research direction to consider large-scale groups or composite DMs in the inverse graph model.

References

Han, Y., Xu, H., Fang, L., & Hipel, K. W. (2022). An integer programming approach to solving the inverse graph model for conflict resolution with two decision makers. *Group Decision and Negotiation, 31*(1), 23–48.

Herrera-Viedma, E., Alonso, S., Chiclana, F., & Herrera, F. (2007). A consensus model for group decision making with incomplete fuzzy preference relations. *IEEE Transactions on Fuzzy Systems, 15*(5), 863-877.

Hipel, K. W., Kilgour, D. M., & Bashar, M. A. (2011). Fuzzy preferences in multiple participant decision making. *Scientia Iranica, 18*(3), 627-638.

Kilgour, D. M., Hipel, K. W., & Fang, L. (1987). The graph model for conflicts. *Automatica, 23*(1), 41-55.

Kinsara, R. A., Kilgour, D. M., & Hipel, K. W. (2015). Inverse approach to the graph model for conflict resolution. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 45*(5), 734–742.

Rêgo, L. C., Silva, H. V., & Rodrigues, C. D. (2021). Optimizing the cost of preference manipulation in the graph model for conflict resolution. *Applied mathematics computation, 392*(1), 125729.

Tanino, T. (1984). Fuzzy preference orderings in group decision making. *Fuzzy Sets and Systems, 12*(2), 117-131.

Tao, L., Su, X., & Javed, S. A. (2021). Inverse preference optimization in the graph model for conflict resolution based on the genetic algorithm. *Group Decision and Negotiation, 30*(1), 1085–1112.

Wang, J., Hipel, K. W., & Fang, L. (2018). Matrix representations of the inverse problem in the graph model for conflict resolution. *European Journal of Operational Research, 270*(1), 282–293.

Wu, Z., Xu, H., & Ke, G. (2019). The strategy of third-party meditation based on the option prioritization in the graph model, *Journal of Systems Science and Systems Engineering*, *28*(4), 399–414.

Xu, H., Hipel, K. W., & Kilgour, D. M. (2009). Matrix representation of solution concepts in multiple-decision-maker graph models. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 39*(1), 96-108.

Zhang, H., Dong, Y., & Chiclana, F. (2023). A graph model with minimum cost to support conflict resolution and mediation in technology transfer of new product co-development. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 53*(11), 7210-7222. doi:10.1109/TSMC.2023.3296487

A meta-level approach to the general equilibrium selection problem

Alexandre Bevilacqua Leoneti[1[0000-0002-0744-8866]] and Takahiro Suzuki[2[0000-0002-3436-6831]]

[1] School of Economics, Business Administration and Accounting, University of São Paulo, Ribeirão Preto, Brazil
ableoneti@usp.br

[2] Department of Civil Engineering, The University of Tokyo. 7-3-1, Hongo, Bunkyo-ku, Tokyo, Japan.
suzuki-tkenmgt@g.ecc.u-tokyo.ac.jp

Abstract

Suzuki and Leoneti (2022) demonstrated that Convergence Theory applied to general meta-level equilibrium selection problem can be used for explaining how cooperation can be realized some non-cooperative games. However, they admitted that whether a certain game converges or not surely depends on the domain of rationalities considered. In this paper we extended the set of rationalities by including equilibria from game theory and social welfare function from social choice theory for investigating the potential of the approach for solving 5,625 different non-cooperative games. It has been found that, under an optimistic perspective, 100% of the games presented the convergency characteristic, which implicate that all general equilibrium selection problem can be solved by using the approach. Others results are discussed with regards to the pessimistic and the egalitarian perspective.

Keywords: convergence theory; equilibrium; game theory; social choice theory

1. Introduction

Convergence Theory (CT) (Suzuki & Horita, 2017; Suzuki & Horita, 2023) can demonstrate how voting rules (rationalities) are justified through the (possibly infinitely-long) meta-level arguments. Suzuki and Leoneti (2022) extended the idea of CT to the so-called general meta-level equilibrium selection problem. The CT approach on the meta-level equilibrium selection problem tries to see when and how such a regression can reach a level at which every rationality (equilibria) in the domain agrees on its ultimate outcomes, which is called a convergence. In other words, a convergence is a situation where no matter which rationality players select as their principle of behaviour at some level, the resulting outcome is the same.

Suzuki and Leoneti (2022) demonstrated that this new approach can be used for explaining how cooperation can be realized in some non-cooperative games. However, they admitted that whether a certain game converges or not surely depends on the domain of equilibria considered. For instance, within the Prisoner Dilemma's (PD) game, considering, e.g., the Nash equilibrium and GMR (General Meta Rationality) rationalities that return a subset of the set of strategies through the game, it begins with $\mathcal{PD}^0$ (the PD game in the usual sense) and it induces the next level of the PD game, $\mathcal{PD}^1$, which is a 2x2 non-cooperative metagame in which players should choose which of Nash or GMR when they apply to select their actions in $\mathcal{PD}^0$. Since $\mathcal{PD}^1$ is itself a game, it requires the next-level game $\mathcal{PD}^2$, and thus it can go on *ad infinitum*.

It has been found that the game converges to the cooperative solution at level game $\mathcal{PD}^2$ when players adopt an optimistic perspective (the maximum among the payoffs when more than one equilibrium is found when moving from the level $\mathcal{PD}^i$ to the level $\mathcal{PD}^{i+1}$). When players choose the pessimistic perspective (the minimum among the payoffs when more than one equilibrium is found when moving from the level $\mathcal{PD}^i$ to the level $\mathcal{PD}^{i+1}$), it was found that the game converges to the non-cooperative solution at level game $\mathcal{PD}^{i}$. Finally, the game does not present the convergence feature on the view of the egalitarian perspective, where the players expect the average payoffs among the equilibria found (Leoneti & Suzuki, 2023).

In this paper we extend the PD's convergence analysis for different games for verifying the extent of the approach for solving the general equilibrium selection problem.

2. Method

By means of a utility with ordinal property, it has been designed all the possible games formed by the combinations of payoffs within the structure of a 2×2 non-cooperative game, totalling 5,625 possible games (Fraser & Kilgour, 1986; Fishburn & Kilgour, 1990) . For each of these games, the occurrence of convergence was investigated by using a set of seven rationalities, including MaxiMin, Nash, GMR, SMR (Symetric Meta Rationality), SEQ (Sequential Meta Rationality), from game theory approach, and Weak Pareto, and Rawls's Maxi-Min from social choice perspective, which received the following identification: (1) MaxiMin, (2) Nash, (3) GMR, (4) SMR, (5) SEQ, (6) Weak Pareto, and (7) Rawls's Maxi-Min. Additionally, it had been created an identification (ID) to the games for a precise mapping. In this sense, the ID is formed by two numbers followed by eight numbers, as "11-11111111", where the two first numbers refer to the rationalities that can be adopted by the players and the remaining numbers are the payoffs for each player for each state. For example, for the PD game when players are using the rationalities (2) Nash and (3) GMR, we have the identification "23-14223341", which is correspondent to the Table 1.

Table 1. Prisoner Dilemma game

		Player 2	
		a	b
Player 1	a	(1,4)	(3,3)
	b	(2,2)	(4,1)

In this sense, the ID of the game is the concatenation of the payoffs with the set of rationalities. According to Leoneti and Suzuki (2023), the approach is restricted to pairwise choices for the sake of simplicity, since it would resemble the original 2x2 non-cooperative game's framework and it would make possible to represent the game graphically following the graph representation by Shubik (1970).

Following, a spreadsheet was created for calculating the results based on the combination of each pair of rationalities. That calculation is required for the generation of the meta-level structures based on a given pair of rationalities. It should be noted that there are four possible combinations to be tested for every pair of rationalities. Then, the possible four combinations should be aggregated by means an aggregation procedure that is necessary for the investigation on the optimistic, pessimistic, and mean-valued perspective for each of the games. Through the aggregation, the immediate superior level of the game can be formed.

For instance, taking the optimistic perspective, and considering the set of Nash and GMR, the superior level of the game "23-14223341" would be formed by the best payoffs of the respectives states {Nash,Nash}, {Nash,GMR}, {GMR,Nash}, {GMR.GMR}, as presented in Figure 1.

Therefore, in Figure 1 we consider a level-1 game, in which players select their rationality to be applied for their choice of strategy in the level-0 game, i.e., Prisoner's Dilemma game (Table 1). For instance, Nash×GMR means that a player is operating under the rationality of Nash equilibrium while the counterpart is operating under the GMR equilibrium rationality. It can be seen that the state b/a would be the only meta-equilibrium (an equilibrium formed by the joint application of the rationalities Nash×GMR) of the level-0 game, resulting in the payoffs (2,2). Therefore, for the level-1 game, this would be the payoff to the state Nash×GMR.

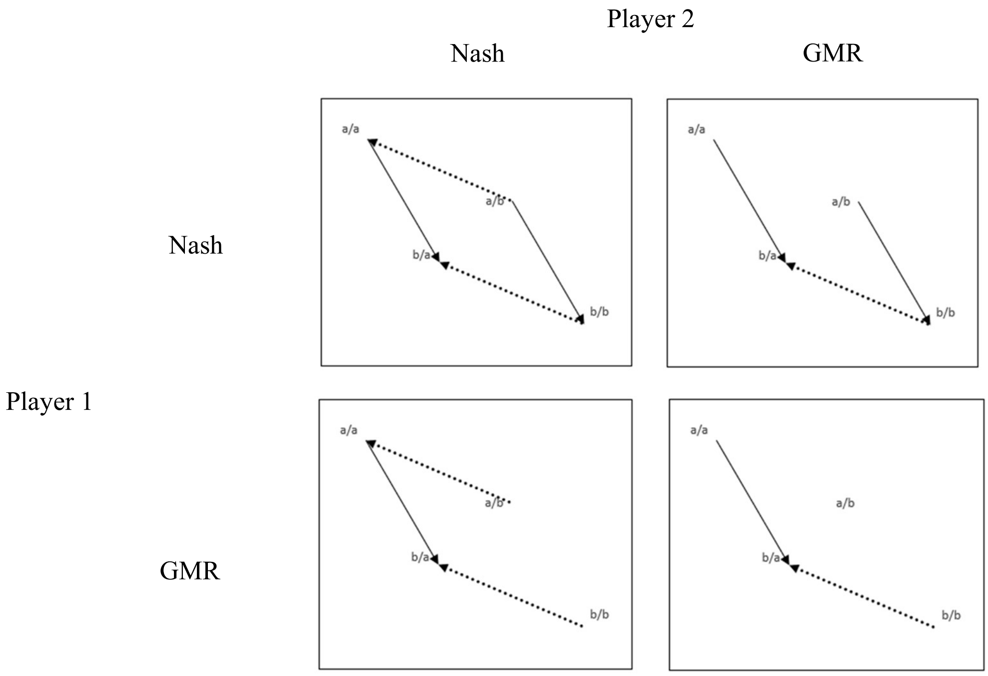


Figure 1: Structure of the level-1 meta game where the continuous line is the movement of Player 1 in level-0 while the dotted line is the movement of Player 2 in level-0

In the case of more than one Nash equilibrium, translating Figure 1 to a payoff matrix requires some assumptions on players' attitude. In this study, we consider three possibilities (assumptions): optimistic, pessimistic, and mean-valued (egalitarian). The optimistic perspective considers that players expect the best social outcome, based on the solutions to occur in the level-0. The pessimistic perspective considers that players expect the worst of the solutions to occur in the level-0, and the mean-valued perspective considers that players expect the average payoffs among the solutions found. Therefore, if, hypothetically, the state a/b was also a meta-equilibrium in the joint application of Nash×GMR, resulting in the payoffs (3,3), then: (i) from the optimistic perspective, the payoff to the state Nash×GMR would be (3,3), (ii) from the pessimistic perspective, the payoff to the state Nash×GMR would be (2,2), and (iii) from the egalitarian perspective, the payoff to the state Nash×GMR would be (2.5,2.5). Finally, to interpret this level-1 game as a 2×2 game, the payoffs are ordered and the order is used as the payoff. Level-3, 4, and upper-level games are constructed in the same way.

Considering 21 possible combinations among the seven rationalities used and the 5,625 possible games, it would be necessary to verify 118,125 possibilities for each of these perspectives in order to generate each level of the meta-structure. In order to decrease the necessary computation time, it was used a mapping of the games. The mapping concerns on a redundancy search based on the first level generated. For instance, consider that, through the optimistic perspective, the game "23-14223341" at level-0 became "23-11111122" at level-1. Therefore, for level-2 and the other upper levels, it was searched the game of level-1 (the input for the level-2) in the domain of level-0, then it was registered what would be the game at level-1 (under the same rationality) for representing what it would be the level-2 and other upper-level games.

The results were gathered in a new spreadsheet, where the pair of equilibria are fixed throughout the levels, and a single row represents a situation in which two players select a specific pair of rationalities at the game with a specific payoff matrix without specifying the levels. So, this row can be applied to any level with the same situation following the aforementioned procedure.

Finally, for verifying if a menu (set of feasible equilibria) converges or not; or a preference profile (one game) converges under a menu or not, it was used statistics analysis. The results were discussed under the perspective of a universal solution to the general equilibrium selection problem.

3. Results and discussion

From the results, it was found that, on the view of the optimistic perspective, 78,545 games from a total of 118,125 presented convergency at level-1, representing about 66% of the games. The number of convergent games increased to 117,317 (99%) at level-2 and to 118,125 at level 3, representing 100% of the games. From the optimistic perspective, it is interesting to note that the games Called Bluff, Cuban Missile Crisis, Protector 2 and 4, Ideological Hegemony, Cycle, Inspector-Evader, Total conflict, Alibi, and Pursuit of the Israelites, from the periodic table of Robinson & Goforth (2005), were among the ones that did not converge at level-2. However, from these games, on the classification of Robinson & Goforth (2005), none were found on the layer 3, which are the games initiated by the number 3 and that are mostly categorized as common interest games. In this sense, it is particularly interesting that all the games classified as the type of pure common interest presented convergence before the level 3.

On the view of the pessimistic perspective, 54,589 games from a total of 118,125 presented convergency at level-1, representing about 46% of the games. The result is significantly lower that from the optimistic perspective. The number of convergent games increased to only 86,001 (73%) at level-2, to 88,259 (75%) at level 3, to 88,387 (75%) at level 4, and keep the same 88,387 games at level 5, showing a cycle behavior that is persistent on upper-levels, indicating that there are some games that does not converge. From the periodic table of Robinson & Goforth (2005), the games Ideological Hegemony 1 and 2, Hamlet and Claudius, Blackmailer A, Cuban Missile Crisis, Cycle, Big Bully, Pursuit of the Israelites, Total Conflict, Inspector-Evader, Samson and Delilah, Vietnam, Hegemonic Stability, Alibi, Coordination, Asym. Coordination, Blackmailer B, Vietnam Bombing, Anti-Prisoner's Dilemma, Protector 1, 2, 3 and 4, Bully, Anti-Chicken, Battle of the Sexes, Asym. Battle of the Sexes, Called Bluff and Prisoner's Dilemma. It is interesting to notice that, among the games that does not present convergence, there is the Prisoner's Dilemma game. That result differs from the one presented by Leoneti & Suzuki (2023) because now it is also considered the social choice rationalities, which were the responsible for the non-convergence. This is an important result, since the convergence could be measured by its strongness, which would be the number of convergences that are attained under different types of rationalities.

Finally, on the view of the mean-valued perspective, only 44,641 games from a total of 118,125 presented convergency at level-1, representing about 38% of the games. The other levels are presented in the Table 2.

Table 2. Convergency of games under the mean-valued perspective

	Convergent	Relative
Level-1	44,641	38%
Level-2	81,345	69%
Level 3	85,651	73%
Level 4	86,039	73%
Level 5	86,047	73%
Level 6	86,047	73%

It can be noticed that the same peculiarities of non-convergency on the pessimistic perspective also occurred on the view of the egalitarian perspective. The games that did not converge under mean-valued perspective are relatively similar to the ones that did not converge on the pessimistic perspective, but now with the inclusion of the Stag Hunt, Hero and Chicken games and exclusion of the Battle of the Sexes game.

4. Conclusions

It has been found that, under an optimistic perspective, the meta-level approach is able to solve all non-cooperative games at up to three levels on the meta-game structure. It had been demonstrated, hence, a robust capacity for solving conflicts that are usually modelled be means of game theory. The result is also sound since it considers, under the set of rationalities, those from the social choice theory. Consequently, if a society has as premise an optimistic perspective, the proposed approach would be effective for supporting their choice.

It should be noticed that on the pessimistic and egalitarian perspectives, high level of convergence was also noticed, which implies in great capacity of the meta-level approach for solving the general equilibrium selection problem under these scenarios. This corroborate previous results that, on the pessimistic and egalitarian perspectives, myopic behaviour can easily prevail in society and could explain, why in a predominant pessimistic society, the myopic rationality is always the convergent state. Future research is suggested for further investigated the meta-level approach under these two perspectives.

References

Fishburn, P. C., & Kilgour, D. M. (1990). Binary 2× 2 games. Theory and Decision, 29, 165-182.

Fraser, N. M., & Kilgour, D. M. (1986). Non-strict ordinal 2× 2 games: A comprehensive computer-assisted analysis of the 726 possibilities. Theory and Decision, 20, 99-121.

Leoneti, A. B.; Suzuki, T. (2023) The general equilibrium selection problem in the Prisoner Dilemma game. In: 23rd International Conference on Group Decision and Negotiation, 2023, Tokyo. Proceedings of 23rd International Conference on Group Decision and Negotiatio.

Robinson D. & Goforth D. (2005). The Topology of the 2×2 Games: A New Periodic Table. Routledge, New York, NY

Shubik, M. (1970) Game theory, behavior, and the paradox of the prisoner's dilemma: three solutions. J. Conflict Resolut. 14(2), 181–193

Suzuki, T. ; Leoneti, A. B. (2022) Meta Level Equilibrium Selection for Two Illustrative Noncooperative Games. Group Decision and Negotiation. 1ed.: Springer, v. , p. 1-.

Suzuki, T., & Horita, M. (2017). Plurality, borda count, or anti-plurality: regress convergence phenomenon in the procedural choice. In Group Decision and Negotiation. Theory, Empirical Evidence, and Application: 16th International Conference, GDN 2016, Bellingham, WA, USA, June 20-24, 2016, Revised Selected Papers 16 (pp. 43-56). Springer International Publishing.

Suzuki, T., & Horita, M. (2023). A Society Can Always Decide How to Decide: A Proof. Group Decision and Negotiation, 1-37.

Alternative Limited Motion Stability within the Graph Model for Conflict Resolution Involving Multiple Decision-Makers

Alecio S. Silva [1], Giannini I. A. Vieira [2] and Leandro C. Rêgo [3, 4*]

[1] Graduate Program in Statistics, Universidade Federal de Pernambuco, Recife, PE, 50670-901, Brazil
[2] Crateús Campus, Universidade Federal do Ceará, Crateús 63700-000, Brazil
[3] Statistics and Applied Math Department, Universidade Federal do Ceará, Fortaleza, CE, 60455-760, Brazil
[4] Graduate Programs in Statistics and Management Engineering, Universidade Federal de Pernambuco, Recife 50670-901, Brazil
*Corresponding author, e-mail: leandro@dema.ufc.br

Abstract

In the literature on the Graph Model for Conflict Resolution, stability concepts hold a special prominence, serving as the tools employed in analyses seeking solutions to conflicts. Among these concepts, some accommodate a variable horizon, including the Limited Move Stability L_h. There are two definitions of this concept for multilateral conflicts depending on whether the focal DM is allowed to move again after his initial move. However, as illustrated through an example, the case where the focal DM moves only once is not indeed a generalization of the L_h originally proposed for bilateral conflicts. Furthermore, these generalizations assume that each opponent of the focal decision-maker acts independently, which differs from the approach of most established stability concepts in the literature. Typically, these concepts consider opponents as forming a coalition. In this article, our goal is to propose a generalization of L_h stability for multilateral conflicts, where we regard the opponents of the focal decision-maker as a coalition.

Keywords: L_h stability; GMCR; conflict resolutions.

1. Introduction

There is great importance in understanding conflicts as social phenomena that contribute to the development of a society in social, cultural, political, and scientific realms. Conflicts arise when involved parties seek to address specific preferences, highlighting the strategic nature of these situations. In order to comprehend and address human behaviors in conflicts, stability analysis is introduced, focusing on the Graph Model for Conflict Resolution (GMCR).

The GMCR incorporates concepts from Game Theory and Graph Theory, providing different approaches to analyze stability in conflicts. It serves as a mathematical tool to model, describe, and analyze strategic conflicts based on the preferences of decision-makers (DMs) involved in the conflict. The set of all these DMs is denoted by N. In GMCR, the set of all possible states that can arise in a conflict according to the actions that can be taken by the DMs is denoted by $S = \{s, s_1, s_2, \dots s_k\}$. There exists a collection of directed graphs, one for each DM, where each graph $D_i = (S, A_i)$, $A_i \subset S \times S$, determines, for each state $s \in S$, the states to which DM i can move the conflict. These states are referred to as accessible states for DM i in a single step from s and form the set $R_i(s)$. For any nonempty set of DMs, H, let $R_H(s)$ be the set of all states that can be reached by a sequence of moves of DMs in H, where no DM is allowed to move twice one right after the other.

The decision of a DM to move the conflict from one state to another is determined by his preferences between these states. These preferences are represented by a binary relation, assumed in this work to be complete, transitive, and asymmetric, and denoted by $\succ_i$. If $s_1 \succ_i s_2$, it represents that DM i prefers state s_1

over state s_2. This relation can be used to derive the weak preference relation, $\succcurlyeq_i$, where $s_1 \succcurlyeq_i s_2$ indicates that $s_2 \nsucc_i s_1$. The set of states that are reachable in a single step and preferable to s by DM i is denoted by $R_i^+(s)$.

In GMCR, several stability concepts have been established. Some of them have fixed horizons, such as Nash stability Nash (1950), General Metarational Stability (GMR) Howard (2003), Sequential Stability (SEQ) Fraser & Hipel (1979), Symmetric Metarational Stability (SMR) Howard (2003), and Symmetric Sequential Stability (SSEQ) Rêgo & Vieira (2017), while others consider a variable horizon h, including Limited Move Stability L_h Kilgour (1985), Generalized Metarational Stability MR_h Zeng et al. (2006), and Maximin$_h$ Stability Rêgo & Vieira (2020).

Here, we revisit Limited Move Stability L_h as an approach that considers a variable horizon h. This stability is fundamental for understanding DMs behavior in conflict situations. The L_h concept has two generalizations, known as 'Case 1' and 'Case 2', proposed by Fang et al. (1993) for conflicts with multiple decision-makers. However, in this article, we highlight limitations in the generalizations of the L_h concept for multilateral conflicts, indicating that these generalizations do not adequately treat opponents as a coalition. In this context, the article aims to present an alternative generalization for the L_h concept in conflicts with multiple DMs.

The paper is organized as follows. Section 2 brings the necessary background for understanding this paper. Section 3 provides the formal definition of the proposed stability concept. In Section 4, we illustrate the proposed concept by means of an example. Finally, the final conclusions are presented in Section 5.

2. Background

In this section, we recall some background information necessary to the understanding of this paper.

2.1 L_h stability

In Limited Move Stability L_h (Kilgour et. al (1985)), every time DMs move in a conflict, they can either stay in the current state and end the conflict or move to some accessible state giving the chance for another DM to move. In this concept, the number of movements made by DMs, their actions and reactions, is limited by h, which is the horizon of the conflict. Thus, DMs anticipate the conflict horizon by h steps, considering possible future changes that other DMs can make to reach states according to their preferences, and preferences are assumed to be common knowledge among DMs.

To understand L_h concept, we now recall is definition for bilateral conflicts. Let $K_i(s)$ be the number of states that are worse than s for DM i, i.e., $K_i(s) = ||\{s_1 : s \succ_i s_1\}||$, and let h be a positive integer. The state that DM i anticipates from state s when i moves in s and the horizon is h is defined as $G_h(i,s)$, as follows:

$$\mathrm{G_h}(i,s) = \begin{cases} s, \text{ if } R_i(s) = \emptyset,\ K_i(s) \geqslant \mathrm{A_h}(i,s) \\ \mathrm{G_{h-1}}\big(j, M_h(i,s)\big), \text{ otherwise,} \end{cases}$$

where $M_h(i,s)$ is some state $s_1^* \in R_i(s)$ that satisfies $K_i(G_{h-1}(j,s_1^*)) = max\{K_i(G_{h-1}(j,s_1)) : s_1 \in R_i(s)\}$, $j \neq i$, $A_h(i,s) = K_i(G_{h-1}(j, M_h(i,s)))$, and $G_0(\cdot,s) = s$.

Thus, Limited Move Stability L_h is defined as follows:

Definition 1: For $i \in N$, a state s $\in S$ is L_h stable for DM $i \in N$ iff $G_h(i,s) = s$.

For conflicts involving multiple DMs, Fang et al. (1993) presents two possible generalizations of Limited Move Stability Kilgour (1985). In one of them, called (Case 1), it is allowed for the focal DM to be part of the sequence of movements occurring after his initial move, meaning he can participate in the sanction immediately following his first move. On the other hand, in (Case 2), the focal DM is not part of the sequence of movements occurring after his initial move. Silva (2023) details these concepts contributing to a deeper understanding of them.

2.2 Pareto Dominance

In the study of multi-objective optimization, an important concept is that of a Pareto set or Pareto-optimal set. As many objective functions are involved in these problems, some of which need to be maximized and/or minimized, it is usually not possible to improve one aspect of the problem without worsening another. For example, in a coalition of DMs involved in a conflict, there may be DMs with significantly diverse preferences, thus characterizing a conflict of interest among the DMs belonging to the coalition. A solution belongs to the Pareto-optimal set if there is no other solution that never performs worse than it in some objective and performs better in at least one of them.

Next, we adapt the definition of Pareto dominance, presented in Pareto (1986), for GMCR. This concept will be used later to define an alternative stability notion for L_h stability in conflicts with multiple DMs.

Definition 2: For a coalition of DMs H, a state $s^* \in S$ is Pareto dominated by a state $s' \in S$ for DMs in H if and only if $s' \succ_i s^*$ for some DM $i \in H$ and $s' \succcurlyeq_i s^*$ for every DM $j \in H$.

3. Alternative Limited Move Stability, L_h^{CC}

It is crucial to emphasize that the DMs comprising the coalition opposing the focal DM may have distinct preferences. Although a criterion may be used to aggregate these preferences into a single one, nothing guarantees that every DM in the coalition will be satisfied with the aggregate preferences, which could cause a deviation from a state prescribed by the aggregate preference. To avoid this problem, in our approach, we do not use an aggregate preference for the coalition. Instead, we assume that the coalition's improvement movement will be directed towards an achievable state, which is not a Pareto-dominated state by another state for the coalition of opponents of the focal DM.

To present an alternative stability concept to the notion of stability L_h, we need to define some sets. Let $N_i = N - \{i\}$ and $O_h^{CC}(i,s)$ be the state anticipated by the focal DM when moving at state s, with a conflict horizon of h, and believing that the conflict will follow according to the notion of stability L_h^{CC}. We formally define this in Equations 1 and 2. Also, let $R_h(N_i,s) = \{s\} \cup \{O_{h-1}^{CC}(i,s') : s' \in R_{N_i}(s)\}$ be the set of all possible final anticipated states by DM i when his opponents move first from state s and that they either stay at s or move to any reachable state, alternating moves from DM i and sequence of moves of his opponents, considering an horizon$h$. The intuition is that the opponents should move to a state anticipating ending the conflict in the best state in $R_h(N_i,s)$. As the opponents do not share the same preference relation, they may agree on any state which is not Pareto dominated in $R_h(N_i,s)$. Formally, let

$$R_h^{CC}(N_i,s) = \begin{cases} \{s\} \cup \{s' \in R_{N_i}(s) : O_{h-1}^{CC}(i,s') \textit{ is not Pareto dominated in the set } R_h(N_i,s) \\ \textit{for coalition } N_i\} \textit{ if } s \textit{ is not Pareto dominated in the set } R_h(N_i,s) \textit{ for coalition } N_i, \\ \\ \{s' \in R_{N_i}(s) : O_{h-1}^{CC}(i,s') \textit{ is not Pareto dominated in the set } R_h(N_i,s) \\ \textit{for coalition } N_i\}, \textit{otherwise.} \end{cases}$$

Therefore, the states belonging to $R_h^{CC}(N_i,s)$ are all states that can be characterized as the final state of a legal sequence of movements, and the anticipated state from them is not Pareto dominated within the set $R_h(N_i,s)$ for the coalition N_i. Additionally, $R_h^{CC}(N_i,s)$ includes the state s if it is not Pareto dominated for coalition N_i in the set $R_h(N_i,s)$. Thus, we define the anticipated state $O_h^{CC}(i,s)$, where, by convention, $O_0^{CC}(\cdot,s) = s$, as follows:

$$O_h^{CC}(i,s) = \begin{cases} s, \textit{if } R_i(s) = \emptyset, \\ s, \textit{if } K_i(s) \geq A_h^{CC}(i,s), \\ O_{h-1}^{CC}\left(N_i, Q_h^{CC}(i,s)\right), \textit{otherwise} \end{cases} \qquad (1)$$

Where $Q_h^{CC}(i,s)$ is some state $s^* \in R_i(s)$ such that $K_i(O_{h-1}^{CC}(N_i,s^*)) = max\{K_i(O_{h-1}^{CC}(N_i,s')) : s' \in R_i(s)\}$ and $A_h^{CC}(i,s) = K_i(O_{h-1}^{CC}(N_i,Q_h^{CC}(i,s)))$. Intuitively, the focal DM always moves trying to reach the best final anticipated state for him. In turn, $O_h^{CC}(N_i,s)$ is given by:

$$O_h^{CC}(N_i,s)=\begin{cases} s, if\ R_{N_i}(s)=\emptyset\ or\ R_h^{CC}(N_i,s)=\{s\}, \\ s, if\ K_i(s)\le A_h^{CC}(N_i,s)\ and\ s\in R_h^{CC}(N_i,s), \\ O_{h-1}^{CC}\left(i,Q_h^{CC}(N_i,s)\right), otherwise \end{cases} \quad (2)$$

Where $Q_h^{CC}(N_i,s)$ is some state $s^*\in R_h^{CC}(N_i,s)$ such that $K_i(O_{h-1}^{CC}(i,s^*))=min\{K_i\left(O_{h-1}^{CC}(i,s')\right):s'\in R_h^{CC}(N_i,s)\}$ and $A_h^{CC}(N_i,s)=K_i(O_{h-1}^{CC}(i,Q_h^{CC}(N_i,s)))$. Intuitively, the focal DM anticipates that the coalition of opponents will move to the worst final anticipated state for him, among those that are not Pareto dominated for the coalition of opponents in the set of all possible final anticipated states.

In Definition 3, we present the Alternative Limited Move Stability with Cooperative Coalition L_h^{CC} for conflicts involving multiple DMs.

Definition 3: For $\boldsymbol{i}\in\ \boldsymbol{N}$, state $\boldsymbol{s}\in\ \boldsymbol{S}$ is $\boldsymbol{L_h^{CC}}$ stable for DM $\boldsymbol{i}$ if and only if $\boldsymbol{O_h^{CC}(i,s)=s}$.

4. The relationship between the L_h and L_h^{CC} stabilities in bilateral conflicts

We now prove that in the particular case when only two DMs howare involved in the conflict, the alternative L_h^{CC} stability coincides with the standard L_h stability for bilateral conflicts. To prove this result, Lemma 1 establishes an equivalence between the anticipated state regarding the L_h concept for two DM's with a horizon of h, denoted as $G_h(\cdot,s)$, and the anticipated state regarding the L_h^{CC} concept for two DMs with a horizon of h, denoted as $O_h^{CC}(\cdot,s)$. This result will be used to demonstrate that the L_h^{CC} concept generalizes the L_h concept for conflicts with multiple DM's.

Lemma 1: For $N=\{i,j\}$, $s\in\ S$, and a positive integer h, we have that $G_h(i,s)=O_h^{CC}(i,s)$ and $G_h(j,s)=O_h^{CC}(j,s)$.

Theorem 1 establishes that the Alternative Limited Move Stability, L_h^{CC}, is a generalization of the Limited Move Stability proposed in Kilgour (1985) for conflicts with multiple DMs.

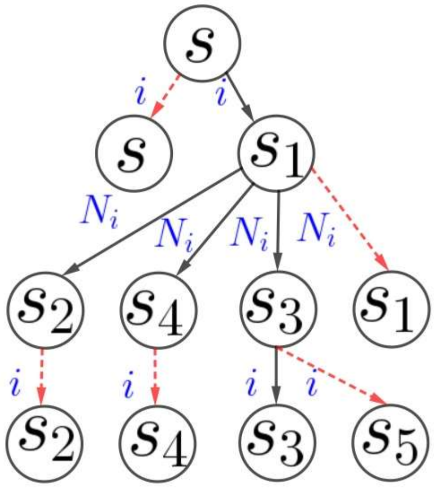
Theorem 1: For a conflict involving two DMs, state s is L_h stable for DM i if and only if s is L_h^{CC} stable for DMi.

5. Application

In this section, we present a brief hypothetical example, originally introduced by Silva (2023), with the aim of clarifying how stability is achieved through this new concept. Example 1 illustrates a hypothetical conflict involving six states $\{s,s_1,s_2,s_3,s_4,s_5\}$ and three DMs i,j, and k.

Example 1: Consider a scenario representing a hypothetical conflict involving three DMs, denoted as i,j, and k, and six states $\{s,s_1,s_2,s_3,s_4,s_5\}$. Assume that the accessibility sets for i,j and k are represented, respectively, by: $R_i(s)=\{s_1\},R_i(s_3)=\{s_5\},R_j(s_1)=\{s_2\},R_j(s_3)=\{s_4\},R_k(s_2)=\{s_3\}$, and the accessibility sets for all DMs in states not previously shown are empty. The preference relations are defined as follows: $s_5\succ_i s_4\succ_i s\succ_i s_3\succ_i s_2\succ_i s_1$; $s_3\succ_j s\succ_j s_1\succ_j s_5\succ_j s_2\succ_j s_4$; and $s_2\succ_k s_1\succ_k s\succ_k s_3\succ_k s_4\succ_k s_5$. Figure 1 illustrates how the conflict can unfold from state s with an initial move by DM i, considering the L_3^{CC} stability.

L_3^{CC} stability analysis for DM i, starting from state s, in the conflict described in Example 1



Source: Silva (2023).

Analyzing the L_3^{CC} stability of state s for DM i in the conflict presented in Example 1. According to it, from state s, DM i needs to decide whether to stay in state s or move to state s_1. Therefore, it is necessary to determine $O_2^{CC}(N_i, s_1)$. To do so, we need to determine $O_1^{CC}(i, s')$ for every $s' \in R_{N_i}(s_1) = \{s_2, s_3, s_4\}$. Since DM i cannot move from s_2 or s_4, we have $O_1^{CC}(i, s_2) = s_2$ and $O_1^{CC}(i, s_4) = s_4$. Finally, as $R_i(s_3) = \{s_5\}$ and $s_5 \succ_i s_3$, we have $O_1^{CC}(i, s_3) = s_5$. Therefore, $R_2(N_i, s_1) = \{s_1, s_2, s_4, s_5\}$. Since states s_4 and s_5 are Pareto dominated by state s_1 for coalition N_i, we have $R_2^{CC}(i, s_1) = \{s_1, s_2\}$. Since $K_i(s_1) < K_i(s_2)$, we have $O_2^{CC}(N_i, s_1) = s_1$, meaning that among all possible final states for the conflict, s_1 is the non-Pareto dominated state for coalition N_i that minimizes the payoff for DM i. Thus, since $s \succ_i s_1$, at state s, DM i has no incentive to move the conflict from state s to state s_1, i.e., $O_3^{CC}(i, s) = s$, so we conclude that state s is L_3^{CC} stable for DM i.

6. Conclusions

In this article, we present an alternative definition for Limited Move Stability in multilateral conflicts, denoted as L_h^{CC}. This definition has the characteristic of treating the opponents of the focal DM as a coalition, unlike the other Limited Move stabilities (Case 1) and (Case 2) existing for multilateral conflicts proposed in Fang (1993).

To better understand the properties of our proposed concept, we show that the alternative definition L_h^{CC} coincides with the Limited Move Stability proposed in Kilgour (1985) in the case of bilateral conflicts. In future work, we intend to analyze the relation between L_h^{CC} and other stability concepts.

Acknowledgments. The authors would like to acknowledge the financial support of Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ) under grants 406697/2023-0 and 308980/2021-2.

References

L. Fang, K. W. Hipel, and D. M. Kilgour. Interactive Decision Making: The Graph Model for Conflict Resolution, volume 3. Wiley, 1993.

N. M. Fraser and K. W. Hipel. Solving complex conflicts. IEEE Trans. Syst., Man, Cybern, 9(12):805–816, 1979.

N. Howard. Paradoxes of Rationality: Games, Metagames, and Political Behavior. Mit Press, 2003.

D. Marc Kilgour. Anticipation and stability in two-person noncooperative games. Dynamic models of international conflict, pages 26–51, 1985.

D. Marc Kilgour, K. W. Hipel, and L. Fang. The graph model for conflicts. Automatica, 23(1):41–55, 1987.

J. F. Nash. Equilibrium points in n-person games. Proceedings of the National Academy of Sciences-USA, 36(1):48–49, 1950.

V. Pareto. Cours d'economie politique. F. Rouge, 1896.

L. C. Rêgo and G. I. A. Vieira. Symmetric sequential stability in the graph model for conflict resolution with multiple decision makers. Group Decision and Negotiation, 26(4):775–792, 2017.

L. C. Rêgo and G. I. A. Vieira. Maximinh stability in the graph model for conflict resolution for bilateral conflicts. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 50(10):3760–3769, 2020

A. S. Silva. Um estudo sobre alguns conceitos de estabilidade com horizonte variável no modelo de grafos para resolução de conflitos. Doutorado em estatística, Universidade Federal de Pernambuco, Recife, 2023.

D. Z. Zeng, L. Fang, K. W. Hipel, and D. M. Kilgour. Generalized metarationalities in he graph model for conflict resolution. Discret Applied Matematics, 154(16):2430–2443, 2006.

Network Analysis of International Conflicts

Fuad Aleskerov [1,2], Anna Semenova [1*]

[1] National Research University Higher School of Economics (HSE), Moscow, Russia
[2] V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences (ICS RAS), Moscow, Russia
[*] amsemenova_4@edu.hse.ru

Abstract

We study international conflicts based on the construction of a network among countries involved, and take into account the intensity of an international conflict as well as the intensity of internal conflicts in the country. In the construction of the network we consider the influence of third countries on the conflict.
We use classic centrality indices as well as new ones considering the parameters of the nodes and group influence of nodes on a single vertex.

Keywords: international conflicts; internal conflicts; intensity of conflicts; network centrality indices; group influence in conflicts networks

1. Introduction

The study of conflicts is a permanent topic of international research. It can also be called basic. For centuries and in our time, international conflicts are one of the most important components of international relations study including studies of democracy and democratization, state capacity, state power (Watts et al., 2017).

But the structure of international relations is incredibly complex, tracking individual relationships and trying to analyze them comprehensively requires a lot of effort. No matter what the subject of research is – the conflict itself, a set of conflicts or events that conflicts can indirectly influence understanding of the structure of conflicts is important for comparative studies. Comprehensive analysis of international conflicts requires a special analytical tool. And most importantly, it will allow not to miss unobvious relationships.

Network analysis becomes increasingly popular and useful in studies of international relations (Maoz, 2012). Research employing this methodology encompasses various fields of study such as distribution of influence in international organizations, transnational economic relations, causes of tie formation in global politics as well as causes of war and peace (Maoz et al., 2003). Network approach enables to avoid studying separately either units or their interactions. It becomes possible to study units and their interactions simultaneously. This promise of network approach is of paramount importance for research on international relations which have been always regarded either from unit (agent) level or structure level.

In this work we attempt to apply network approach to the study of armed conflicts and assessing the involvement of countries in international conflicts, taking into account internal conflicts on their territory.

We present a new approach to study international conflicts based on the construction of a network among countries involved. We take into account the intensity of an international conflict as well as the intensity of internal conflicts in the country. Moreover, in the construction of the network we take into account third countries influence on the conflict.

To assess the involvement of countries in international conflicts we use not only classic centrality indices but also new ones taking into account the parameters of the nodes and group influence of vertices. Using these new indices, we can analyze the impact and involvement of the countries in the international conflicts

(Aleskerov & Yakuba, 2020). The model has applied to the study of international conflicts for different time periods after the World War II.

It should be noted that this is not the first attempt to build such a network of international conflicts. In 2016 we proposed the model of a network of the interstate conflicts from 1946 to 2015. We found the most influential states during certain time periods by means of two classical centrality indices (Eigenvector and Page Rank) and a new index of centrality (Short-Range Interaction Centrality) (Aleskerov et al., 2016). Additionally, several years ago, in 2021, similar results but based on different centrality measures were presented on the Conference GDN – 2021. Important difference of these very results comparing with the previous ones is that we have taken into account the problems inside countries involved in a conflict, namely civil war, terrorist attacks, etc.

To create the model we used the Uppsala Conflict Data Program (UCDP) at the Department of Peace and Conflict Research, Uppsala University and the Centre for the Study of Civil War at the International Peace Research Institute in Oslo (PRIO) datasets: the "UCDP/PRIO Armed Conflict" (Gleditsch et al., 2002; Davies et al., 2023) and "UCDP Battle-Related Deaths" (Lacina & Gleditsch, 2005; Davies et al., 2023) datasets which provide information about armed conflicts from 1946 to 2022. UCDP defines state-based conflict as: "a contested incompatibility that concerns government and/or territory where the use of armed force between two parties, of which at least one is the government of a state, results in at least 25 battle-related deaths in a calendar year" (Davies et al., 2023).

We use three types of conflict: interstate armed conflict, internal armed conflict and internationalized internal armed conflict. To evaluate the intensity of the conflict we use information about the number of battle-related deaths in the conflict which was transformed to ordinal scale.

The model has applied to the study of conflicts for different time periods after the World War II. This paper presents the results obtained between 2011 and 2021.

2. International Conflict Network Simulation

The network structure of conflicts is a directed graph that models the interactions of states based on an armed conflict or support in an armed conflict in a certain time period, and takes into account the intensity of conflict, as well as the intensity of an internal conflict in states participating in an international conflict. The nodes of the graph are generally recognized states.

The edges of the graph are conflicts between generally recognized states. Support relationships in our model is also taken into account, but indirectly – as an indirect conflict between the supporting state and the opposite state. In addition, we assume that so-called supporters of conflicting states are in distant indirect conflict.

So, there are only three types of connection between nodes in our network: direct conflict (type 1), indirect conflict (type 2) and distant indirect conflict (type 3).

Suppose we have a conflict in which four states are involved: state A in direct conflict with state B, state C supports state A, state D supports state B.

The edges between nodes in our network will be formed for a conflict-year according to the following rules (Table 1):

- *Direct conflict:* direction on both sides edge is formed between states A and B for a conflict-year in which they were the main participants to the same conflict (from different sides);
- *Indirect conflict:* direction on both sides edge is formed between states C and B if state C is an ally of state A which is in conflict with a state B (both A and B are main participants), and the same for the state A and D;
- *Distant indirect conflict:* direction on both sides edge is formed between states C and D if state C is an ally of state A which is in conflict with a state B, and state D is an ally of state B (both A and B are main participants).

Table 1. Illustration of the three types of communication in the network

State	State	Type of conflict	Name of conflict
A	B	Type 1	Direct conflict
A	D	Type 2	Indirect conflict
B	C	Type 2	Indirect conflict
C	D	Type 3	Distant indirect conflict

So, if state A is in conflict with state B in a conflict-year, then there are edges AB and BA. There is a direct conflict between A and B, an indirect conflict between C and B (edges BC and CB), the same rule for B and D if D supports A. And there is also distant indirect conflict between states C and D (edges CD and DC).

All edges in the network structure are weighted by intensity of international and intrastate conflicts. To evaluate the intensity of a conflict we have used data on number of battle-related deaths, which are transformed into an ordinal scale.

In our model, the intensity of international conflict (Bd_{ext}) is measured on a scale from 1 to 13, where 1 – less than 100 battle related deaths, and 13 – more than 50,000 battle related deaths (Table 2).

Table 2. Intensity of international conflict by period

Bd_{ext}	Number of battle related deaths	Periods		
		2011 – 2017	**2018 – 2021**	
13	>50000	0	0	
12	>30000	0	0	
11	>20000	0	1	
10	>10000	0	0	
9	>7000	0	1	
8	>5000	1	1	
7	>3000	2	2	
6	>1500	2	2	
5	>1000	4	0	
4	>500	1	1	
3	>300	2	2	
2	>100	3	9	
1	<100	10	3	**All episodes**
	Total number of episodes (by period)	25	22	**47**

Similarly, the model includes information about the intensity of internal conflict within a country (Bd_{int}), which is measured on the scale from 1 to 5. A score of 1 indicates less than 100 battle-related deaths, while a score of 5 indicates more than 5,000 battle-related deaths (see Table 3).

Table 3. Intensity of internal conflict by period

Bd_{int}	Number of battle related deaths	Periods		
		2011 – 2017	**2018 – 2021**	
5	>5000	4	0	
4	>1000	14	0	
3	>500	14	11	
2	>100	77	42	
1	<100	93	57	**All episodes**
Total number of episodes (by period)		202	110	**312**

The intensity of conflicts in the model (Int_{conf}) is calculated using the obtained scales of international and internal conflict intensity, according to the following rules.

Intensity of conflict-based communication (Int_{conf}) equals to the intensity of international conflict (Bd_{ext}), measured on a scale from 1 to 13, added to the intensity of internal conflict (Bd_{int}) of the opposite side, measured on a scale from 1 to 5 (if there is an internal conflict present). In other words, the presence of internal conflict in a country increases the influence of other countries on that country in international conflicts.

It is assumed that in a conflict between two states, the impact of state A on state B will be greater if state B is experiencing internal conflict. This means that the presence of an internal conflict weakens the state's ability to withstand the impact of an interstate conflict, in proportion to the intensity of the internal conflict. The intensity of a conflict is calculated by including the presence of internal conflict, whether it is direct, indirect, or distant indirect.

To determine the intensity of indirect and distant indirect conflict, we introduce intensity correction into the model (Table 4). If the conflict is indirect, the intensity of the interaction is reduced by 2 levels. In the case of a distant indirect conflict, the intensity decreases by 4 levels. However, for conflicts between two or more countries, conflict intensity cannot be less than 1.

Table 4. Intensity correction by type of conflict

Type of conflict	Name of conflict	Intensity correction
Type 1	Direct conflict	No correction
Type 2	Indirect conflict	– 2 intensity level
Type 3	Distant indirect conflict	– 4 intensity level

It should be noted that each year of a conflict lasting several years is considered a separate episode. For example, if a conflict lasts for three years and has at least 25 victims each year, it will be recorded as three episodes in the dataset. Therefore, we calculated the average value for each conflicting dyad value over each period for both the intensity of an international conflict (Bd_{ext}) and the intensity of an internal conflict (Bd_{int}). After transforming this value to ordinal scale and considering intensity correction based on the type of conflict and the presence of internal conflict in the countries, we constructed the conflict network for each period.

The conflict relationship between 1946 and 2021 was divided into eight distinct periods. Period 1 – 1946-1953; Period 2 – 1954-1962; Period 3 – 1963-1979; Period 4 – 1980-1987; Period 5 – 1988-2000; Period 6 – 2001-2010; Period 7 – 2011-2017, Period 8 – 2018-2021. The periods are close to the stages of the Cold War before the collapse of the Soviet Union. Each period is characterized by one major conflict. This paper presents the results obtained during the period between 2011 and 2021, specifically for Period 7 and Period 8.

For each time period, we assess the most influential countries involved in the corresponding conflicts.

3. Network Analysis Methodology

We have a network of interstate conflicts, where the nodes of the network are generally recognized states. Arcs in the network represent conflicts (direct, indirect, indirectly distanced), the weight of an arc is the intensity of a conflict, which is determined by the number of battle-related deaths in both interstate and intrastate conflicts, as well as the intensity correction. Table 5 presents the descriptive statistics of conflict networks by period. The resulting conflict network is illustrated in Figure 1, dark grey nodes represent states with internal conflict.

Table 5. Descriptive statistics of conflict networks by period

Network	Nodes	Arcs	Density	Mean Weight	SD Weight
2011-2017	27	42	0.060	4.452	2.539
2018-2021	22	38	0.082	4.737	2.310

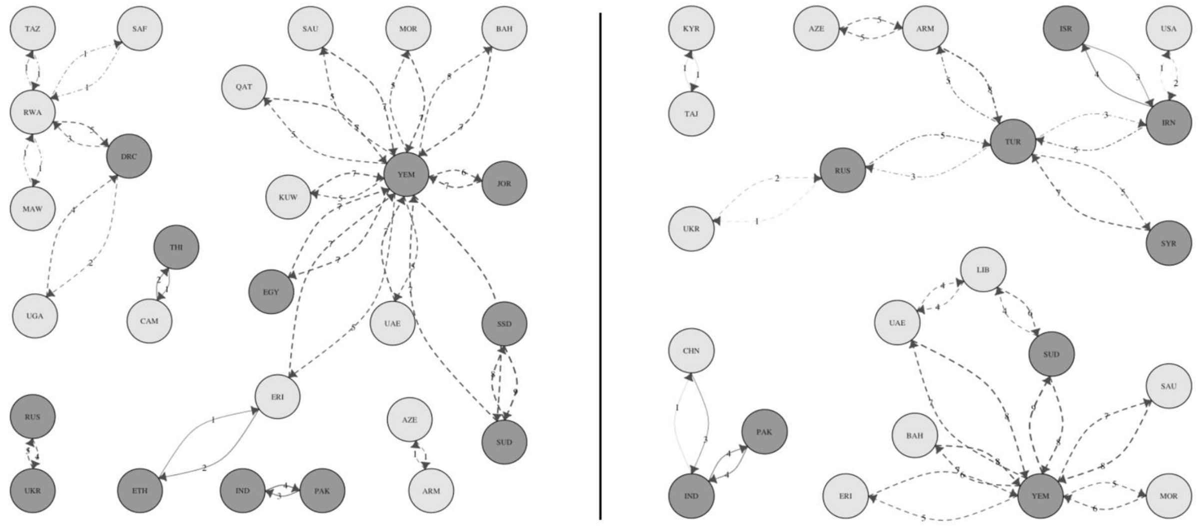


Figure 1: Networks of interstate conflicts for Period 7 (left) and Period 8 (right) (solid line – direct conflict, dashed line – indirect conflict, dotted dashed line – distant indirect conflict)

The goal of this research is to evaluate the involvement of each state in conflicts. To assess the involvement of countries in international conflicts we use classic centrality indices – degree, closeness, betweenness, eigenvector and Page Rank centrality measures (Freeman, 1979; Brin & Page, 1998).

We also consider new centrality indices that take into account the parameters of the nodes and group influence of vertices. Using these new indices, we can analyze the impact and involvement of the countries in the international conflicts. Group influence within a network can be modelled using the new Bundle and Pivotal centrality indices proposed in (Aleskerov & Yakuba, 2020). We apply these indices to the conflict network.

The state can participate in direct, indirect, or distant interstate conflicts with varying intensity levels. A critical level of interaction intensity, referred to as a quota, is set to the intensity level equals to 3. If the total intensity of interaction with a certain subgroup of states exceeds the quota, then that subgroup is considered critical. Maximum number of participants in a subgroup is set to 3.

The Bundle Index calculates the number of critical sets for each state and normalizes it across all states in the selected period. Additionally, some states may leave the critical sets, causing the set to no longer be critical. These states are referred to as pivotal. The proposed Pivotal index evaluates the number of pivotal states across all critical sets and normalizes the values among the states. To compare the results obtained using the new indices, we also included a classic centrality index, the Copeland index.

Using not only classic centrality indices but also new ones, we evaluate the participation of states in interstate conflicts, considering internal conflicts during the selected time periods.

4. Results

The network model represents conflict relations between generally recognized states from 2011 to 2021. The model is based on data regarding countries' participation in international conflicts, including internal conflicts on their territory, and information about battle-related deaths in both international and internal conflicts. The model reflects the connections between countries based on their participation in conflicts and the intensity of these ties. The table attached for both time periods shows the involvement of countries in conflicts, determined by eight centrality indices.

Appendix 1 presents the results of calculating the centrality measures for Period 7, sorted by degree centrality. According to the classical indices, Yemen, Rwanda, Sudan, Eritrea, and DR Congo are the most central actors in the conflict network. Egypt ranks third and Jordan ranks fourth in the eigenvector centrality values.

At the same time, if we look at the indices that take into account group interactions, the Bundle index that takes into account the critical groups in the network reveals that Rwanda is most involved country, followed by Yemen by a significant margin. When examining the Pivotal centrality index values, which consider the states that, upon exiting, cause the critical group to no longer be critical, it becomes apparent that, again, Yemen and Rwanda are involved in conflicts to the greatest extent.

Appendix 2 presents the centrality measures for Period 8, sorted by degree centrality. Based on the classical indices, Yemen, Turkey, Iran, Russia, and Armenia are the most central actors in the conflict network. In terms of eigenvector centrality values, Sudan ranks second and the UAE ranks fourth. The application of the new centrality indices reveals that Yemen and Sudan continue to be the most prominent, while Iran is increasing in significance.

Using new indexes that take into account group interactions allows us to assess the true involvement of countries in conflicts, including indirect participation and internal conflicts.

References

Aleskerov, F., & Yakuba, V. (2020). Matrix-Vector Approach to Construct Generalized Centrality Indices in Networks. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3597948

Aleskerov, F., Kurapova, M., Meshcheryakova, N., Mironyuk, M., & Shvydun, S. (2016). A network approach to analysis of international conflicts. *Political science (RU)*, (4), 111-137.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, *30*(1-7), 107-117. https://doi.org/10.1016/S0169-7552(98)00110-X

Davies, S., Pettersson, T., & Öberg, M. (2023). Organized violence 1989–2022, and the return of conflict between states. *Journal of Peace Research*, *60*(4), 691–708. https://doi.org/10.1177/00223433231185169

Freeman, L. C. (1979). Centrality in social networks: Conceptual clarification. *Social Networks 1: critical concepts in sociology. Londres: Routledge*, *1*, 238-263. https://doi.org/10.1016/0378-8733(78)90021-7

Gleditsch, N. P., Wallensteen, P., Eriksson, M., Sollenberg, M., & Strand, H. (2002). Armed Conflict 1946-2001: A New Dataset. *Journal of Peace Research*, *39*(5), 615–637. https://doi.org/10.1177/0022343302039005007

Lacina, B., & Gleditsch, N. P. (2005). Monitoring Trends in Global Combat: A New Dataset of Battle Deaths. *European Journal of Population*, *21*(2/3), 145–166. https://doi.org/10.1007/s10680-005-6851-6

Maoz, Z. (2012). How Network Analysis Can Inform the Study of International Relations. *Conflict Management* *and Peace Science*, *29*(3), 247–256. https://doi.org/10.1177/0738894212443341

Maoz, Z., Terris, L. G., Kuperman, R. D., & Talmud, I. (2007). What Is the Enemy of My Enemy? Causes and Consequences of Imbalanced International Relations, 1816–2001. *The Journal of Politics*, *69*(1), 100–115. https://doi.org/10.1111/j.1468-2508.2007.00497.x

Watts, S. (2017). *Understanding Conflict Trends: A Review of the Social Science Literature on the Causes of Conflict*.

Attachment 1

Classic Centrality Measures of Period 7 (2011-2017)

Country	Degree	Betweenness	Closeness	Eigenvector	Page Rank
Yemen	0,769	0,197	0,152	1,000	0,189
Rwanda	0,308	0,028	0,333	0,000	0,078
Sudan	0,154	0,034	0,086	0,484	0,054
Eritrea	0,154	0,034	0,082	0,250	0,029
DR Congo	0,154	0,012	0,294	0,000	0,073
Russia	0,077	0,000	0,250	0,000	0,037
Armenia	0,077	0,000	1,000	0,000	0,037
Uganda	0,077	0,000	0,143	0,000	0,030
Ethiopia	0,077	0,000	0,077	0,025	0,011
South Sudan	0,077	0,000	0,050	0,193	0,030
Morocco	0,077	0,000	0,076	0,249	0,020
Egupt	0,077	0,000	0,077	0,348	0,026
Jordan	0,077	0,000	0,076	0,298	0,023
Saudi Arabia	0,077	0,000	0,076	0,249	0,020
Kuwait	0,077	0,000	0,076	0,249	0,020

New Centrality Measures of Period 7 (2011-2017), q = 3, k = 3

Country	Copeland In index share	BI index share, s=3	PI index share, s=3	PI' index size share, s=3
Yemen	0,374	0,858	0,256	0,182
Sudan	0,091	0,015	0,051	0,036
DR Congo	0,048	0,015	0,051	0,036
South Sudan	0,043	0,005	0,026	0,018
Egypt	0,037	0,005	0,026	0,018
Rwanda	0,032	0,039	0,256	0,455
Eritrea	0,032	0,010	0,051	0,055
Jordan	0,032	0,005	0,026	0,018
Russia	0,027	0,005	0,026	0,018
Morocco	0,027	0,005	0,026	0,018
Saudi Arabia	0,027	0,005	0,026	0,018
Kuwait	0,027	0,005	0,026	0,018
Bahrain	0,027	0,005	0,026	0,018
Qatar	0,027	0,005	0,026	0,018
UAE	0,027	0,005	0,026	0,018

New Centrality Measures of Period 7 (2011-2017), q = 3, k = 5

Country	Copeland In index share	BI index share, s=5	PI index share, s=5	PI' index size share, s=5
Yemen	0,374	0,955	0,256	0,182
Sudan	0,091	0,004	0,051	0,036
DR Congo	0,048	0,004	0,051	0,036
South Sudan	0,043	0,001	0,026	0,018
Egypt	0,037	0,001	0,026	0,018
Rwanda	0,032	0,013	0,256	0,455
Eritrea	0,032	0,003	0,051	0,055
Jordan	0,032	0,001	0,026	0,018
Russia	0,027	0,001	0,026	0,018
Morocco	0,027	0,001	0,026	0,018
Saudi Arabia	0,027	0,001	0,026	0,018
Kuwait	0,027	0,001	0,026	0,018
Bahrain	0,027	0,001	0,026	0,018
Qatar	0,027	0,001	0,026	0,018
UAE	0,027	0,001	0,026	0,018

Attachment 2

Classic Centrality Measures of Period 8 (2018-2021)

Country	Degree	Betweenness	Closeness	Eigenvector	Page Rank
Yemen	0,571	0,086	0,137	1,000	0,142
Turkey	0,381	0,110	0,195	0,000	0,127
Iran	0,286	0,062	0,129	0,000	0,061
Russia	0,191	0,033	0,114	0,000	0,037
Armenia	0,191	0,033	0,092	0,000	0,064
Sudan	0,191	0,000	0,092	0,582	0,050
UAE	0,191	0,024	0,090	0,444	0,039
India	0,191	0,005	0,400	0,000	0,066
Libya	0,191	0,005	0,075	0,230	0,032
Ukraine	0,095	0,000	0,094	0,000	0,012
United States	0,095	0,000	0,104	0,000	0,012
Israel	0,095	0,000	0,098	0,000	0,027
Syria	0,095	0,000	0,087	0,000	0,041
Eritrea	0,095	0,000	0,080	0,280	0,022
Morocco	0,095	0,000	0,080	0,280	0,022

New Centrality Measures of Period 8 (2018-2021), q = 3, k = 3

Country	Copeland In index share	BI index share, s=3	PI index share, s=3	PI' index size share, s=3
Yemen	0,244	0,477	0,176	0,162
Turkey	0,139	0,163	0,118	0,108
Sudan	0,083	0,035	0,059	0,054
UAE	0,061	0,035	0,059	0,054
Armenia	0,056	0,035	0,059	0,054
Iran	0,044	0,070	0,118	0,162
Libya	0,044	0,035	0,059	0,054
Saudi Arabia	0,039	0,012	0,029	0,027
Bahrain	0,039	0,012	0,029	0,027
India	0,039	0,035	0,059	0,054
Russia	0,028	0,023	0,059	0,081
Syria	0,028	0,012	0,029	0,027
Eritrea	0,028	0,012	0,029	0,027
Morocco	0,028	0,012	0,029	0,027
Azerbaijan	0,028	0,012	0,029	0,027

New Centrality Measures of Period 8 (2018-2021), q = 3, k = 5

Country	Copeland In index share	BI index share, s=5	PI index share, s=5	PI' index size share, s=5
Yemen	0,244	0,574	0,176	0,162
Turkey	0,139	0,139	0,118	0,108
Sudan	0,083	0,028	0,059	0,054
UAE	0,061	0,028	0,059	0,054
Armenia	0,056	0,028	0,059	0,054
Iran	0,044	0,056	0,118	0,162
Libya	0,044	0,028	0,059	0,054
Saudi Arabia	0,039	0,009	0,029	0,027
Bahrain	0,039	0,009	0,029	0,027
India	0,039	0,028	0,059	0,054
Russia	0,028	0,019	0,059	0,081
Syria	0,028	0,009	0,029	0,027
Eritrea	0,028	0,009	0,029	0,027
Morocco	0,028	0,009	0,029	0,027
Azerbaijan	0,028	0,009	0,029	0,027

24th International Conference on Group Decision and Negotiation &
10th International Conference on Decision Support System Technology
Human-Centric Decision and Negotiation Support for Societal Transitions

Fairly Allocating Indivisible and Divisible Items to Two Individuals

Steven J. Brams[1], D. Marc Kilgour[2*], and Christian Klamler[3]

[1] New York University, New York, USA
[2] Wilfrid Laurier University, Waterloo, CANADA
[3] University of Graz, Graz, AUSTRIA

[*] corresponding author: mkilgour@wlu.ca

Abstract
An ideal allocation of items to two players satisfies three fundamental properties of fairness: it is envy-free (neither player prefers the other's assignment to its own), equitable (the players' values for their assignments are equal --in their own measure), and Pareto-optimal (any other allocation is less preferred by at least one player). The AW algorithm can achieve this ideal, but its success is guaranteed only if all items are infinitely divisible. This paper explores the possibility of replacing this algorithm with one that applies even when most of the items to be allocated are indivisible.

Keywords: allocation; indivisible; envy-free; equitable; Adjusted Winner

1. Introduction

An ideal allocation of an unlimited number of distinct items to two players satisfies three fundamental properties of fairness: it is envy-free (EF), equitable (EQ), and Pareto-optimal (PO). EF means that neither player prefers the other's assignment to its own. EQ means that the players' values for their assignments are equal (in their own measure). PO means that any other allocation is less preferred by at least one player. If there are n items, then there are 2n possible allocations that do not divide any item; some of these are always PO, but almost always no EQ allocation exists, and sometimes there may be no EF allocation at all (Brams et al., 2023).

Adjusted Winner (AW) (Brams and Taylor, 1996) is an algorithm that apparently contradicts these statements. It always finds an EF-EQ-PO allocation. How can this be? AW works by dividing one item between the players. Unfortunately, AW works only when this item, which cannot be identified in advance, is continuously divisible. We consider explicitly whether an EF-EQ-PO allocation exists when all items except one, or a few, are indivisible and, if so, how to find it. For now, we assume exactly one indivisible item, which we call m (money). Using examples, we show that in this new context it remains possible that there will be no EF allocation, and that sometimes there may exist allocations that are EF-PO but not EQ.

We assume that the two players, A and B, can assign a value to each item, and that each player's value for any subset of items is the sum of the values for the individual items in that subset. In particular, m is an item, and both players must assign it a value – which, of course, may be different. For fairness, both players' total values for all items must be equal – we assume that each player's values are non-negative and, including the value for m, add to 100. Note that our assumption implies that neither player has any synergies, positive or negative, for any items, including m, that it may receive. Simple examples now give a sense of the problem.

1.1. Examples

To illustrate, suppose there is one indivisible item, i, as well as one divisible item, m. Suppose that both A and B think i is worth more than 50%, so that m is worth less than 50% to each player. In any allocation, item i must be assigned to exactly one player because it is indivisible. Then, no matter how m is assigned – including all possible ways of splitting m – the player who does not receive i will envy the player who does. More generally, as will be seen below, the availability of an EF-EQ-PO allocation is associated with relatively high values for m.

But if the situation were reversed, and the players agree that m is worth more than 50% and i less, then an EF-EQ-PO allocation is possible: assign i to the player who values it more, and then split m so that both players value their assignment equally. For instance, if the players values are as in Example 1:

Example 1: One item (*m*) divisible and one item (*i*) indivisible

	Item *i*	Item *m*
A	30	70
B	10	90

then the allocation in which A receives i plus 3/8 of m is worth 30 + (3/8) × 70 = 56.25 to A and (5/8) × 90 = 56.25 to B. This allocation, and only it, is EF-EQ-PO.

We believe that the allocation we have just described is better than any allocation produced by the famous "I cut, you choose" procedure. If A is the focal player, then A would offer i + (2/7) m or (5/7)m, which are both worth 50.00 to A. B would choose the latter bundle, which would be worth 64.28 to B. The situation is similar if B were the proposer. (B would propose i + (4/9) m or (5/9)m, which are both worth 50.00 to B. A would choose the former, seeing it as worth 61.11.) While these allocations are EF-PO, they seem unfair to the divider, who gets exactly 50% while the chooser generally gets more. The observation that "I cut, you choose" is not equitable motivates us to search for EF-EQ-PO allocations.

2. Adjusted Winner (AW)

Adjusted Winner (AW) is a two-person fair-division algorithm. The input to AW is A's and B's assignment of a fixed number of points (for us, 100) to divisible items to reflect the values they attribute to each item. The output is an allocation of all items that is EF-EQ-PO. AW proceeds in six steps:

1. Initially, A wins the items on which it puts more points, and B wins those on which it puts more points.
2. Tied items on which A and B put the same number of points are awarded, insofar as possible, to the player with fewer total points.
3. If each party's total number of points is the same, the procedure ends.
4. If one player (say, A) has more points than the other (B), then A gives back items to B in a certain order, and may split one, until both parties' points are equal.
5. The giveback starts with the item having the smallest ratio of A's (the initial winner's) points to B's points; if necessary, it continues to the item with the next-smallest ratio, and so on.
6. As soon as the giveback of an item from A to B would award B more total points than A, that item is split so as to equalize the points that A and B receive.

Altogether, these three properties imply that the number of points (out of 100) that each player receives is 50 or more, and this number is the same for both players (i.e., the allocation is equitable) and cannot be improved upon (the allocation is PO). We illustrate AW with an example below that we return to in section 3.

Example 2 (4 divisible items)

	a	*b*	*c*	*d*	Sum
A	<u>40</u>	20	15	<u>25</u>	100
B	20	<u>25</u>	<u>35</u>	20	100

The procedure begins with A's and B's choice of points, totalling 100, for the four divisible items {a, b, c, d}, as shown in the table. For each item, we have underscored the points of the player who puts more points on the item. Initially, A receives {a, d}, and B receives {b, c}, giving A 65 points in total and B 60 points in total, and putting A ahead of B. Consequently, A must give back some of its points to B, starting with the item initially assigned to A with the smallest ratio of A's points to B's. This is item d, because 25/20 < 40/20.

But the transfer of item d in its entirety is too much, as it leaves A with 40 points and gives 80 to B. To ensure that A and B receive the same total number of points, thereby satisfying EQ, we set A's points equal to B's when A receives the fraction x of item b, and B receives the complementary fraction 1 – x,

$$40 + 25x = 60 + 20(1 - x)$$

whose solution gives x = 8/9 = 0.89 of d to A and 1 – x = 1/9 = 0.11 of d to B.

Substituting into the above equation, we see that each player receives 62.22 points. This is the EF-EQ-PO solution in Example 1: Each player receives the same number of points equal to or greater than 50 (the allocation satisfies EF and EQ), and there is no allocation that gives strictly more points to both players (the allocation satisfies PO).

The key to AW is the splitting of exactly one item, called the equitability adjustment (Brams and Taylor, 1996, 1999). In practical examples, such as art objects, tickets to events such as concerts or sports, or hard candies, not all items can be meaningfully split between the two parties. Moreover, the item that AW splits cannot be predicted in advance of the application of the algorithm, so if some items are divisible and some are not, the applicability of AW is in doubt.

3. Adjusted Winner with Money (AWm)

3.1. Existence of EQ Allocations

We now consider a two-person allocation problem with n + 1 items, in which n items are indivisible and one item, m, is divisible. First of all, consider the possible application of AW to this problem. If AW specifies that m is the item to be divided, then we have an EF-EQ-PO allocation, and there is no more work to do. But if AW requires that some indivisible item, rather than m, be divided between the two players, then what can we do? We now develop a sufficient condition for the existence of an EF-EQ-PO allocation, provide an algorithm that identifies an EF-EQ allocation. Later, we discuss what happens if this allocation is not PO, and whether it can be "adjusted" to find an EF-EQ-PO allocation.

First, we define the problem formally. Assume a set I = {i1, i2, …, in} of n indivisible items along with one divisible item, m, to be allocated to two players, A and B, whose utilities for any allocation of the items are the sums of their utilities for the items they receive plus the portion of m, mA or mB, that the player receives. Thus, the problem is defined by A's utilities, (a1, a2, …, an, mA), and B's utilities, (b1, b2, …, bn, mB). As usual, all of these utilities are assumed positive. In particular, player A's utility of a fraction x of m is the product, xmA, and player B's xmB. In an allocation, each indivisible item must be allocated in its entirety either to A or to B, whereas m can be split between them.

First, we observe that no EF allocation may exist. This occurs, for example, when there is one indivisible item that both players agree is worth more than all of the others (including the divisible item m) put together. If so, then in any allocation the player who does not receive the uniquely valuable item will envy the one who does, so envy-freeness cannot be achieved.

Next, we note that even when an EF allocation exists, no allocation may be EQ. This can occur because the value that each side puts on money is too low to equalize their values of the other items, as illustrated by Example 3:

Example 3 (2 indivisible items and one divisible item)

	a	*b*	*m*	Sum
A	60	30	10	100
B	49	39	12	100

The unique EF-PO allocation is a to A and {b, m} to B, which is worth 60 to A and 51 to B. But it is not EQ and cannot be made EQ by shifting money between the players. Simply put, this problem admits no EQ allocations.

We now formulate a necessary condition for the existence of an EF-EQ allocation. Because the total of each player's utilities for the indivisible items and money must be 100, it follows that, if an EQ allocation exists, then it must be possible to find subsets of the indivisible items, I, SA ⊆ I and SB = I - SA, such that either uB(SB) ≤ uA(SA) ≤ uB(SB) + mB or uA(SA) ≤ uB(SB) ≤ uA(SA) + mA. Or, stated negatively, it is impossible to find an EQ allocation if and only if, for every allocation (SA. SB) of the indivisible items, either uB(SB) + mB < uA(SA) or uA(SA) + mA < uB(SB).

We assume that each player's utilities add to 100, i.e. uA(I) + mA = uB(I) + mB = 100. Suppose that uB(SB) ≤ uA(SA) ≤ uB(SB) + mB. Then there is an EQ allocation in which A receives SA and the portion $x = \frac{m_B-(u_A(S_A)-u_B(S_B))}{m_B+m_A}$ of m. This allocation is EF if and only if the common value is at least 50, i.e.,

$$\frac{m_B u_A(S_A) + m_A u_B(S_B) + m_A m_B}{m_A+m_B} \geq 50$$

(If uA(SA) ≤ uB(SB) ≤ uA(SA) + mA, the expression for x changes, but the common value is the same.) It follows that there is an EF-EQ allocation if and only if either uB(SB) ≤ uA(SA) ≤ uB(SB) + mB or uA(SA) ≤ uB(SB) ≤ uA(SA) + mA, and, in addition,

$$\frac{m_B u_A(S_A) + m_A u_B(S_B) + m_A m_B}{m_A+m_B} \geq 50$$

Next, consider what happens when AW is applied to our problem. The item that AW splits between the two players may turn out to be m. In that case, the AW allocation is EF-EQ-PO, and is the solution that we seek. We note that, in any case, the common value of the AW allocation is a maximum on the value to a player of any allocation that is EF-EQ-PO; all EF-EQ-PO allocations that do not split any indivisible item (though they may split m) must have a lower common value.

4.2 The Algorithm

We now consider what happens when the item split by AW is not money – in other words, when AW cannot solve our problem. Enumerate the indivisible items so that the ratios ak/bk are decreasing in k. (Also assume that all of these ratios are distinct.) Now consider Example 4, below, which repeats Example 2 in the sense that items a, b, and c are now assumed to be indivisible, and indexed by k = 1, 2, 3, whereas item d is now m, the only divisible item. In Example 4, the ordering of indivisible items is abc or k = 123.

Example 4 (3 indivisible items and *m*)

	$j = 1$	$j = 2$	$j = 3$		
	a	*b*	*c*	*m*	Sum
A	20	25	30	25	100
B	10	30	40	20	100

Next, consider assigning the indivisible objects so that A receives items with lower subscripts (starging at j = 1) and B receives items with higher subscripts, ending at j = n. Then the two players' values are

$$V_A^k = \sum_{j=0}^{k} a_j\,; \quad V_B^k = \sum_{j=k+1}^{n+1} b_j$$

for k = 0, 1, …, n. As a convention, we take a0 = 0, so that VA0 = 0 represents the situation where all indivisible items are assigned to B, and bn+1 = 0, so that VBn+1 = 0 represents the situation where all indivisible items are assigned to A. These cumulative values, plus the difference VAk − VBk, are shown in the last three rows of the next table. Notice that VAk − VBk is strictly increasing in k, negative for k = 0, and positive for k = n.

Example 4 with cumulative allocations

	$j = 0$	$j = 1$	$j = 2$	$j = 3$	*m*	Sum
A		20	25	30	25	100
B		10	30	40	20	100
V_A^k	0	20	45	75		
V_B^k	80	70	40	0		
V_A^k - V_B^k	-80	-50	5	75		

In Proposition 1, below, we show that, if the following condition is satisfied, there is an allocation of the indivisible items to A and B, and a split of *m*, that is EF and EQ,

Condition C: There exists a value of *k*. $1 \leq k \leq n$, such that either $-m_A \leq V_A^k - V_B^k \leq 0$ or $0 \leq V_A^k - V_B^k \leq m_B$.

To interpret Condition C in words, assume that $0 \leq V_A^k - V_B^k \leq m_B$. Then, B's valuation for *m*, m_B, is at least equal to the difference between A's cumulative valuation of the indivisible items up to *k* and B's cumulative valuation of indivisible items from *k* + 1 to *n*. In other words, if $S_A = \{i_1, i_2, \ldots, i_k\}$ is assigned to A and $S_B = \{i_{k+1}, i_{k+2}, \ldots, i_n\}$ is assigned to B, then we have $u_B(S_B) \leq u_A(S_A) \leq u_B(S_B) + m_B$, which, as noted above, guarantees that an EQ allocation exists. The situation is similar if $-m_A \leq V_A^k - V_B^k \leq 0$, for then $V_A^k \leq V_B^k \leq V_A^k + m_A$, or equivalently, $u_A(S_A) \leq u_B(S_B) \leq u_A(S_A) + m_A$.

We now show that Condition C is guaranteed to produce an "AW-like" allocation that is EF-EQ. But for now, we cannot draw any conclusion about whether it is PO.

Proposition 1. *If Condition C holds for k, then there exists an EF-EQ allocation in which A receives items* $i_1, i_2, \ldots, i_k$ *plus a fraction x of m, and B receives* $i_{k+1}, i_{k+2}, \ldots, i_n$ *plus the complementary fraction,* $1 - x$, *of m. Moreover, the value of x satisfies*

$$x = \begin{cases} \dfrac{V_B^k - V_A^k + m_B}{m_A + m_B} & if\ \ m_A \geq V_B^k - V_A^k \geq 0 \\ \dfrac{V_A^k - V_B^k + m_A}{m_A + m_B} & if\ \ m_B \geq V_A^k - V_B^k \geq 0 \end{cases}$$

Proof. See the full version of this paper.

The algorithm set out in Proposition 1 is called AWm. Note that it cannot be applied if Condition C fails. If Condition C holds, the AWm allocation is EF-EQ. If Condition C holds for more than one value of *k*, Proposition 2 describes how to determine the specific value of *k* producing the allocation with the highest utilities for each player – the AWm allocations for other values of *k* are EF-EQ, but cannot be PO.

Even when an AWm allocation exists, it may fail to be PO, as illustrated by Example 6, where Condition C holds, and AWm yields the EF-EQ allocation in which A receives indivisible items 1, 2, and 3, plus (9/40)*m,* and B receives item 4 plus (31/40)*m*; this allocation gives each player's utility 66.75. However, the allocation of items 1 and 3 plus (39/40)*m* to A and items 2 and 4 plus (1/40)*m* to B gives each player a utility of 69.25 and is the unique EF-EQ-PO allocation in this example.

Example 6 (with cumulative allocations)

	$j = 0$	$j = 1$	$j = 2$	$j = 3$	$j = 4$	m	Sum
A		10	20	30	10	30	100
B		4	10	17	59	10	100
V_A^k	0	10	30	60	70		
V_B^k	90	86	76	59	0		
$V_A^k - V_B^k$	-90	-77	-47	1	70		

5. Computer Simulations

If all possible valuations of the indivisible items and m are equiprobable, Condition C is satisfied 67% of the time, the fraction increasing to 96% when valuations of m are at least equal to the average valuation of an indivisible item. Although information on the opponent's preferences can make AWm manipulable, a player is assured of obtaining at least 50% of its valuation of all items if it is truthful and Condition C is satisfied.

6. Conclusions

We illustrate AWm by applying it to the division of marital property in a real-life divorce, suggesting how it can be applied to other fair-division problems with a combination of indivisible and divisible items. AWm is not always applicable, but it does represent a good start on the determination of satisfactory procedures for allocating a combination of indivisible and divisible items.

References

Brams, S.J., D.M. Kilgour, C. Klamler, and F. Wei (2023). "Two-Person Fair Division of Indivisible Items: Bentham vs. Rawls on Envy," Journal of Philosophy 120 (8), 441-456.

Brams, S.J., and A.D. Taylor (1999). The Win-Win Solution: Guaranteeing Fair Shares to Everybody. New York: W.W. Norton.

Brams, S.J., & A.D. Taylor (1996). Fair division: From cake-cutting to dispute resolution. New York: Cambridge University Press.

Allocating Indivisible Items Using Accelerated Fallback Bargaining

Fahimeh Ziaei[1] and D. Marc Kilgour[1*]

[1] Wilfrid Laurier University, Waterloo, CANADA
[*] corresponding author: mkilgour@wlu.ca

Abstract

In studies of collective decision-making, the problem of allocating indivisible items to two parties fairly and efficiently is now recognized as among the most difficult. The Fallback Bargaining algorithm has been suggested as a promising approach to address this problem, as it achieves fairness by finding relatively few scenarios that achieve both parties' interests insofar as possible. However, a significant drawback emerges when many items are to be allocated, as the length of the Fallback Bargaining input comes to resemble a prolonged negotiation process. We introduce a novel approach aimed at accelerating the Fallback Bargaining algorithm, enhancing its overall performance and making it more applicable in real-world allocation scenarios.

Keywords: 2-person fair division; indivisible items; Pareto-optimal; Fallback Bargaining; Social Choice

1. Introduction

The problem of fair allocation arises in a wide range of contexts; in particular, a decision maker may have complete information about the preferences of others, or may know only their relative preferences, or may be completely uncertain about their values. Fairly allocating divisible resources is possible when there is full information about all players' preferences, but even then, the challenge increases dramatically when what is to be allocated are indivisible items. In earlier papers [14,15], we considered whether and how indivisible items could be allocated to two players who knew each other's preference orderings to satisfy criteria such as envy-free, Pareto-optimal, maximin, maximum Borda sum, and Borda maximin. The conclusion was that the Fallback Bargaining algorithm [13] has the ability to satisfy most of these criteria of efficiency and fairness.

Research on fair-division problems and their applications falls within the general theory of fair allocations of Thomson [13]. The general framework for modelling a set of players and indivisible items, where each player's preferences over the items are given as utilities, is debated by Sen [1] and Kilgour and Vetschera [10]. In some studies, the focus is on situations in which participants have complete knowledge of each other's (additive) preferences over items; the objective is to compare algorithms, aiming to find allocations that satisfy three properties of fairness and efficiency.

The efficiency and fairness of allocations is measured by basic criteria---are they envy-free, Pareto-optimal, and maximin [5,6,7,13] Voting procedures may help, but sometimes no envy-free allocation exists [11]. Allocations can also be assessed using Borda properties, based on Borda counts [8]. Many practical allocation contexts are exemplified by two-player problems, which must be solved prior to addressing more ambitious applications, like the allocation of portfolios of assets, the drafting of players to sports teams, and the division of assets in a divorce.

A maximin allocation, which maximizes the minimum rank of any item received by any player, may be attractive, as often players seek to avoid “bad" items [2]. For instance, in a scenario where a team's performance depends mainly on the capability of the least capable team member, a manager might evaluate possible teams using a maximin criterion. This concept has been applied by Brams and King [6] and Herreiner and Puppe [9], and is central to the Fallback Bargaining algorithm proposed by Brams and Kilgour [4]. In a previous paper, we examined the relationship between the correlation of two players' rankings and features of Fallback Bargaining including depth of agreement and probability of a (two-way) tie [15].

The main objective of this study is to use the players' rankings to identify outcomes of Fallback Bargaining algorithm without actually applying the algorithm. Our idea is to leverage players' preference orderings to expedite the Fallback process.

2. Properties of allocations

In this paper, two players must share a set consisting of an even number of indivisible items that each strictly ranks from best to worst. The following assumptions describe the processes we consider:

1. *Self-interest.* Each player aims to obtain the best items it can, not to hurt the opponent.
2. *Independence.* Each player acts independently. There are no coalitions or hidden agreements.
3. *Partial information.* Both players know each other's preference ordering, though not each other's utilities.
4. *Balance*. Each player receives an equal number of items.
5. *Synergy-free.* There are no synergies, positive or negative, among the items received by any player.

Consider a set S consisting of $2n$ distinct items to be allocated to players A and B. Suppose $X_A \subseteq S$ and $X_B \subseteq S$ satisfy $X_A \cap X_B = \emptyset$ and $X_A \cup X_B = S$. Then $X = (X_A, X_B)$ is the allocation in which X_A is assigned to A and subset X_B is assigned to B. Because our allocations must be balanced allocations, we require that $| X_A| = X_B| = n$.

We assume that players have preferences on S. For player m in M = {A, B}, $x \prec_m y$ means that item x in S is less preferred by m than item y in S. In this case, we also write $y \succ_m x$. We assume that player m's preferences are complete, asymmetric, and irreflexive, and therefore form a linear (strict) ordering. Player m's rank for item x in S is

$$r_m(x) = | \{s \text{ in } S: s \succ_m x\}|$$

But we will need to consider the possibility of indifference of distinct items. If player m is indifferent between item x and item y, we write $x \sim y$. In this case, we replace the rank function $r(m)$ by

$$r^*_m(x) = |\{s \text{ in } S: s \succ_m x \}| + 0.5 * |\{s \text{ in } S \setminus \{x\}: s \sim_m x \}|$$

Thus, $r_m(x)$ equals the number of items in S that m prefers to x, and in $r^*_m(x)$, this rank is augmented by half the number of items indifferent to x. Each player's preference ordering on S is assumed to be public information. To assess possible allocations, we must take account of players' preferences on subsets of S. To some extent, a player's preferences on subsets are implied by the player's preferences on individual items. We say that $X \subseteq S$ is ordinally less than $Y \subseteq S$ for m, denoted $X \prec_m Y$, if there exists an injective mapping $f: X \rightarrow Y$ such that for all x in X, $x \prec_m f(x)$. If $X \prec_m Y$, then $Y \succ_m X$, and we say that Y is ordinally more than X. We emphasize that each player has some preferences between subsets that cannot be captured using the ordinally less criterion, but these preferences are private information.

For any subset X ⊆ S, define the ranksum of X for m to be

$$r_m(X) = \Sigma\ (x \text{ in } X)\ r_m(x)$$

and the augmented ranksum of X for m to be

$$r^*_m\ (X) = \Sigma\ (x \text{ in } X)\ r^*_m\ (x)$$

It is easy to show that, if $X \prec_m Y$, then $r_m\ (X) < r_m\ (Y)$ and $r^*_m\ (X) < r^*_m\ (Y)$.

An allocation (X_A, X_B) is envy-free if $X_A \succ_A X_B$ and $X_B \succ_B X_A$. It is Pareto-optimal if there is no other allocation (Y_A, Y_B) such that $Y_A \succ_A X_A$ and $Y_B \succ_B X_A$.

For any $T \subseteq S$, $T \neq \emptyset$, let $\min_m \{T\}$ be player *m*'s least preferred (highest ranked) item in *T*. Thus, $\min_m \{T\} = x$ if and only if *x* in *T* satisfies $y \succ_m x$ for all *y* in $T - \{x\}$. Also, we define player *m*'s maximin value for *T* to be $r_m(\min_m \{T\})$, the rank of *m*'s least preferred item in *T*. The maximin value of an allocation (X_A, X_B) is $r(X_A, X_B) = \max\{r_A(\min_A \{X_A\}), r_B(\min_B \{X_B\})\}$, the rank of the least preferred item that (X_A, X_B) assigns to either player. An allocation (X_A, X_B) is maximin if there is no other allocation (Y_A, Y_B) such that $r(Y_A, Y_B) < r(X_A, X_B)$.

Our examples follow the convention that items are named in order of A's preference, so that A's preference ordering of S is always

A: $1 \succ_A 2 \succ_A \ldots \succ_A 2n$

Of course, B has $(2n)!$ possible orders of preference on *S*, so there are $(2n)!$ distinct allocation problems with $2n$ items.

3. Using Fallback Bargaining to select an Allocation

3.1. Fallback Bargaining

The Fallback Bargaining algorithm is a step-by-step procedure to select from many possible agreements. Let the set of possible bargains, over which the two players have preferences, be *S*, where $1 < |S| <$ infinity. Fallback Bargaining begins with each player reporting to the referee their preference ordering of *S* in decreasing order of preference. Then it proceeds as follows:

1. Consider each player's most preferred outcome. If it is the same for both players, then it is implemented and constitutes a depth $j = 1$ agreement. The process stops. Otherwise, set $j = 2$ and proceed to Step 2.
2. Assume that no item is common to both players' top j - 1 items. There is agreement at depth j agreement if there is an item common to both players' top j items. Any such item constitutes a depth j agreement, and the process stops. If there is no item common to the players' top j items, and if $j < |S|$, increase j by 1 and repeat Step 2. Otherwise, terminate with no agreement.

It can be shown that, even if players' preferences are strict, Fallback Bargaining must end at depth at most floor($|S|/2$ +1), so the procedure must terminate if $|S| <$ infinity. Assuming preferences are strict, there can be at most two Fallback outcomes. The Fallback outcomes are all values of *s* in *S* that minimize $\max \{r_A(s), r_B(s)\}$, which have been called the maximin outcomes. Here, "maximin" refers to maximizing the preference for the least desirable item, as opposed to choosing *s* in *S* to minimize r(s).

Note that the Fallback Bargaining process must stop, as there is a number, at most $j = n + 1$, such that at least one item is common to both players' top j, but no item is common to their top j - 1 allocations. Any such item is a depth j agreement. The Fallback Bargaining method has been proven to satisfy many desirable criteria of choice, including possibly Pareto-optimal, possibly envy-free, maximin, and Borda maximin.

We now extend Fallback Bargaining to weak preferences. Assume that one player is indifferent between two items. (If both players are indifferent, we consider the items identical.)

Example 1:

$$A: 1 >_A 2 \sim_A 3 >_A 4$$

$$B: 3 >_B 2 >_B 1 >_B 4$$

The Fallback outcome seems to be item 2 at depth 2. But, taking into account that $2 \sim_A 3$, there exists another item, 3, that is Pareto-superior to 2.

We address this failure to find a Pareto-optimal outcome with a convention: If one player (say, A) is indifferent between two items, then A's preference listing is modified so that the two indifferent items are in the order of the opponent's (B's) preference. Thus, Example 1 becomes

Example 1A:

$$A: 1 \succ_A 3 \sim_A 2 \succ_A 4$$

$$B: 3 \succ_B 2 \succ_B 1 \succ_B 4$$

Thus, the Fallback procedure delivers the Pareto-optimal outcome 3 at depth 2.

Now return to the assumption that each player has a strict preference on *S* and consider the problem of finding a balanced allocation of all items of *S*. We consider how to use Fallback Bargaining to choose an allocation. We also assume that the referee receives the players' rankings of *S*. Consequently, the referee knows that one subset of *S* is preferred to another if and only if it is ordinally less.

3.2 Accelerated Fallback Bargaining Method

The Fallback Bargaining process can be used to decide on an allocation of *S*. In this case, it must be applied to subsets of size $2n$, and each player must have an ordering of all these subsets. Fallback Bargaining can be carried out by an impartial referee (or a computer). However, there are several challenges to overcome.

One drawback of the Fallback Bargaining procedure is that two subsets may tie, both being the first mutually acceptable subsets at the same level. Fortunately, at most two allocations can tie in this way if the original orderings are strict; in general, ties cannot proliferate even as *S* becomes large.

A second drawback is that Fallback Bargaining must be applied not to items but to bundles (possible assignments) and, given our assumptions, preferences over bundles may not be complete. If each player submits their preference ordering, then there will be many bundles that are tied—at least in public information. This violates the standard Fallback assumption that preference orderings are strict. Below we propose an amendment to the Fallback procedure to deal with weak, as opposed to strict, preferences.

A third drawback is that navigating the journey of agreement-building between players using the Fallback Bargaining approach may be a long and complex process. If $2n$ is large, then there are many possible assignments to each player, and the Fallback Bargaining process may be very long. For instance, if the players' preferences are exactly opposite, agreement will require $\binom{2n}{n}$+ 1 steps. To illustrate, consider a scenario where there are $2n = 12$ items to be allocated. Regardless of their preference orderings, the players must collectively explore 925 different allocations—a daunting task.

To address this drawback, we introduce a method that involves examining the order of players' preferences for items. Through this approach, we can identify all the items that each player is to receive; these items can be allocated immediately to the player who is to receive them.

This document outlines the Accelerated Fallback Bargaining Method designed for efficiently allocating indivisible items between two players. Follow these steps:

1. **Step 1**: Calculate the rank difference, Δ, for each item, defined as $\Delta(x) = |r_A(x) - r_B(x)|$. This represents the absolute difference between the ranks assigned to items by the two players.
2. **Step 2**: If Δ is 2 or greater for any item, allocate that item to the player who ranks it higher, and then remove the item from the list.
3. **Step 3**: For the remaining items, explore all possible allocations between the two players and calculate the Borda sum for each allocation, including the items allocated in Step 2.
4. **Step 4**: Identify the allocation that results in the minimax Borda sum for both players.

Example 2:

$$\text{Player A: } 1 \succ_A 2 \succ_A 3 \succ_A 4 \succ_A 5 \succ_A 6$$
$$\text{Player B: } 6 \succ_B 1 \succ_B 2 \succ_B 3 \succ_B 4 \succ_B 5$$

We first calculate Δ for each item:

$\Delta(1) = |1 - 2| = 1$
$\Delta(2) = |2 - 3| = 1$
$\Delta(3) = |3 - 4| = 1$
$\Delta(4) = |4 - 5| = 1$
$\Delta(5) = |5 - 6| = 1$
$\Delta(6) = |6 - 1| = 5$

Since $\Delta(6)$ is the only value exceeding 2, so item 6 is allocated to player B. We then proceed to consider all remaining possible allocations and their respective Borda scores:

(123,456) with Borda scores (12,6)
(124,356) with Borda scores (11,7)
(125,346) with Borda scores (10,8)
(134,256) with Borda scores (10,8)
(135,246) with Borda scores (9,9)
(145,236) with Borda scores (8,10)
(234,156) with Borda scores (9,9)
(235,146) with Borda scores (8,10)
(245,136) with Borda scores (7,11)
(345,126) with Borda scores (6,12)

Definition 1: In case there is a tie over two or more assignments in the preference ordering of one player and no tie for the other, the rank ordering of allocations in the Fallback procedure is determined by the preference of the player with strict preference.

For instance, consider applying Fallback to the next example:

Example 3:

$$\text{A: } 1 \succ_A 2 \sim_A 3 \succ_A 4$$
$$\text{B: } 2 \succ_B 3 \succ_B 4 \succ_B 1$$

Note alternative 2 is Pareto-optimal whereas alternative 3 is not. But if alternative 3 were listed prior to alternative 2, then the Fallback outcome would be 3 and not 2.

4. Conclusions

It has been established that, among various algorithms employing different approaches to achieve the optimal allocation of an even number of indivisible items based solely on two players' preference orderings, the Fallback Bargaining algorithm stands out as the only reliable method. This algorithm effectively meets the criteria of Pareto Optimality, Envy-Freeness, Maximin, as well as Bordasum and Borda Maximin.

The simultaneous actions of players in the Fallback Bargaining algorithm are instrumental in attaining an allocation that satisfies most of the efficiency and fairness criteria, including Borda Maximin. However, a notable drawback is the potential for a long bargaining process, making it lengthy in real-world contexts. In this paper, our aim was to accelerate this process by introducing a step-by-step method which minimizes the players' preference ordering lists, streamlining the bargaining process for quicker resolutions. The proposed approach can also be extended to address problems involving ties among preference orderings.

References

A, Sen.: The Possibility of Social Choice. American Economic Review, 89, 349--378, (1999).

Brams, S.J., Kilgour, D.M. and Klamler, C., 2017. Maximin envy-free division of indivisible items. Group Decision and Negotiation, 26, pp.115-131.

Brams, S.J., Kilgour, D.M. and Sanver, M.R., 2007. A minimax procedure for negotiating multilateral treaties. Diplomacy Games: Formal Models and International Negotiations, pp.265-282.

Brams, S. J., and Kilgour, D. M.: Fallback bargaining. Group Decision and Negotiation, 10(4), 287-316 (2001).

Brams, S.J., Kilgour, D.M. and Klamler, C.: Maximin envy-free division of indivisible items. Group Decision and Negotiation, 26(1), 115-131 (2017).

Brams, S.J. and King, D.L., 2005. Efficient fair division: help the worst off or avoid envy?. Rationality and Society, 17(4), pp.387-421.

Bouveret, S., Lemaître, M.: Characterizing conflicts in fair division of indivisible goods using a scale of criteria. Autonomous Agents and Multi-Agent Systems, 30(2), 259—290 (2016).

Darmann, A., Klamler, C.: Proportional Borda allocations. Social Choice and Welfare, 47(3), 543-558 (2016).

Herreiner, D. and Puppe, C., 2002. A simple procedure for finding equitable allocations of indivisible goods. Social Choice and Welfare, 19, pp.415-430.

Kilgour, D. M.,Vetschera, R.: Two-player fair division of indivisible items: Comparison of algorithms. European Journal of Operational Research, 271(2), 620-631 (2018).

Lang, J.: Collective Decision Making under Incomplete Knowledge: Possible and Necessary Solutions. In: 29th International Joint Conference on Artificial Intelligence and Seventeenth Pacific Rim International Conference on Artificial Intelligence, pp. 4885—4891. Yokohama, Japan (2020)

Pruhs, K., Woeginger, G. J.: Divorcing made easy. In: International Conference on Fun with Algorithms, pp. 305--314. Springer, Berlin, Heidelberg (2012).

Thomson, W. F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Introduction to the theory of fair allocation. Procaccia, Handbook of Computational Social Choice (2016).

Ziaei, F. and Kilgour,D. M.: Comparing Algorithms for Fair Division of Indivisible Items. In: 22th International Proceedings of Group Decision and Negotiation, GDN 2022, 51-62.

Ziaei, F. and Kilgour, D.M. (2023). Comparing Algorithms for Fair Allocation of Indivisible Items with Limited Information. In: Maemura, Y., Horita, M., Fang, L., Zaraté, P. (eds) Group Decision and Negotiation in the Era of Multimodal Interactions. GDN 2023. Lecture Notes in Business Information Processing, vol 478. Springer, Cham. https://doi.org/$10.1007/978-3-031-33780-2-9$.

Ziaei, F and Kilgour, D. M, The Consequences of Preference Correlation for Fair Division of Indivisible Items, in 2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Honolulu, Oahu Island, October 1-4, 2023. Not published. eCF Paper Id: othersmain. SMC23.1271.8bb6c894.

Group Decision and Negotiation - Negotiation Support Systems and Studies -

Conference papers and extended abstracts

A dynamic model of outcome expectations in bilateral negotiations
Extended abstract

Rudolf Vetschera[1*] and Luis C. Dias [2]

[1] University of Vienna, Vienna, Austria
[2] University of Coimbra, Coimbra, Portugal
[*] rudolf.vetschera@univie.ac.at

Abstract

We study a model of a bargaining process between two myopic negotiators, who exchange several offers during a prolonged process. Negotiators are not perfectly rational, but focus on the current step, taking into account their expectations about the eventual outcome of the negotiation. These expectations are determined by their confidence about reaching an agreement that is favorable to them. In this paper, we extend previous work by analyzing a situation in which confidence changes over time. We find that a party whose confidence increases over time obtains better outcomes, but this effect might be cancelled out if confidence of the other part also increases. We also identify a trade-off between confidence and risk attitude.

Keywords: Zeuthen-Hicks bargaining, expectations, negotiator confidence

1. Introduction

Perfectly rational bargainers, who anticipate the entire bargaining process and its eventual equilibrium outcome, would never engage in a lengthy and costly bargaining process, but would jump immediately to the equilibrium solution, as predicted e.g. in the seminal model of Rubinstein (1982). Models that describe a prolonged bargaining process therefore need to consider that parties are not perfectly rational or lack necessary information. This works builds on an extension (Dias & Vetschera, 2022) of the Zeuthen-Hicks bargaining model (Harsanyi, 1965; Bishop, 1964). This model considers myopic negotiators, who only consider a single bargaining step (and the opponent's possible response to that one step) at a time. At each bargaining step, the negotiator has the option to accept the opponent's offer, terminate the negotiation, or continue with a counter-offer. To evaluate the option of continuing the negotiation, the negotiator needs to form a subjective estimate of the eventual outcome of the negotiation. Negotiators could be more or less confident when forming this estimate, so confidence plays an important role in negotiation behavior, as is also illustrated by empirical studies that demonstrate the effect of (over-) confidence on negotiation behavior (Caputo, 2013).

A negotiator's confidence might change over time during a negotiation. In this study, we analyze how such changes influence the bargaining process and its outcomes. We begin in section 2 by explaining an analytical model and some results for the case of constant confidence. To explore effects of changes in confidence, we performed a simulation study, of which we present some results in section 3. A brief conclusion in section 4 completes this extended abstract.

2. Model and analytical results

We consider a two-party negotiation, and for simplicity identify the parties as “buyer” and “seller”. They bargain over a single issue, which we call the “price”. Denote by s and b the two offers (suggested prices) currently on the table, s from the seller and b from the buyer, and consider the decision of the seller whether to continue the negotiation or accept the buyer’s offer. If the negotiation continues, the seller will expect to eventually obtain some outcome z_s, which is between the two offers, i.e.

$$z_s = \beta_s s + (1 - \beta_s) b$$

where β_s represents the seller’s confidence. Assuming that negotiations will fail with probability p, leading to zero utility for the seller, the seller will decide to continue the negotiation rather than making a counter-offer, if

$$(1 - p) u_s(z_s) > u_s(b)$$

where $u_s()$ refers to the seller’s utility function. From this condition, one can obtain a critical probability of failure for the seller, and similarly for the buyer. According to the logic of Zeuthen-Hicks bargaining, the party whose critical probability is lower will make a concession that is large enough to revert that condition,

In Dias & Vetschera (2022), we show that this process will lead to the Asymmetric Nash Bargaining Solution (ANS)

$$x^* = arg\,max\, u_s\,(x)^{\beta_s} u_b(x)^{\beta_b}$$

An interesting special case occurs is the two utility functions are power functions $u_s(x) = x^{e_s}$ and $u_b(x) = (1 - x)^{e_b}$, where e_s and e_b represent the risk aversion coefficients of seller and buyer, respectively. In that case, the ANS is given by

$$x^* = \frac{e_s \beta_s}{e_s \beta_s + e_b \beta_b}$$

This result shows an interesting interaction between risk aversion and confidence, i.e., a high level of risk aversion can be compensated by higher confidence and vice versa low confidence by less risk aversion.

3. Simulation results

The analytical model introduced above assumes constant confidence throughout the negotiation. To analyze the effects of changing levels of confidence, we conducted an extensive simulation. For each simulation experiment, we generated concave utility functions for the buyer and the seller using the method of (Dias & Vetschera 2019). We then simulated the bargaining process as described in the previous section for increasing or decreasing levels of confidence, using a concave, linear, or convex shape of the trajectory of confidence. For each parameter setting, 1,000 bargaining processes with different utility functions were simulated. For brevity, we present only two interesting results in this extended abstract, more detailed results and a description of the simulation framework can be found in (Dias & Vetschera 2024).

Our simulation indicates that negotiations in which both parties exhibit increasing confidence take significantly longer than negotiations in which only one party has increasing or both have decreasing confidence, as shown in Figure 1.

Another interesting phenomenon concerns the development of concessions over time as shown in Figure 2. This figure shows the concessions made by the seller in each quarter of the negotiation, expressed as a fraction of the total possible price range (which was standardized to prices between zero and one in the simulation). For simplicity, this figure only presents the case that both confidence levels change linearly over time, however, the shape of the trajectory only has a small influence on this variable.

If the level of confidence of both parties increases during the negotiation, concessions are made in a similar way throughout the negotiation. In contrast, if both levels of confidence decrease, the parties make larger concessions as the negotiation proceeds. Different directions of change have a strong effect. If the seller's confidence increases, while the buyer's confidence decreases, the seller gradually makes smaller concessions. The most striking effect occurs in the opposite case: If a seller with decreasing confidence faces a buyer with increasing confidence, the seller's concessions increase rapidly over time.

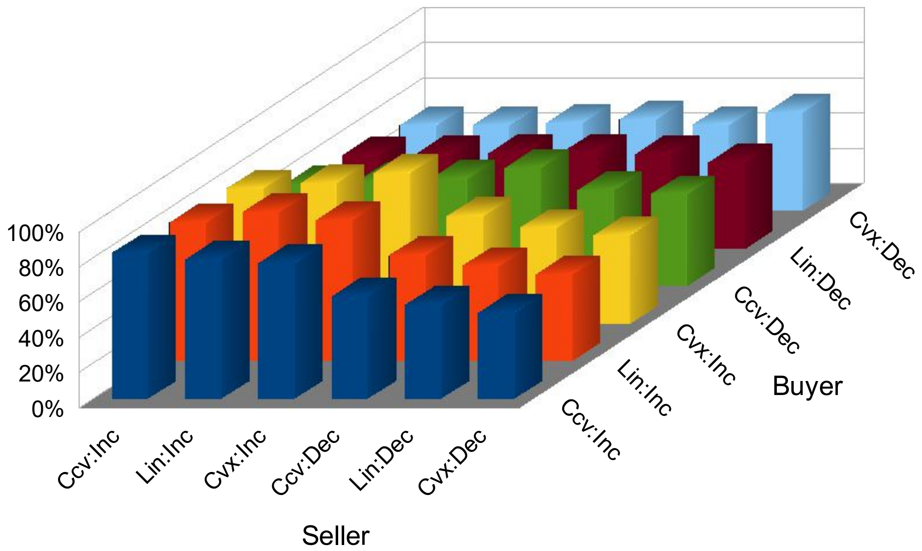


Figure 1: Duration of negotiations (in % of theoretical maximum) for different developments of confidence (Shape: Ccv=Concave, Lin=Linear, Cvx=Convex: Direction: Inc=Increasing, Dec=Decreasing)

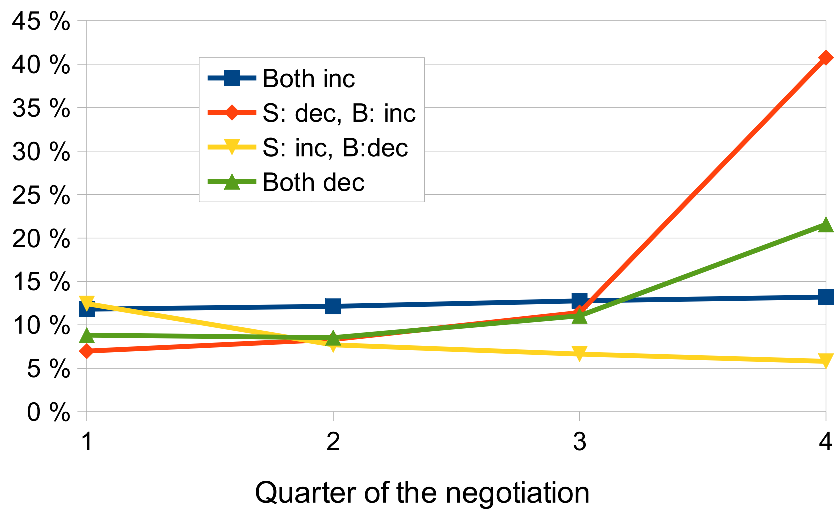


Figure 2: Concessions by seller (in % of total range) during different quarters of the negotiation for linear change of confidence (Party: S=Seller, B=Buyer; Direction: inc=increasing, dec=decreasing confidence)

4. Conclusions

A negotiator's confidence in his or her abilities to achieve a favorable outcome of a negotiation is an important factor shaping the bargaining process. This influence is particularly important if negotiators proceed more or less cautiously in small steps towards an agreement. Such a stepwise process might be due to negotiators being not perfectly rational, or lacking information that can be obtained only in a prolonged interaction with the opponent. Which steps are taken at which time depends on the expectations and thus the confidence of negotiators.

In this work, we have studied a model that allows to represent these considerations formally. It is possible to show analytically that this model leads to a specific bargaining solution, the ANS, where the impact of negotiator confidence is also visible in the outcomes. We have also shown that at least for some specifications of the negotiators' utility functions, there is a clear substitution effect between risk tolerance and confidence of a negotiator. Using a dynamic simulation model, we then have studied the impact of changing levels of confidence during the negotiation. These results show again a clear trade-off: increasing levels of confidence can benefit a negotiator or at least (when facing another negotiator with increasing confidence) will avoid a spiral of increasing concessions, but will lead to considerably longer negotiations.

The model we presented in this extended abstract is still a rather simplistic view of the role of changing confidence in negotiations. In particular, our model considers changes in confidence to be exogenous, while in practice, the course a negotiation takes will possibly influence the expectations and also the confidence of a negotiator. Furthermore, the results we obtained so far are based on analytical modelling and simulations, it remains to be seen whether they can be confirmed in empirical studies with human subjects. While the dynamic model has not yet been tested empirically, results using a static model of confidence have confirmed validity of the overall framework (Vetschera & Dias, 2023). We therefore consider this model a useful first step in analyzing the interplay between negotiator confidence and the negotiation process.

References

Bishop, R.L. (1964). A Zeuthen-Hicks theory of bargaining. *Econometrica*, 32 ,410-417

Caputo, A. (2013). A literature review of cognitive biases in negotiation processes. *International Journal of Conflict Management*, 24 (4), 374-398

Dias, L.C., & Vetschera, R. (2019). On generating utility functions in stochastic multicriteria acceptability analysis. *European Journal of Operational Research*, 278 , 672-685

Dias, L.C., & Vetschera, R. (2022). Two-party bargaining processes based on subjective expectations: A model and a simulation study. *Group Decision and Negotiation*, 31, 843–869

Dias, L.C. & Vetschera, R. (2024) A dynamic model of outcome expectations in bilateral negotiations, *Submitted for publication*

Harsanyi, J.C. (1956). Approaches to the bargaining problem before and after the theory of games: A critical discussion of Zeuthen's, Hicks', and Nash's theories. *Econometrica*, 24 (2), 144-157

Rubinstein, A. (1982). Perfect equilibrium in a bargaining model. *Econometrica*, 50 (1), 97-109

Vetschera, R., & Dias, L.C. (2023). Confidence in bargaining processes and outcomes: Empirical tests of a conceptual model. *EURO Journal on Decision Processes*, 11, 100028,

.

joint conference
group and negotiation DECISION support systems technology
Porto 2024

24th International Conference on Group Decision and Negotiation &
10th International Conference on Decision Support System Technology
Human-Centric Decision and Negotiation Support for Societal Transitions

An Artefact for Self-Reflection in E-Negotiation Training

Marlene Meyer [1*] and Mareike Schoop [1]

[1] University of Hohenheim, Schwerzstrasse 40,
70599 Stuttgart, Germany
[*] marlene.meyer@uni-hohenheim.de

Abstract

The usage of electronic negotiation systems increases. Thus, a structured electronic training is essential to enhance negotiators performance and negotiation skills supporting self-reflection in the post-negotiation phase. We investigate to what extent a negotiation support system (NSS) could provide such feedback and developed seven feedback elements as artefacts; four of which were shown to be useful in the post-negotiation phase of e-negotiations in general and in particular for self-evaluating the behaviour.

Keywords: e-negotiation; e-training; self-reflection; negotiation skills; NSS

1. Introduction

"An important aspect of experience and training is for negotiators to reflect after completing negotiations" (Thompson, 2022, p. 215). A negotiation is called electronic negotiation (e-negotiation) if an electronic medium supports the negotiation, and the support includes at least one decision-making or communication task (Ströbel & Weinhardt, 2003). Negotiation Support Systems (NSSs) enable electronic negotiations; they should consequently provide electronic feedback. Providing feedback after a negotiation ends can improve the performance in future negotiations (Thompson, 2022). The post-negotiation phase focuses on the evaluation of the outcome (Kersten & Noronha, 1999) and thus also on the self-reflection of the finished negotiation; this will be the focus of this research contribution.

The cognitive fit theory (Vessey, 1991) investigates the effects of graphical and tabular representations in problem-solving tasks. If the information emphasised in the representation type matches the required information to solve the task, the performance of the task will be improved (Vessey, 1991). Therefore, feedback should be provided in a way that fits the task. Initial investigations for feedback elements have already been conducted, e.g. by designing gamified elements, such as rankings and pareto graphs (Schmid & Schoop, 2022) or by assessing several feedback mechanisms to support the participants during the negotiation process, i.e., in the preparation, in the negotiation, and in the post-negotiation phases (Meyer et al., 2020). By negotiating electronically the offers are automatically documented and thus can be used to analyse the ended negotiation (Holtom & Kenworthy-U'Ren, 2006). Providing support (such as feedback elements) to participants in a training for self-assessment and for peer assessment, has not yet widely been reported in the feedback literature (Panadero & Lipnevich, 2022).

Personalised negotiation training can lead to improved acquisition of skills and fairer negotiation outcomes (Melzer & Schoop, 2016) and to improved negotiation outcomes (Johnson & Gratch, 2022). A system can provide learning content for personalised learning (Melzer & Schoop, 2015), e.g., assisting participants in self-reflection. Further, the participants are responsible to manage their learning process and to achieve their learning goals (Melzer & Schoop, 2015). To the best of our knowledge, there exists no elaborated system that support such self-reflection. Therefore, our aim is to design an artefact that support self-reflection in e-negotiation. To do so, possible feedback elements have to be designed and evaluated as a first step:

RQ 1: Which feedback elements can support participants to reflect upon their finished e-negotiations?

To create feedback elements for training purposes, it has to be determined how participants learn. The experiential learning model by Kolb (1976) describes an individual's learning process and can be applied in negotiation training (Lewicki, 1997; Melzer & Schoop, 2016). The model is defined as a cyclic process that comprises four phases (adapted to negotiations): concrete experience (negotiation cases), observation and reflections (debriefing), formulation of abstract concepts and generalisations (lectures and readings), and testing implications of concepts in new situations (personal goal setting) (Kolb, 1976; Lewicki, 1997). Since the current paper focuses on supporting participants to reflect on their finished negotiations in a training (cf. debriefing phase of the learning cycle), various methods can be applied, e.g., debriefings of the negotiations, feedback, usage of tools, or written analyses (Roloff et al., 2003). As Negoisst (Schoop, 2021; Schoop et al., 2003) already provides some information in a tabular-like manner, our research aims to provide graphical feedback for personalised training in terms of participants' negotiation styles.

Secondly, the identified feedback elements have to be examined whether they support the participants improving their negotiation skills in self-reflection. Since personalised training can improve skills (Melzer & Schoop, 2016), the feedback elements have to provide information about the specific negotiation. To analyse which skills can be improved by these feedback elements, the second research question is formulated:

RQ 2: Which skills support the self-reflection of participants through using the designed feedback elements in the post-negotiation phase?

2. Background

2.1. TKI Styles

The Thomas Kilman Conflict Mode Instrument (TKI) as a tool to support reflection and analysis has been used in training (Holtom & Kenworthy-U'Ren, 2006) and is used in the current paper to analyse individual conflict styles that show negotiation strategies. TKI describes individual conflict behaviour using two dimensions – assertiveness and cooperativeness (see Figure 1).

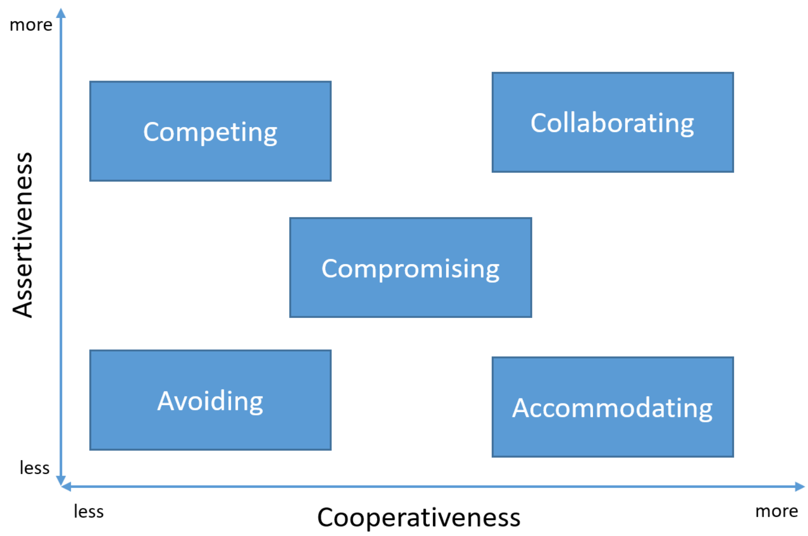


Figure 1: TKI Styles (Thomas & Kilmann, 1976, 2008)

Assertiveness is defined as how an individual is concerned with their own interests; cooperativeness is defined as how an individual is concerned with their partner's interests. Different relations between assertiveness and cooperativeness lead to the following five conflict styles (Thomas & Kilmann, 1976, 2008). Since these conflict styles can be stated as negotiation styles (Ganesan, 1993), they are defined as negotiation behaviour in our research.

2.2. Negotiation skills

Personalised training can improve skills (Melzer & Schoop, 2016). Consequently, it has to be examined which skills can be improved by the feedback elements. Thus, negotiation skills have to be defined which were adapted from Meyer et al. (2020) and consolidated in Table 1 focusing on self-reflection in the post-negotiation phase.

Table 1. Negotiation Skills for Self-Reflection

Negotiation Skills	Description
Adaptivity	Adapting e.g., negotiation behaviour through improved understanding of the negotiation partner (ElShenawy, 2010).
Communicativeness	Sharing information to the partner to decrease confusion and misinterpretation (Lewicki et al., 2015).
Confidence	Acquiring as much knowledge as possible for the negotiation to be well informed (Lewicki et al., 2015).
Effectiveness	Identify, prioritise, set, and achieve objectives stated for future negotiations (Lewicki et al., 2010).
Empathy	The ability of building on self-awareness, understanding the feelings of others and considering their views in formulating messages (Lewicki et al., 2015).
Pragmatism	The ability of understanding various meanings of syntax, semantics, and communication style, concerning the intention of additional, hidden information (Lewicki et al., 2010).
Preparedness	Achieving an understanding of goals and interests of oneself and the partner (Lewicki et al., 2010).
Rationality	The ability to reduce irrationality and avoid decision biases (Lewicki et al., 2010).
Strategic	The ability to plan effectively and to set goals (Lewicki et al., 2010).

3. Methodological Approach

The overall research approach is a design science research approach based on Gregor and Hevner (2013). Due to the lack of designed elements providing support in a training for self-assessment and for peer assessment (Panadero und Lipnevich 2022), potential feedback elements were designed based on the literature and adapted for self-reflection in the post-negotiation phase into an artefact. To evaluate the artefact, feedback elements were created and evaluated as a model, as well as quantitative data, and qualitative data were collected (see Figure 2).

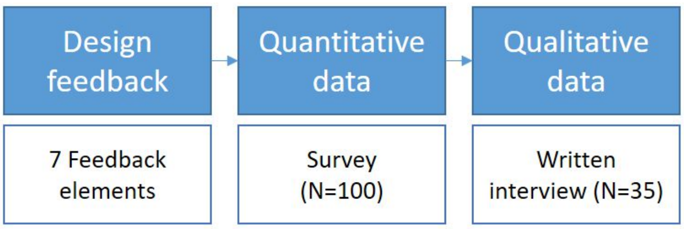


Figure 2: Methodology

Based on Meyer et al. (2020), we designed seven feedback elements to support negotiators in evaluating their finished negotiations in a NSS. NSSs are communication and information systems supporting negotiators during the negotiation process by gathering information, structuring information, and generating alternatives for the user (Bichler et al., 2003). The feedback elements are designed to improve negotiation skills (cf. Table 1) and to improve self-reflection in the post-negotiation phase.

We used a convergent mixed method approach (Creswell & Plano Clark, 2018). To do so, we conducted a survey (quantitative) by asking questions based on a 7-likert scale (1 = totally disagree, 7 = totally agree) with 100 students (43 male, 56 females, 1 gender unknown; with a mean age of 24.19) from European universities after they had participated in a negotiation experiment using the NSS Negoisst (Schoop, 2021; Schoop et al., 2003) to identify whether the designed feedback elements would have been useful for self-reflection on behaviour in the post-negotiation phase.

Secondly, we identified the top four feedback elements from the quantitative data followed by written interviews, i.e., open answers in a survey, with a focus group (n=35) to investigate in-depth whether these elements could assist the participants in self-reflecting and improving their negotiation skills. The focus group consists of a subgroup of the participants who completed the quantitative survey. We will name them S1-S35. They were asked for each of the four feedback elements to discuss 1) whether the feedback element could have assisted them in the reflection after the negotiation; if so, how and why; if not, why not and what should be improved; 2) which negotiation skills (cf. Table 1) could be improved by the element. Finally, we compared the quantitative results with the qualitative results.

4. Survey Results

Seven feedback elements were designed and considered in a survey. For each feedback element, we asked whether it supports the participant (1) in the post-negotiation phase and (2) in increasing their self-reflection.

4.1. Designed and Ranked Feedback Elements

The four feedback elements that were considered most be useful in the post-negotiation phase for self-reflection (see Table 2) and will be explained in more detail. The remaining feedback elements are (5) *achieving aspiration level (AL) and reservation level (RL)*, (6) *goals for issue selections*, and (7) *dashboard*. Feedback element (5) aims to assist in identifying the zone of possible agreements of received offers by (de)selecting issues to identify which selection is suitable for achieving AL/ RL. Feedback element (6) considers the selections for each issue in each message sent by the negotiation partner to enable evaluating whether an issue in an offer suits the defined goal. Feedback element (7) considers an overview of the current negotiation by specifying the duration and status of the negotiation (e.g., ongoing, accepted), the current utility value, whether AL and RL are reached, and recommendations to achieve the goals.

Table 2. Results Feedback Elements

Rank	Name	Min	Max	Mean Ended Negotiation (SD)	Mean Self-Reflection (SD)
1	TKI evaluation	2	7	5.61 (1.37)	5.63 (1.26)
2	Behaviour evaluation	1	7	5.18 (1.41)	5.70 (1.30)
3	History graph	1	7	5.11 (1.41)	5.38 (1.15)
4	Feedback from partner	1	7	5.05 (1.55)	5.52 (1.62)
5	Achieving AL and RL	1	7	4.74 (1.44)	4.89 (1.35)
6	Goals for issue selections	1	7	4.60 (1.49)	4.90 (1.33)
7	Dashboard	1	7	4.09 (1.76)	4.82 (1.53)

TKI Evaluation

This feedback element considers the overall conflict style of the participant (Thomas & Kilmann, 1976, 2008) and how the participant behaved in each offer. Figure 3 show triangles depicting the participant's own behaviour. The applied negotiation style in each offer is shown as a dot. For each of the five TKI styles the percentile according to Shell (2001) is calculated to define whether one's style is in the top 25%, middle 50% or bottom 25%. Figure 3 shows an example of a competing-compromising style with collaborating interests (triangles). The negotiator acted in a strongly competitive manner in the first two offers.

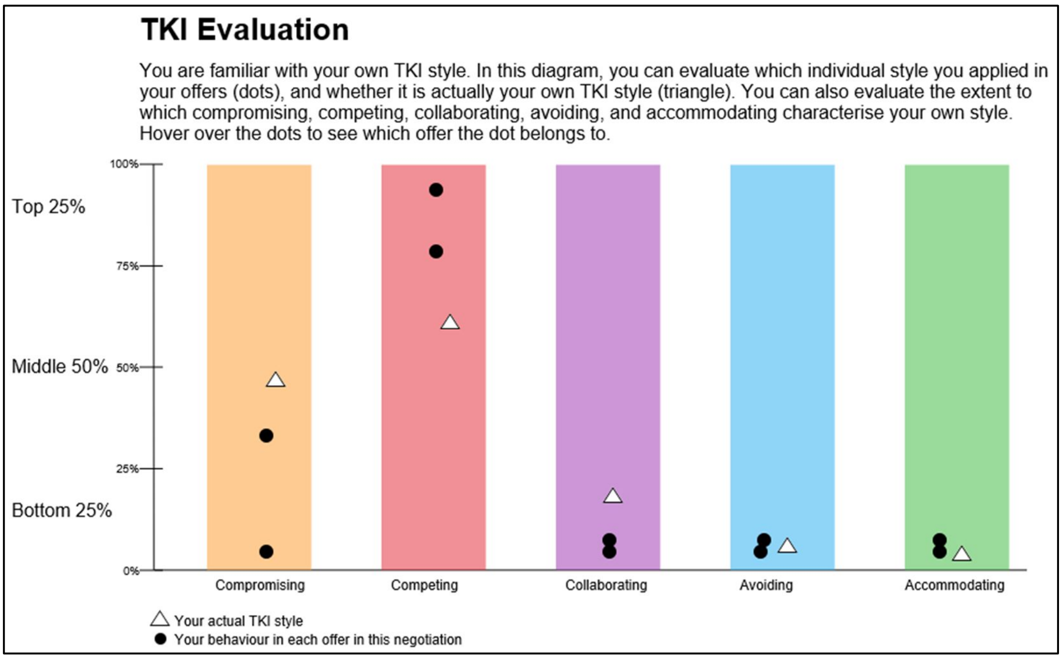


Figure 3: TKI Evaluation

Behaviour Evaluation

The feedback element *behaviour evaluation* evaluates a participant's behaviour during the negotiation by rating their own offers on a 7-likert scale (from totally agree to totally disagree) as being compromising, avoiding, collaborating competing and/or accommodating. Dependencies between behaviour types are not considered, rather all identified characteristics in the sent offers (text and concessions) are reviewed. In Figure 4, the participant had a high compromising, collaborating, and accommodating behaviour without avoiding and competing in the early phases of the negotiation. The behaviour changed over time with a decrease of accommodating, compromising and collaborating behaviour and an increase in competing behaviour. There are multiple reasons for those changes, e.g., the partner's attitude or offers, as well as changed goals.

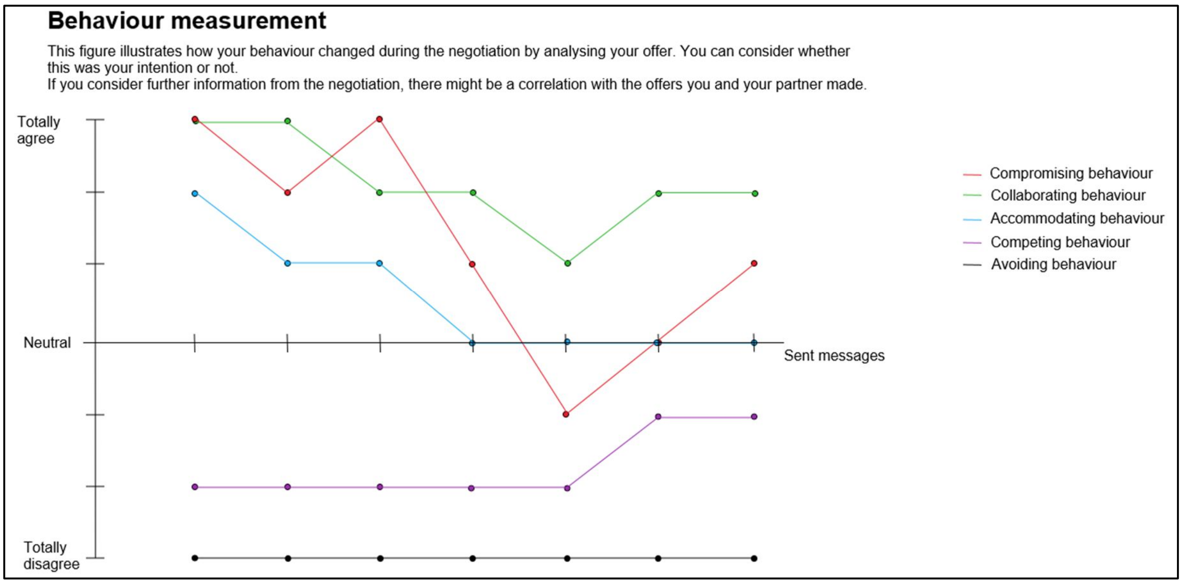


Figure 4: Behaviour Measurement

History Graph

The *history graph* as a feedback element considers the utility values of each offer as well as the participant's aspiration level and reservation level (see Figure 5). Participants can use this element to evaluate whether an offer meets their defined goals.

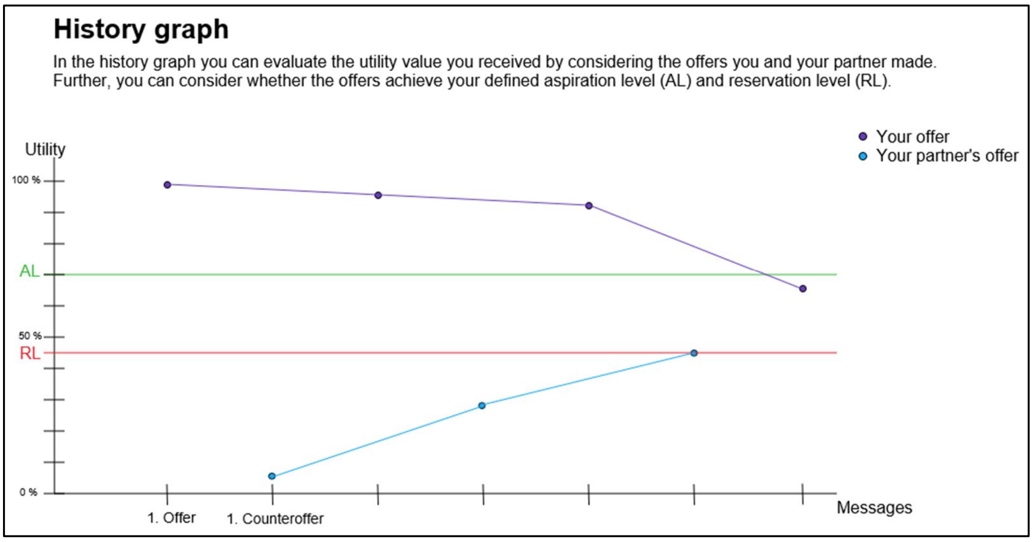


Figure 5: History Graph

Feedback from partner

This feedback element is especially designed to be considered when the negotiation ended to provide feedback to the partner (see Figure 6). Such feedback requires each participant to reflect on the negotiation and also shows how the partners perceived each other.

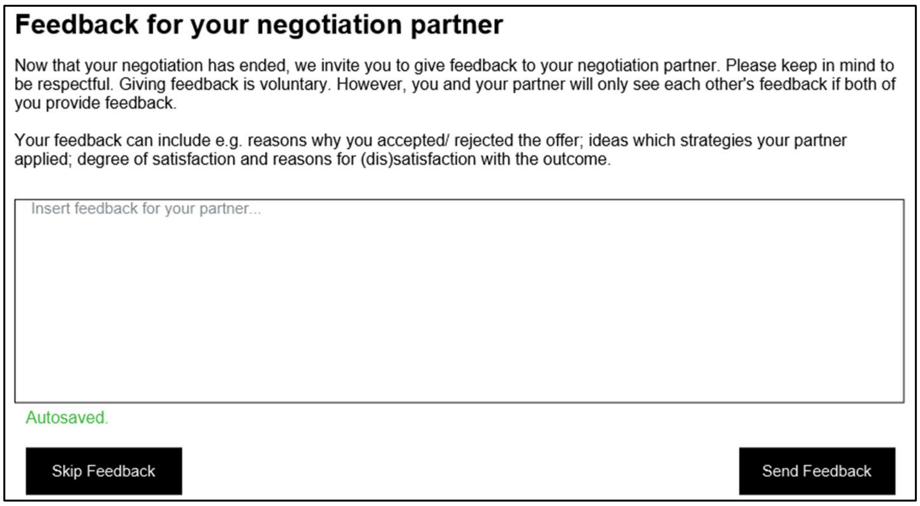


Figure 6: Feedback from Partner

4.2. Improving Confidence, Understanding, and Outcome

Further, the participants were asked whether the feedback elements would support their confidence and understanding and the negotiation (Table 3). All four elements were not taken to improve the outcome significantly. All four elements resulted in a mean between neutral and slightly agree to improve the participants' confidence in future negotiations. Whether the feedback element would support to better understand its partner behaviour in future negotiations, *feedback from partner* resulted in a mean between neutral and slightly agree, *history graph* and *behaviour evaluation* in a mean neutral, and *TKI evaluation* in a mean slightly disagree.

Table 3. Survey Results

Supporting	Name	Min	Max	Mean	SD
Outcome	Feedback from partner	1	7	4.82	1.424
	Behaviour evaluation	1	7	4.76	1.334
	History graph	1	7	4.64	1.382
	TKI evaluation	1	7	4.33	1.531
Confidence	Behaviour evaluation	1	7	4.71	1.297
	Feedback from partner	1	7	4.59	1.393
	History graph	1	7	4.58	1.312
	TKI evaluation	1	7	4.13	1.680
Understanding	Feedback from partner	1	7	4.5	1.521
	History graph	1	7	4.38	1.516
	Behaviour evaluation	1	7	3.81	1.680
	TKI evaluation	1	7	3.09	1.602

The highest ranked feedback elements in Table 2 show the negotiation behaviour (1 and 2), the change in utility values over time (3) and the perception by the partner (4). The element with feedback given by the partner (namely *feedback from partner*) gives insights in the partner's evaluation of the other's behaviour. This is limited to the personal styles of all negotiators, defined goals, and perceptions of the negotiation partner. However, it can indicate the reasons for how the negotiation ended.

The four feedback elements ranked as highest in self-reflection and usefulness after the negotiations ended tend to be seen as improving the outcome in future negotiations, as well as improving the confidence of the participant. Only direct *feedback from the partner* and the *history graph* seems to support the participant to better understand their partner and their partner's preferences. Since *behaviour evaluation* and *TKI evaluation* mainly focus on the participant's behaviour or TKI, they do not tend to support a better understanding of the partner.

Considering the ranking, *TKI evaluation* ranked highest regarding post-negotiation and second highest in self-reflection. However, based on the outcome, confidence, and understanding of the partner *TKI evaluation* is ranked as lowest. *Behaviour evaluation* is ranked as second highest in post-negotiation and highest in self-reflection, as well as highest in confidence and second highest in outcome. *Feedback from partner* is ranked fourth in post-negotiation and third in self-reflection, but highest in outcome and understanding, and second highest in confidence. *History graph* is ranked third in post-negotiation and fourth in self-reflection, as well as third in outcome and confidence; and as second in understanding. Thus, the ranking of the feedback elements varies regarding to their purposes, i.e., the simultaneous usage of various feedback elements to enable comprehensive reflection on the negotiation in the post-negotiation phase should be considered.

5. Written Interviews

The second evaluation was conducted via written interviews. This chapter consolidates the interview results by presenting whether self-reflection is supported, by discussing possible improvements, and by considering skill improvements.

5.1. Evaluating Feedback Elements

The participants were asked to discuss whether the feedback elements can support them reflecting on their finished negotiations.

The *feedback from the partner* is considered to be helpful in reflecting their negotiations by 16 of 35 participant if context-based feedback is received (S5, S6). Further it allows "[…] sharing perspectives, discussing decision rationales, and providing constructive feedback on the negotiation process." (S7). 13 participants are ambivalent; they recognise the potential by receiving feedback from a human partner but state that the usefulness of the element depends on the quality of the feedback which depends on the negotiation partner, i.e., how they behave in conflict situations (S10) and how much information is received by the partner (S34). However, the form of a free-text field is seen to be too unstructured and subjective (S1). Multiple participants mention that the feedback has to be constructive to be useful for reflection. Six participants state that the feedback element is not beneficial as their biggest concern is the dependency on whether they receive constructive feedback (S3, S18), e.g., some might perceive the negotiation (even in training) as a competition and thus do not provide open and honest feedback (S14). Further "[…] especially in […] failed negotiations, there will be cases in which no constructive dispute follows. Also, in my experience, dissatisfied people tend to write feedback rather than satisfied ones […]" (S29).

24 of 35 participants state the feedback element *history graph* to be useful for self-reflection. The existing history graph in Negoisst is enriched by adding the AL and the RL. The element is described to be "[…] simple and easy to understand […]" (S14) and further "[…] a powerful tool for users to analyze, reflect, and refine negotiation skills for future scenarios. […]" (S7). It can be used to deduce when and to what extent concessions are made (S28). Further three participants claim that the element is only useful to reflect and adjust the strategy during the negotiation (S12, S13). Five participants recognise its potential, however they criticise the necessity of stating AL and RL as they were aware of their preferences (S5, S19). For the remaining three participants the element just visualises the past negotiation and the utility value (S4, S8) and thus "[…] does not play a big role in reflection after the negotiation […]" (S8).

25 of 35 participants state that the feedback element *behaviour evaluation* supports them in reflecting their negotiation. In particular the focus on the timeline is mentioned, i.e., to see when the behaviour changed and where the behaviour could be improved in the negotiation (S20). Further, the element enables to identify why the behaviour changed (S10) and thus negotiators can "[…] understand the outcome of the negotiation in relation to [the] behaviour […]" (S13), as well as considering the effectiveness of an offer and the influence on the behaviour (S22). Four of 35 participants indicate the potential of the element, but also mentioned some drawbacks such as missing information to support the graph (e.g., why values changed) (S29) or factors influencing the outcome of the graph (e.g., emotions, attitudes) (S33). Six participants state that the current designed feedback element cannot support their self-reflection due to an overloaded (S19) and confusing (S2) design. Further, it is difficult to obtain valuable information by comparing several negotiations or to reflect effectively on failed negotiation (S3) since "[…] there is no comparison with a reference value […]" (S12), such as the personal TKI style.

22 of 35 participants consider the feedback element *TKI evaluation* to be useful in supporting self-reflection particularly through the lean design of the element which "[…] avoids overloading the user with excessive figures and impressions […]" (S7) and enables novices to identify their individual style and how it could be improved (S23). The element ensures to reflect whether the participant behaved as intended and if not, it states in which offer and to what extent it deviates from the style planned and the own TKI style (S2). Ten participants are ambivalent, i.e., they consider the potential of the element by identifying strengths and weaknesses of their

own behaviour (S1), but have difficulties in understanding the design of the element (S29) due to the missing connection between the negotiation style and the success of individual offers (S1), missing a retrospective in general (S24) or in "[…] negotiations skills or [in] the development of more effective offers […]" (S1). One participant mentions that the feedback element could lead the negotiators to consider only their applied TKI style in previous negotiation(s) rather than adapt their behaviour to the current situation (S12). Two participants state that the element could not assist them in self-reflection (S19); while for one participant the focus on TKI is too strong as TKI has some weaknesses (S18).

5.2. Improvements

Based on the discussed weaknesses of the feedback elements, the participants were asked to provide suggestions for improving the elements to assist them in self-reflection.

25 participants suggest improvements for *feedback from the partner*. Nine participants state that the feedback should be mandatory instead of voluntary to ensure that both receive feedback (S1, S13, S14, S21). Twelve participants suggest structuring the form of providing feedback by predefined open and closed questions (S14). Further some propose that scales or star ranking could be added to quantify the feedback (S1, S6, S15). Seven state that the partner should answer questions, e.g., by enabling to add questions (S24) or a chat function to clarify misunderstandings (S25, S27). The written feedback should be shared even if the partner did not provide feedback to foster a culture of feedback (S26). Some rather prefer to provide the feedback face-to-face if possible (S3, S29).

Fourteen participants suggest improvements for the *history graph*. Eight of them want the utility values of the partner to be displayed (S5, S16, S29, S32). Further, the results of other participants who already archived a result could be plotted (S16), e.g., as an average for comparison (S17). The element should additionally be displayed during the negotiation (S1), and it should be enabled to adjust AL and RL at any time (S14). Three like to see a history graph for each issue individually (S20, S24) either by hovering over the utility value (S2), or including separate history graphs (S20, S24). The history graph should include a time axis, i.e., to frame the time passed between offers (S15).

Seventeen participants suggest improvements for the *behaviour evaluation*. Eight of them suggest including the behaviour values of the partner (S24, S33, S35) as the partner's style have an influence on the own behaviour (S1). Seven state that further values should be added to enable a more profound analyses, such as personal TKI style (S1, S5, S12), analyses of previous negotiations (S2), utility values of the offers (S6, S24, S25), and set goals (S25). For five participants, the labelling should be improved by colour-coding the offers based on the styles' colour (S21), by fully labelling the y-axis (S19) and/or changing the parameter from 0 (not agree) to 100 (totally agree) (S23). Four participants like to receive more profound information for each identified behaviour (S2, S17), suggestions when to change the strategy to conclude the negotiation (S17), and whether they achieve their goals (S17). Further information could be added by linking the dots with the corresponding offer (S21) or hovering over the dots (S29). Trends could be analysed, e.g., how the participant behaves in certain time frames in the negotiation (S4).

Eighteen participants suggest improvements for the *TKI evaluation*. Ten participants consider improving the design of the dots by using numbers (S1, S17), different colours (S2, S19) or adjusting the information richness by linking the dots with the corresponding offers (S7, S11, S30). A filter mechanism (e.g., checkboxes) enables the participants to decide which offers they want to consider in-depth (S30, S34). Seven participants also want to evaluate their partner's behaviour (S5, S7), as it has an influence on the own behaviour (S1, S25) as well as like to compare the styles (S33). Two participants state that previous negotiations should be considered by 1) adding an average behaviour of previous negotiations into the graph of the currently ended negotiation to consider whether one behaves on average regarding to the previous negotiations (S2); 2) analysing whether the behaviours in previous negotiations changed over time (S15); and 3) a list of personal applied strategies successfully used in previous situations (S15). Additional factors should be included, such as providing concrete examples how the own TKI could look like in offers (S11), the own utility values per offer and style (S6, S25), and an average behaviour during the negotiation to consider if one behaves on average align to the TKI (S21). Further information could be added by including explanations how the behaviour affects the outcome (S7), how different expressions of behaviours could lead to different outcomes in certain phases of the negotiation (S17), how the behaviour resulted from the given offers (which factors) (S2, S24), as well as how a certain offer differentiate to the TKI style (S20). One participant likes a chronological structure how the negotiation styles have changed over time (S7).

5.3. Skill improvements

Further, the participants were asked to discuss which skills of Table 1 could be improved by the feedback elements. 29 of 35 participants discussed whether the feedback elements could support skill improvement. Figure **7** consolidates how often a skill is mentioned to be improved by a certain feedback element. The participants were allowed to name as many skills as they found useful. As the number of skills mentioned for each element could vary, the number of skills mentioned was set in relation to the total number of skills mentioned per element.

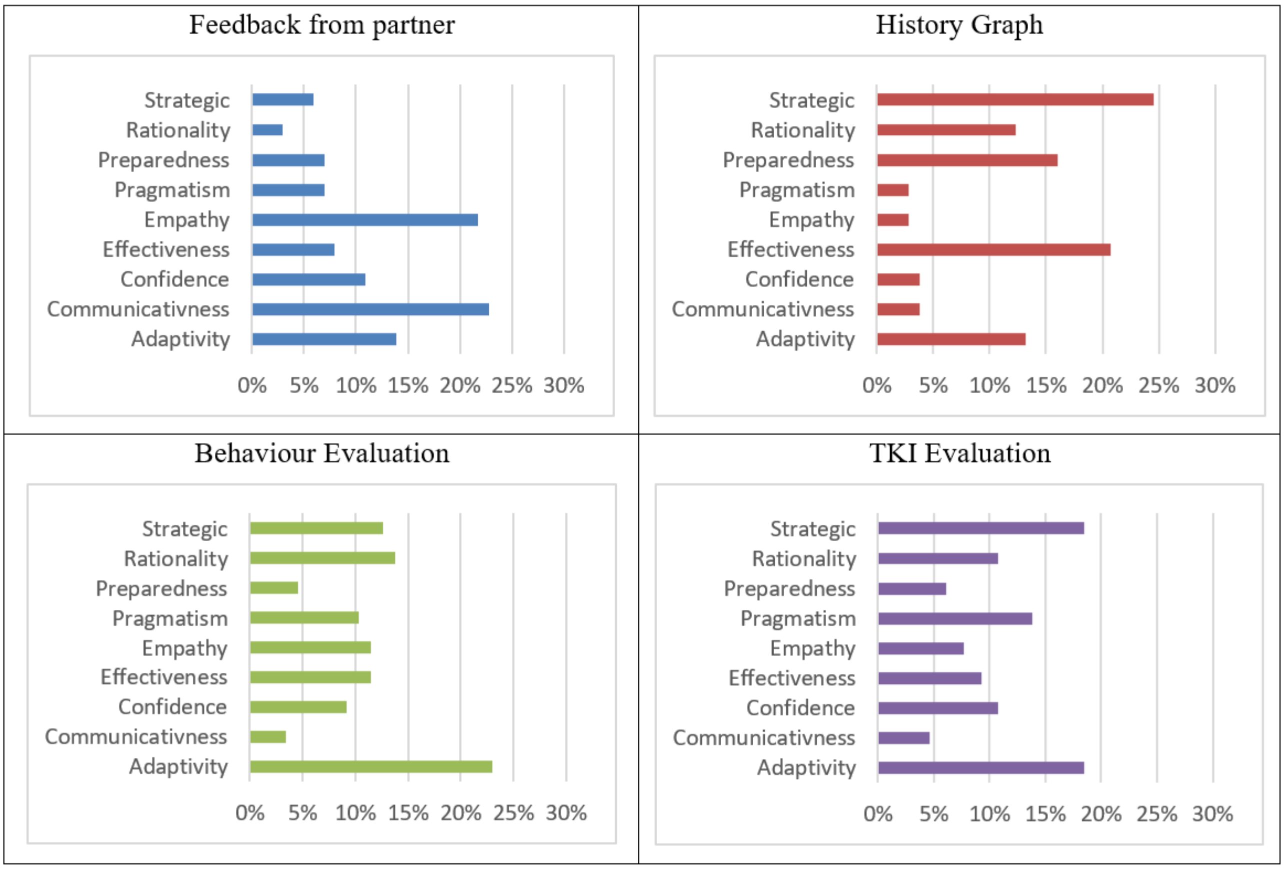


Figure 7: Skill Improvements

For *Feedback from the partner* the skills communicativeness (22.77%) and empathy (21.78%) are stated to be improved; followed by adaptivity (13.86%) and confidence (10.89%). The skills effectiveness (7.92%), preparedness (6.93%), pragmatism (6.93%), strategic (5.94%), and rationality (2.97%) are stated as lowest to be improved. One participant states that none of the given skills could be improved.

The participants state that the *history graph* could improve the skills strategic (24.53%) and effectiveness (20.75%); followed by preparedness (16.04%), adaptivity (13.21%), and rationality (12.26%). The skills confidence (3.77%), communicativeness (3.77%), pragmatism (2.83%), and empathy (2.83%) are stated to be improved as lowest.

Behaviour evaluation is stated to improve the skill adaptivity (22.99%); followed by rationality (13.79%), strategic (12.64%), empathy (11.49%), effectiveness (11.49%), and pragmatism (10.34%). The skills confidence (9.2%), preparedness (4.6%), and communicativeness (3.45%) are stated to be improved as lowest. Three participants state that none of the given skills could be improved.

TKI evaluation is stated to improve the skills strategic (18.46%) and adaptivity (18.46%); followed by pragmatism (13.85%), rationality (10.77%), and confidence (10.77%). The skills effectiveness (9.23%), empathy (7.69%), preparedness (6.15%), and communicativeness (4.62%) are stated to be improved as lowest. One participant states that none of the given skills could be improved.

6. Discussion

When analysing the ranking in Table 2, only the top four elements were considered further, as the remaining elements are either (1) describing the offers in more detail which would be more useful during the negotiation to react to offers or (2) providing a general overview of the negotiation which are already available in the system on different pages in Negoisst and thus would be repetitive.

The results in the written interviews for *feedback from the partner* emphasise with the overall ranking (fourth rank). In contrast, the survey results in outcome (first rank), confidence (second rank) and understanding (first rank) are contradicting, which could rely on a missing in-depth reflection during the survey. The main concern in the written interviews was the reliability of the partner's willingness to provide constructive feedback which is valid and thus should be counteracted by e.g., mandatory participation and structured questions. Since mandatory participation could be seen as external regulation of extrinsic motivation (Ryan & Deci, 2000), the effects have to be examined further.

History graph was ranked as third or second in the survey regarding to outcome, understanding and confidence, and second in the written interviews. Additional, participants stated using the element during the negotiation instead of after the negotiation would improve the usefulness of the element. As a similar version for during the negotiation already exists in Negoisst, the improved functionality can also be added during the negotiation. The information richness of the history graph could be expanded by the suggestions; however, they would simultaneously increase the complexity of the feedback element. The suggestion mentioned most was to add the utility values of the partner which is most helpful as feedback (Thompson, 2022). Despite some participants wanting to gather more experiences (Schmid et al., 2020) and trying out different behaviour, repeating negotiations would no longer be possible as the partner's private information would be revealed. Only if case studies can be created autonomously, one can consider providing partners' preferences in the post-negotiation phase during the training.

Behaviour evaluation was ranked second in outcome, first in confidence, and third in understanding in the survey which align to the results in the written interviews. The main concern was that information are missing regarding to relation between behaviour result and offers. The missing relation could be added by linking the dots in the figure with the corresponding offers. Besides this, behaviour evaluation was mentioned as useful in the written interviews (25 of 35 participants).

TKI evaluation was ranked as highest in the survey and only three participants in the interview considered it as not useful for self-reflection. The main improvements are designing specific and more precise information, such as the relation to the corresponding offers. To counteract information overload in the graph checkboxes as filter mechanisms should be added. The suggestion to get to know the partner's behaviour rely on the risk of self-revelation (cf. partner's utility values in history graph), i.e., the partner would benefit from the information when repeating negotiations. Only if the partners are agents with randomised behaviour (not predictable by knowing previous behaviours), the partner's behaviour can be provided after the negotiation ended in training. A further improvement is to provide a behaviour value based on the average of the behaviours of previous negotiations (S2). Since the negotiation can be quite contradictory and a participant can adapt their behaviour to specific situations, an average could only be calculated if similar negotiation settings were used to calculate the average. The current setting of the training component is designed to provide various situations to train into the broad, i.e., in order to calculate the average of the negotiations, negotiations with similar case studies have to be applied.

The behaviour in behaviour evaluation is classified by the negotiators by means of self-assessment. The actual TKI styles used during the negotiation, as presented in the feedback element TKI evaluation, are determined by the system. Apart from this, *behaviour evaluation* and *TKI evaluation* are similar. Multiple participants suggested a consolidation of those two feedback elements. Others like the lean design of the TKI evaluation element. Thus, both elements should be used in the artefact with the relation as well as the separate focus of them by adding explanations.

The results in chapter 5.3 show that the feedback elements support different skills, which can be explained by the different focuses: The history graph considers the set goals on preference level; feedback from the partner considers how the participant's behaviour is received by the partner, behaviour evaluation supports self-reflecting each of the participant's offer, and TKI evaluation considers how the participant's behaviour is received by the system (a neutral party). Based on the results (cf. Figure **7**), the usefulness of improving skills by the elements varies for each participant. To enable a comprehensive skill improvement for all participants, all four feedback elements should be considered in the artefact.

Behaviour evaluation, history graph, and TKI evaluation were considered to support over 60% of the participants, and thus, they should be part of the artefact for self-reflection. Suggested improvements could increase the usefulness of these elements. Feedback from the partner was considered to support self-reflection only 45.71% of the participants, but further 37.14% of the participants suggested improvements which would increase the support of self-reflection. Since the element feedback from the partner has a high potential to increase the usefulness, the suggestions have to be considered. It has to be noted, that whether a training form motivates the participant, depends on the individuum and the training purposes (Mathieu & Martineau, 1997), i.e., the usefulness of a feedback element relies on it. Reflecting after the negotiation is an important aspect to be more effective and be more likely to reach integrative agreements (Thompson, 2022) and based on the high potential of all four elements as well as the relation with each other, all four elements should be implemented as an artefact to support self-reflection.

7. Outlook and Future Work

The aim of the paper is to design an artefact for self-reflection in an e-negotiation training. To do so, seven feedback elements were designed based on existing literature to be evaluated regarding to their usefulness of supporting self-reflection after a negotiation. Four feedback elements, namely feedback from the partner, history graph, behaviour evaluation, and TKI evaluation, were considered to be useful and thus were evaluated in-depth via written interviews. All four elements were considered to support self-reflection and improve various skills. Suggestions to improve the elements to better support self-reflection were discussed. As the elements have different focuses on self-reflection and support different skill improvements, all four feedback elements should be provided together as an artefact in the post-negotiation phase.

This research has some limitations. The written interviews are a form of asynchronous interaction. Direct reactions are not possible although the interviewees were able to reach out via an e-learning platform and via email to receive answers to their questions. Further, the participants were familiar with the task of written interviews since they already participated in previous ones. As the feedback elements have so far only been designed, they were only available after the negotiation, i.e., the actual usage and possible measuring of improvements by using those feedback elements are not investigated. A potential correlation between personality and particular types of feedback elements were not considered. We address this issue by designing different types of feedback elements, but potential correlations need to be analysed in-depth during future research.

As next steps, the artefact will be integrated in the NSS Negoisst and further evaluations with control groups will be conducted regarding the actual skills improvements though the feedback elements.

References

M. Bichler, G. Kersten, S. Strecker, Towards a Structured Design of Electronic Negotiations. *Group Decision and Negotiation*, *12*(4), 311–335. https://doi.org/10.1023/A:1024867820235, 2003.

J. W. Creswell, V. L. Plano Clark. *Designing and conducting mixed methods research* (3rd ed.). Sage, 2018.

S. Ganesan, Negotiation Strategies and the Nature of Channel Relationships. *Journal of Marketing Research*, *30*(2), 183. https://doi.org/10.2307/3172827, 1993.

S. Gregor, A. R. Hevner, Positioning and Presenting Design Science Research for Maximum Impact. *MIS Quarterly*, *37*(2), 337–355. https://doi.org/10.25300/MISQ/2013/37.2.01, 2013.

B. C. Holtom, A. L. Kenworthy-U'Ren, Electronic Negotiation: A Teaching Tool for Encouraging Student Self-Reflection. *Negotiation Journal*, *22*(3), 303–324. https://doi.org/10.1111/j.1571-9979.2006.00103.x, 2006.

E. Johnson, J. Gratch, The Impact of Personalized Feedback on Negotiation Training. In A. Sinatra, A. C. Graesser, X. Hu, B. Goldberg, A. J. Hampton, & J. H. Johnston (Eds.), *Design Recommendations for Intelligent Tutoring Systems: Competency-Based Scenario Design* (9th ed., pp. 89–102). U.S. Army Combat Capabilities Development Command – Soldier Center, 2022.

G. Kersten, S. Noronha, Negotiation via the World Wide Web: A Cross-cultural Study of Decision Making. *Group Decision and Negotiation*, *8*(3), 251–279. https://doi.org/10.1023/A:1008657921819, 1999.

D. A. Kolb, Management and the Learning Process. *California Management Review*, *18*(3), 21–31. https://doi.org/10.2307/41164649, 1976.

R. Lewicki, Teaching Negotiation and Dispute Resolution in Colleges of Business: The State of the Practice. *Negotiation Journal*, *13*(3), 253–269. https://doi.org/10.1111/j.1571-9979.1997.tb00131.x, 1997.

J. Mathieu, J. Martineau, Individual and Situational Influences on Training Motivation. In J. K. Ford & S. W. J. Kozlowski (Eds.), *Series in applied psychology. Improving training effectiveness in work organizations* (1st ed.). Psychology Press, 1997.

P. Melzer, M. Schoop, A Conceptual Framework for Task and Tool Personalisation in IS Education. *ICIS 2015 Proceedings,* 2015.

P. Melzer, M. Schoop, The Effects of Personalised Negotiation Training on Learning and Performance in Electronic Negotiations. *Group Decision and Negotiation*, *25*(6), 1189–1210. https://doi.org/10.1007/s10726-016-9481-y, 2016.

M. Meyer, A. Schmid, P. Melzer, M. Schoop, Gamified Feedback in Electronic Negotiation Training. *UK Academy for Information Systems Conference Proceedings 2020*(19), 2020.

E. Panadero, A. A. Lipnevich, A review of feedback models and typologies: Towards an integrative model of feedback elements. *Educational Research Review*, *35*(5), 2022.

M. E. Roloff, L. L. Putman, L. Anastasiou, Negotiation Skills. In B. R. Burleson & J. O. Greene (Eds.), *LEA's communication series. Handbook of communication and social interaction skills* (pp. 801–833). L. Erlbaum Associates, 2003.

R. M. Ryan, E. L. Deci, Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology*, *25*(1), 54–67. https://doi.org/10.1006/ceps.1999.1020, 2000.

A. Schmid, P. Melzer, M. Schoop, Gamifying Electronic Negotiation Training - A Mixed Method Study of Students' Motivation, Engagement and Learning. *Proceedings of the 28th European Conference on Information Systems (ECIS),* 2020.

A. Schmid, M. Schoop, Gamification of Electronic Negotiation Training: Effects on Motivation, Behaviour and Learning. *Group Decision and Negotiation*, *31*(3), 649–681. https://doi.org/10.1007/s10726-022-09777-y, 2022.

M. Schoop, Negoisst: Complex Digital Negotiation Support. In D. M. Kilgour & C. Eden (Eds.), *Handbook of Group Decision and Negotiation* (Vol. 12, pp. 1149–1167). Springer International Publishing. https://doi.org/10.1007/978-3-030-49629-6_24, 2021.

M. Schoop, A. Jertila, T. List, Negoisst: A negotiation support system for electronic business-to-business negotiations in e-commerce. *Data & Knowledge Engineering*, *47*(3), 371–401. https://doi.org/10.1016/S0169-023X(03)00065-X, 2003.

G. R. Shell, Bargaining Styles and Negotiation: The Thomas-Kilmann Conflict Mode Instrument in Negotiation Training. *Negotiation Journal*, *17*(2), 155–174. https://doi.org/10.1023/A:1013280109471, 2001.

M. Ströbel, C. Weinhardt, The Montreal Taxonomy for Electronic Negotiations. *Group Decision and Negotiation*, *12*(2), 143–164. https://doi.org/10.1023/A:1023072922126, 2003.

K. W. Thomas, R. H. Kilmann, Thomas-Kilmann Conflict Mode Instrument. *Group & Organization Studies*, *1*(2), 249–251. https://doi.org/10.1177/105960117600100214, 1976.

K. W. Thomas, R. H. Kilmann, Thomas-Kilmann Conflict Mode: Profile and Interpretive Report, 2008.

L. L. Thompson. *The mind and heart of the negotiator* (7th ed.). Pearson, 2022.

I. Vessey, Cognitive Fit: A Theory-Based Analysis of the Graphs Versus Tables Literature *Decision Sciences*, *22*(2), 219–240. https://doi.org/10.1111/j.1540-5915.1991.tb00344.x, 1991.

joint conference group and negotiation DECISION support systems technology Porto 2024

24th International Conference on Group Decision and Negotiation &
10th International Conference on Decision Support System Technology
Human-Centric Decision and Negotiation Support for Societal Transitions

Negotiation Specific User Experience Metrics for Electronic Negotiation Platforms

Ulkar Garajali [1*]
[1] University of Vienna, Vienna, Austria
[*] ulkar.garajali@univie.ac.at

Abstract

The digital transformation has significantly impacted traditional negotiation processes, leading to the rise of electronic negotiation (eNegotiation) platforms. These platforms have evolved from basic online dispute resolution systems to advanced AI-powered negotiation platforms, playing a pivotal role in streamlining business negotiations in the modern age. Despite extensive research on their mechanics and outcomes, there is a noticeable gap in the literature regarding the user experience (UX) of these platforms. This paper underscores the importance of integrating UX principles into the development of eNegotiation platforms. Drawing on various research methodologies, it aims to propose quantifiable metrics specific to negotiation tasks conducted in virtual settings. These metrics can be applied in the evaluation of working eNegotiation platforms. A literature review will be conducted to identify existing metrics. To confirm the relevance of these identified metrics and gather further insights on negotiation goals and objectives, as well as key eNegotiation performance indicators for typical and expert users, representatives of these groups will be interviewed. Combining the results of the interviews and the literature review, a set of negotiation-specific metrics will be proposed. To assess the validity of these metrics, a high-fidelity eNegotiation platform prototype will be designed and evaluated. This evaluation will involve measuring the performance of the metrics through questionnaires. The results from these questionnaires will be used to propose a set of validated negotiation-specific metrics.

Keywords: electronic negotiations; user experience; user experience principals; user experience metrics; negotiation-specific metrics

1. Introduction

Negotiations, a foundational element of human interaction, have been extensively studied and analyzed for decades. Central to negotiations are parties with differing interests, each aiming to forge an agreement that meets their objectives while remaining acceptable to the other party (Pruitt & Carnevale, 1993). The journey to this agreement is shaped by various factors, including negotiation strategies, communication methods, and media (Putnam & Roloff, 1992).

This study focuses on a specific medium of negotiation communication: electronic negotiation (eNegotiation) platforms. Electronic negotiations refer to the use of software and internet-based platforms to facilitate conflict management and resolution processes (Kersten, 2010). These negotiations can take various forms, including auctions and bilateral bargaining tables, and are supported by a range of tools and software platforms (Mäkiö, 2004; Ströbel, 2003). The distinction between a negotiation tool and a negotiation platform lies in their functionality and scope. A negotiation tool, such as the CONTRACT NEGOTIATOR, Decision Station are decision-support tools designed to assist negotiators in identifying and evaluating issues during contract negotiations for enhanced utility in electronic environments. (Law, 2005; Vahidov & Kersten, 2004). In contrast, negotiation platforms, like the Invite, INSPIRE, and PLAMUN, are flexible and customizable e-negotiation software that allow for the creation of multiple negotiation protocols and the simultaneous running of different protocols (Law 2005; Srivastava 2003; Kersten & Cray, 1997; Schoop, 2021). Outside of academia corporations like IBM offer enterprise resource planning (ERP) software such as SAP Ariba where corporate buyers and sellers can negotiate multimillion deals.

These platforms are designed to support a wide range of negotiation activities and interactions, including multi-user negotiations (Srivastava 2003; Jennings 2005). Examining these examples we can deduce that negotiation tools are specific aids or applications that support certain tasks or steps within a negotiation process, while negotiation platforms provide an all-encompassing environment that hosts a full negotiation lifecycle, integrating various tools and services needed to conduct negotiations effectively.

eNegotiation platforms offer a range of services to their users, including personalization and customization of products and services, multi-issue negotiations, and negotiation automation (Chen, 2005; Kersten, 2004; Bui, 2005). These platforms also provide support for complex negotiations, such as communication support, decision support, and contract management (Schoop, 2010). They can be used in e-business negotiations, providing context-dependent advice and autonomous activities (Kersten, 2003).

The interest in eNegotiations remains relevant. Hernández (2021) and Dobrijević (2021) both highlight the increasing use of audio-visual communication technologies and negotiation support systems in online negotiations. Myoo (2022) and Schmid (2019) discuss the role of virtual worlds and gamified electronic negotiation training in enhancing negotiation skills. Spector (2022) and Dobrijević (2020) focus on the impact of information technology and the use of artificial intelligence in international and electronic negotiations. Batra (2022) and Mahajan (2020) propose the development of intelligent negotiation bots and chatbots to improve customer service and enhance the bargaining process in e-commerce. These studies collectively underscore the growing importance of electronic negotiation platforms and the potential for their continued evolution.

While there has been extensive interest in the mechanics, algorithms, and outcomes of eNegotiation platforms, there is a noticeable gap regarding their user experience (UX). The emphasis has typically been on the systems' effectiveness and efficiency, rather than on how users perceive and interact with these platforms. Notably, there is a lack of universally accepted metrics for evaluating the UX of negotiation platforms. While general UX metrics (such as usability and learnability) are applicable, negotiation-specific metrics are less established.

The recent surge in interest in a user-centric approach to online systems highlights the importance of ensuring that the User Experience (UX) of electronic negotiation (eNegotiation) platforms is carefully considered and optimized. This approach has the potential to bring significant benefits to both academic researchers and practitioners in the field. By focusing on user needs and preferences, eNegotiation platforms can enhance user satisfaction, increase platform adoption, and improve negotiation outcomes. Studies have shown that secure, user-friendly platforms not only maximize clients' utility and shorten negotiation times but also ensure data security during negotiations, underscoring the importance of UX in eNegotiation environments (Al-Jaljouli et al., 2018). Therefore, the integration of user-centric design principles in the development and refinement of eNegotiation platforms is crucial for their success and effectiveness.

In the literature review, many studies evaluate UX of online systems using dimensions and criteria (Aranyi & van Schaik, 2016; Liu et al., 2013; Alarifi et al., 2017; Huang et al., 2020; Tcha-Tokey et al., 2018). However, these studies often lack comprehensive UX evaluation models incorporating dimensions, criteria, and metrics. Establishing these components is crucial for organizations or stakeholders to conduct more specific evaluations of product or system usage.

2. Theoretical Framework

Negotiations are emotional undertakings (Shapiro, 2002; Druckman & Olekalns, 2008; Filipowicz et al., 2011). However, many platforms are designed more for utilitarian (logical or computational efficiency), without adequately considering the hedonic (emotional experience) aspects of negotiations. Different users have different negotiation styles and preferences. Platforms may not always account for this diversity, offering a one-size-fits-all solution that might not cater to the nuances of different users. Some UX evaluations might be conducted in controlled environments, which do not necessarily replicate the pressures and unpredictabilities of real-world negotiations. Not all platforms may be designed with accessibility in mind, excluding users with disabilities. Furthermore, cultural inclusivity is often an overlooked aspect.

Research indicates that platforms with intuitive interfaces and positive UX can significantly enhance user engagement (Onișor & Ionita, 2016). Engagement is crucial for negotiation platforms, where sustained attention and strategic thinking are required. A positive UX can bolster a user's confidence in both the platform and the negotiation process, encouraging trust and full utilization of the platform.

2.1. UX Evaluation in Online Systems: Metrics

The concept of UX and its evaluation has been extensively researched in usability and interaction design. Despite numerous contributions, there remains a lack of consensus on a standard definition of UX (Hassenzahl & Tractinsky, 2006; Sharp et al., 2007; Kieschnick, 2008; Pretorius et al., 2005; Rubinoff, 2004; Goddard, 2009). This ambiguity stems from UX's multifaceted nature, integrating elements from various fields (Mashapa & van Greunen, 2010). Hassenzahl and Tractinsky (2006) describe UX as a consequence of the user's internal state, system characteristics, and interaction context. In contrast, Sharp et al. (2007) emphasize UX's subjective aspect, focusing on the user's feelings during system interaction. The International Standard Organisation (ISO) offers a broader definition, viewing UX as the user's perceptions and responses resulting from the use or anticipated use of a product or system.

Measuring and monitoring UX, just like any organizational asset, is crucial (Sauro & Kindlund, 2005). Given UX's complexity, its evaluation methods are diverse, including expert inspection, user testing, and inquiry methods (Mashapa & van Greunen, 2010). Expert inspections involve experienced practitioners identifying design issues affecting UX, while user testing methods are performance-based, observing users as they interact with the system. Inquiry methods gather user opinions through post-test questionnaires and interviews.

In UX domain, metric is used as an indicator of the experience of the user while interacting with a product, system, or service. With UX metric, the researcher can discover a user’s feeling and experience with the evaluated product or discover an improvement in the product (Albert and Tullis, 2013).

Various metrics have been proposed for effective UX evaluation (Nielsen, 1994; Kieschnick, 2008; Pretorius et al., 2005; Rubinoff, 2004; Goddard, 2009; de Kock et al., 2009). These metrics address different UX aspects, from enjoyment of task completion to system error messaging and user control (Mashapa & van Greunen, 2010).

Albert and Tullis (2013) group these metrics into four types; performance metrics, issue-based metrics, self-reported metrics, and behavioral and physiological metrics. Hussain et al. (2018) categorize UX metrics into 3 types of metrics: interaction metric, self-reported metric, and emotion and stress metric.

This study hopes to derive the set of metrics specific to negotiation tasks that can be used for the evaluation of eNegotiation platforms’ UX from a large number of metrics available in UX evaluation literature.

3. Method

UX measurement methods are categorized into expert inspection, user testing, and inquiry methods (Tullis & Albert, 2008; Rubin & Chisnell, 2008; Banati et al., 2006). A range of studies have highlighted the value of interviewing both typical and expert users for user experience evaluation. Distinguishing experts from typical users, experts possess extensive domain-specific knowledge and superior problem-solving skills (Popovic 2000, Kolodner 1982). This expertise is particularly crucial in the use of complex technologies, where it has a greater impact on ease-of-use perceptions and technology use (Goeke 2016). Äijö (2001) found that non-expert evaluators can provide valuable insights into individual user reactions, while Boland (2014) emphasized the importance of aligning expert-derived user needs with system functions.

To evaluate eNegotiation platforms from a UX perspective, we will adopt the user-centric research paradigm proposed by Matamala et al. (2018) starting with profiling typical users of eNegotiation platforms through qualitative interviews. This approach ensures a deep, holistic understanding of users' genuine needs and expectations. The interviewees primarily hold positions in the sales and procurement departments of their respective organizations.

Following Albert & Tullis (2022) we will also conduct qualitative interviews with expert users to identify their expectations from an ideal eNegotiation platform. Subjects for this group will be chosen from members of the academic community studying group decision making and negotiations. These users often utilise eNegotiation platforms for academic data collection or negotiation training. They are also the ones

Based on the insight collected a high-fidelity prototype e-negotiation platform will be designed. Inspired by Sauro and Lewis (2016; 2012), we will incorporate quantitative methods to gather objective data on user behaviors and attitudes toward the newly designed eNegotiation platform prototype. Task success rate (measuring the percentage of tasks that users can complete successfully), time-on-task (recording the time users take to complete tasks), system usability scale (measuring user satisfaction with the platform through a standardized questionnaire) and net promoter score (assessing the likelihood of users recommending the platform to others) will be recorded for analysis.

Considering the diverse user base of e-negotiation platforms, inclusivity in our research approach is essential. Methodologies will be tailored to include user groups with distinct needs, such as seniors or those with limited tech proficiency.

4. Expected Contributions

A strong UX design often leads to quicker user onboarding and wider adoption of online systems (Coatta & Gosper, 2010). In negotiation, where cognitive effort is significant, a UX-centric platform can minimize extraneous cognitive load, allowing users to focus on the negotiation itself.

UX-centric studies are needed to provide actionable feedback on the platform's strengths and weaknesses, guiding iterative improvements to meet users' evolving needs. More specifically negotiations involving multi-attribute products, where preferences for individual attributes are refined as information is exchanged, can greatly benefit from quantifying user experience metrics. This approach allows negotiating parties to incrementally utilize knowledge as the negotiation progresses, leading to potentially more mutually acceptable configurations of the product and its price (Mia, Mudur, & Radhakrishnan, 2005). Yuan (2003) highlighted the benefits of multimedia communication, with text and audio being preferred over text alone. Bichler (2003) and Mitterhofer (2012) both stressed the importance of structured design and the integration of behavioral and analytic decision support in electronic negotiations.

Therefore, understanding the user needs and having a robust mechanism to ensure that those needs are met can ensure the continuous usage of a platform. By prioritizing user experience and incorporating both behavioral and analytic decision supports, these platforms can not only streamline the negotiation process but also facilitate a more effective decision-making environment.

Selected References

1. Alarifi, A., Alsaleh, M., & Alomar, N. (2017). A model for evaluating the security and usability of e-banking platforms. *Computing*, 99(5), 519–535.
2. Aranyi, G., & van Schaik, P. (2016). Testing a model of user-experience with news websites. *Journal of the Association for Information Science and Technology*, 67(7), 1555–1575.
3. Banati, H., Bedi, P., & Grover, P. (2006). Evaluating Web Usability from the User's Perspective. *Journal of Computer Science* 2 (4), 314-317.
4. Coatta, T., & Gosper, J. (2010). UX Design and Agile: A Natural Fit? *Queue*, 8, 50-58.
5. de Kock, E., van Biljon, J.& Pretorius, M., (2009). Usability evaluation methods: mind the gaps. ACM International Conference Proceeding Series Proceedings of the 2009 *Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists* (pp. 122-131). Vanderbijlpark, Emfuleni, South Africa : ACM New York, NY, USA .
6. Ebner, N. (2012). ODR: The next green giant. *Journal of Online Dispute Resolution*.
7. Elkins, A. C., Derrick, D. C., & Burgoon, J. K. (2013). Predicting users' perceived trust in embodied

conversational agents using vocal dynamics. *Group Decision and Negotiation*, 22(5), 801-823.

8. Filipowicz, A., Barsade, S., & Melwani, S. (2011). Understanding emotional transitions: The interpersonal consequences of changing emotions in negotiations. *Journal of personality and social psychology*, 101(3), 541.

9. Goddard, P. H. (2009). User Experience Metrics: Connecting the language of UI design with the language of business. *Human Factors International White Paper*.

10. Hassenzahl, M., & Tractinsky, N. (2006). User Experience – a Research Agenda. *Behaviour and Information Technology*, Vol. 25, No. 2 , 91-97.

11. Huang, Y., & Wilkinson, I. F. (2013). The dynamics and evolution of trust in business relationships. *Industrial Marketing Management*, 42(3), 455-465.

12. Huang, Z., Hong, Y., & Xu, X. (2020). Design and research on evaluation model of user experience on mobile terminal products. In *International Conference on Applied Human Factors and Ergonomics* (pp. 198–206).

13. Ivory, M.Y. (2003). Usability Testing Methods. In Automated Web Site Evaluation. Human-Computer Interaction Series (vol. 4). *Springer, Dordrecht*. https://doi.org/10.1007/978-94-017-0375-8_3

14. JamilHussain,WajahatAliKhan,TaehoHur,HafizSyedMuhammadBilal,Jaehun Bang, Anees Ul Hassan, Muhammad Afzal, and Sungyoung Lee. 2018. A Multimodal Deep Log-Based User Experience (UX) Platform for UX Evaluation. Sensors 18, 5(18 May 2018), 1622. https://doi.org/10.3390/s18051622

15. Katsh, E., & Rabinovich-Einy, O. (2017). Digital Justice: Technology and the Internet of Disputes. *Oxford University Press.*

16. Kersten, G. E., & Cray, D. (1997). Perspectives on representation and analysis of negotiation: Towards cognitive support systems. *Group Decision and Negotiation*, 6(5), 471-485.

17. Kersten, G. E., & Cray, D. (1997). Perspectives on representation and analysis of negotiation: Towards cognitive support systems. *Group Decision and Negotiation*, 6(5), 471-485.

18. Kersten, G. E., & Noronha, S. J. (1999). WWW-based negotiation support: design, implementation, and use. *Decision Support Systems*, 25(2), 135-154.

19. Kersten, G.E. (2003). The science and engineering of e-negotiation: an introduction. 36th *Annual Hawaii International Conference on System Sciences.*

20. Kieschnick, T. (2008). User Experience Metrics. Retrieved April 11, 2009, From the World Wide Web. http://timiti.blogspot.com/2008/03/user-experience metrics.html

21. Liu, X. X., Liu, Q. Y., Wang, W., Huang, C. M., & Li, W. S. (2013). Research on the E-commerce's user experience evaluation model. *Applied Mechanics and Materials*, 427–429, 2859–2863.

22. Mashapa, J., & Greunen, D.V. (2010). User experience evaluation metrics for usable accounting tools. *Research Conference of the South African Institute of Computer Scientists and Information Technologists*.

23. Nielsen, J. (1994). Heuristic Evaluation. In J. Nielsen, & R. Mack, Usability Inspection Methods. *New York: Wiley.*

24. Onișor, L., & Ionita, D. (2016). Crowdsourcing Platforms: Users' Experience Exposed.

25. Pretorius, M. C., Calitz, A. P., & van Greunen, D. (2005). The Added Value of Eye Tracking in the Usability Evaluation of a Network Management Tool. *Proceedings of the 2005 annual research conference of the South African Institute of Computer Scientists and Information Technologists (SAICSIT).* White River, South Africa.

26. Pruitt DG, Carnevale PJ (1993) Negotiation in Social Conflict. *Open University Press, Buckingham, England*

27. Putnam, L. L., & Roloff, M. E. (1992). Communication perspectives on negotiation. In Communication and Negotiation (Vol. 20, pp. 1-18). *SAGE Publications, Inc.*, https://doi.org/10.4135/9781483325880

28. Rubin, J., & Chisnell, D. (2008). Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests. *Indianapolis: Wiley Pub*.

29. Rubinoff, R. (2004). How to Quantify the User Experience. Available from http://www.sitepoint.com/article/quantify-user experience; accessed on 15 May 2009

30. Sauro, J., & Kindlund, E. (2005). A Method to Standardize Usability Metrics into a Single Score. *Proceedings of CHI 2005* (pp. 401-409). New York, NY: ACM.

31. Schoop, M., Jertila, A., & List, T. (2003). Negoisst: a negotiation support system for electronic business-to-business negotiations in e-commerce. *Data & Knowledge Engineering*, 47(3), 371-401.

32. Shapiro, D. L. (2002). Negotiating emotions. Conflict Resolution Quarterly, 20(1), 67-82.

33. Sharp, H., Roger, Y., & Preece, J. (2007). Interaction Design: Beyond Human Computer Interaction. *John Wiley & Sons.*

34. Sonnleitner, A., Pawlowski, M., Kässer, T., & Peissner, M. (2013). Experimentally Manipulating Positive User Experience Based on the Fulfilment of User Needs. *IFIP TC13 International Conference on Human-Computer Interaction.*

35. Ströbel, M., & Weinhardt, C. (2003). The Montreal taxonomy for electronic negotiations. *Group Decision and Negotiation,* 12(2), 143-164.

36. Tcha-Tokey, K., Christmann, O., Loup-Escande, E., Loup, G., & Richir, S. (2018). Towards a model of user experience in immersive virtual environments. *Advances in Human-Computer Interaction*, 2018, 1–10.

37. Tullis, T., & Albert, B. (2008). Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics. *Morgan Kaufmann*.

38. Vahidov, R., & Kersten, G. E. (2004). Decision station: Situating decision support systems. *Decision Support Systems*, 38(2), 283-303.

39. William Albert and Thomas Tullis. 2013. Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics. *Morgan Kaufmann*. https: //doi.org/10.1016/C2011- 0- 00016- 9

24th International Conference on Group Decision and Negotiation &
10th International Conference on Decision Support System Technology
Human-Centric Decision and Negotiation Support for Societal Transitions

Potential of Generative Artificial Intelligence in Digital Negotiations

Lennart Rukower[1] and Mareike Schoop[1*]
[1] Dept. of Intelligent Information Systems, University of Hohenheim, Stuttgart, Germany
[*] schoop@uni-hohenheim.de

Abstract

Generative Artificial Intelligence (GAI) has seen a huge rise in popularity both for academic research and for practical use. Chatbots such as ChatGPT are being used to generate texts, to enhance writing, to summarise literature, or to program certain tasks. There is much potential and also some limitations which calls for a critical assessment of GAI's potential. When looking at GAI, one cannot help but wonder whether chatbots of the newest generation will play a significant role in digital negotiations. We found that whilst GAI has been analysed for various topics such as medicine and engineering, there is no literature assessing the potential of GAI in digital negotiations. This is the topic of the current paper. To this end, we performed a structured literature review looking beyond digital negotiation. We will discuss the results and implications. Furthermore, we will enhance the literature-based results with experiments putting GAI in (digital negotiation) action.

Keywords: generative AI; Artificial Intelligence; Digital Negotiations; Negotiation Support Systems; Negoisst

1. Introduction

In recent years, much progress has been made in the field of GAI. With the emergence of large language models such as OpenAI's generative pre-trained transformers, GAI is able to generate human-sounding text across multiple domains (Bandi et al., 2023). ChatGPT has become very popular in research and for private use (OpenAI, 2022, 2024a). The potential of GAI is already being explored or applied in many different fields such as medicine, law, logistics or education (Galli & Sartor, 2023; Garcia Valencia et al., 2023; Hancock et al., 2020; Kmiecik, 2023; Li et al., 2024).

OpenAI claims that ChatGPT can teach someone to negotiate, and even offers a customised model: "The Negotiator", that can help you "advocate for yourself and get better results" in negotiation scenarios such as salary negotiations or negotiating a car purchase (OpenAI, 2024a).

Apart from this marketing statements, it is important to find out how GAI will change the landscape of digital negotiations and of negotiation support systems. Will chatbots be able to conduct complex digital negotiations? Will they outperform human negotiation experts? Can GAI be used to support human negotiators in negotiation support systems? To find out more, the research question is:

What is the potential of generative artificial intelligence for digital negotiations?

To answer this question, we have conducted a systematic literature review (SLR) which shows that related literature is sparse as this is a new research field. Therefore, we will also conduct an experimental study using ChatGPT in different types of digital negotiation to assess its negotiation capabilities.

The paper is structured as follows. Chapter 2 will briefly review fundamentals of digital negotiations and negotiation support whereas chapter 3 will briefly introduce GAI. Chapter 4 shows the methodology of the performed SLR and chapter 5 presents its results. A discussion (chapter 5) and a conclusion and outlook (chapter 6) conclude the paper.

2. Digital Negotiations & Negotiation Support

This section defines digital negotiations and negotiation support systems, using Negoisst as an example of the latter.

2.1. Digital Negotiations

"In a negotiation […] there are two or more participants in a situation of some kind of interdependence, […] each having some individual goals which may be partially incompatible. In some form of […] the negotiation process, […] alternatives are investigated, […] of which one is mutually agreed upon as the acceptable outcome of the process" (Weigand et al., 2003, p. 6). A digital negotiation, often also called electronic negotiation, is a negotiation performed over a digital medium which provides the means for distributed and asynchronous negotiations. It involves digital support for at least one of the following areas: communication, decision-making, and document management (Schoop, 2021a).

2.2. Negotiation Support Systems

A negotiation support system (NSS) aims to support human negotiators while leaving the decision-making power with them (Schoop, 2021a). Its goal is to assist negotiating parties to efficiently reach an agreement. To achieve this, four support areas exist: communication support, decision support, document management, and conflict management (Dannenmann & Schoop, 2011; Schoop, 2010; Schoop et al., 2004).

2.3. Negoisst

Negoisst is a negotiation support system that offers a holistic approach to all four support areas (Dannenmann & Schoop, 2011; Schoop, 2021a). Therefore, it is currently the most sophisticated support system for negotiations. Negoisst supports communication by providing semantic and pragmatic enrichment, which helps to convey meaning and intention in a solely asynchronous textual exchange of messages. Decision Support in Negoisst focuses on preference elicitation and on computing the utility to rate exchanged packages during a digital negotiation visualising the utilities in various ways. Document management provides a contract template library for various types of negotiations. Conflict management makes negotiators aware of existing conflicts and helps to solve them either directly of involving a third party (Schoop, 2021a).

3. Generative Artificial Intelligence

GAI models are capable of generating realistic and novel data such as texts, images, videos, or music (Bandi et al., 2023) that is "indistinguishable from human-created output" (Ooi et al., 2023, p. 1). These models have various applications, including image synthesis, text generation and acting as human-like chatbots (Bandi et al., 2023). GAI models can be categorised as unimodal or multimodal models. Unimodal GAI models produce data of the same type as the input data. For example, a model that generates only text based on only textual instructions is called a text-to-text model. Multimodal GAI models can generate data types that are different from the input data types. For example, a multimodal model can generate an image based on a textual prompt (Cao et al., 2023).

3.1. ChatGPT

ChatGPT is a text-to-text large language model (LLM) developed by OpenAI which uses the Generative Pre-trained Transformers (GPT) model. It is trained on large amounts of data and uses Reinforcement Learning from Human Feedback to finetune the GPT model which is an approach to align the generated output with preferences from humans in an effective way (Bandi et al., 2023; Cao et al., 2023; OpenAI, 2022). ChatGPT offers a chatbot interface to chat with the underlying LLM. The latest version of ChatGPT, i.e. ChatGPT 4 can also generate images by using OpenAI's text-to-image GAI model DALL-E 3 (OpenAI, 2023). To assess ChatGPT 4's power, various tests have been performed. For example, ChatGPT 4 passed the uniform bar exam according to Katz et al. (2023) showing how advanced the underlying model already is.

4. Methodology

To evaluate the potential of GAI in digital negotiations, we performed two SLRs. The first SLR focused on GAI and digital negotiations. The results suggest that there is currently no research on the potential of GAI in the context of digital negotiations. Therefore, we conducted a second SLR focusing on GAI in decision support and communication support as the two main support areas in digital negotiations. In this SLR, we identified the potential of GAI in decision support and communication support in various domains beyond digital negotiations. To put ChatGPT 4 in action, we will perform three experiments with ChatGPT 4 in different negotiation roles. In the following section we explain how we performed the SLRs and will show the experimental plan.

4.1. Systematic Literature Review

The SLRs were conducted based on the methodology of vom Brocke et al., 2009 and Xiao & Watson, 2019. The goal of the initial SLR was to collect relevant literature on GAI and digital negotiations that could be used to extract the potential and to assess the current state-of-the-art. A first sample search, using the string *"generative artificial intelligence" AND ("digital negotiation" OR "electronic negotiation")* on Google Scholar did not yield any results. Due to the lack of results, a more comprehensive search string was then constructed and additional databases were used, namely Google Scholar, SpringerLink, ACM Digital Library, IEEE Xplore, ScienceDirect and AIS eLibrary. The final search string for the database keyword search was:

((“electronic negotiation” OR “digital negotiation” OR “electronic negotiations” OR “digital negotiations” OR “negotiation support” OR “negotiation support system”) AND (“generative artificial intelligence” OR ChatGPT))

The database keyword search resulted in 10 results. These 10 hits were then further analysed by systematically screening the abstracts and filtering the full texts to the topics GAI and negotiations. However, this screening did not yield any relevant results.

A second SLR was thus conducted with the goal of providing a pool of literature on GAI in the context of communication support and decision support as the two primary areas of negotiation support. A sample search using the search string *“generative artificial intelligence” AND (“communication support” OR “decision support”)* resulted in a substantial amount of literature. The same six databases were used for the second SLR, and the final search string was constructed:

negotiation AND (“communication support” OR “decision support”) AND (“generative artificial intelligence” OR ChatGPT)

The database keyword search yielded 617 results. 26 duplicates were immediately removed. To narrow down the findings further, we systematically screened the abstracts and filtered the full texts for GAI and communication or decision support. This resulted in 25 papers which were then further analysed by examining the full texts in detail. Papers that did not provide any information on the potential of GAI for communication or decision support were filtered out. After this detailed examination, a total of 7 papers remained, which were used as basis for defining the potential of GAI in digital negotiations. During the screening and examination phase, the focus was on text-to-text GAI as Negoisst and other negotiation support systems support the exchange of textual messages (Schoop, 2021a). The process of the performed SLRs is shown in Figure 1:

Process of the conducted SLR.

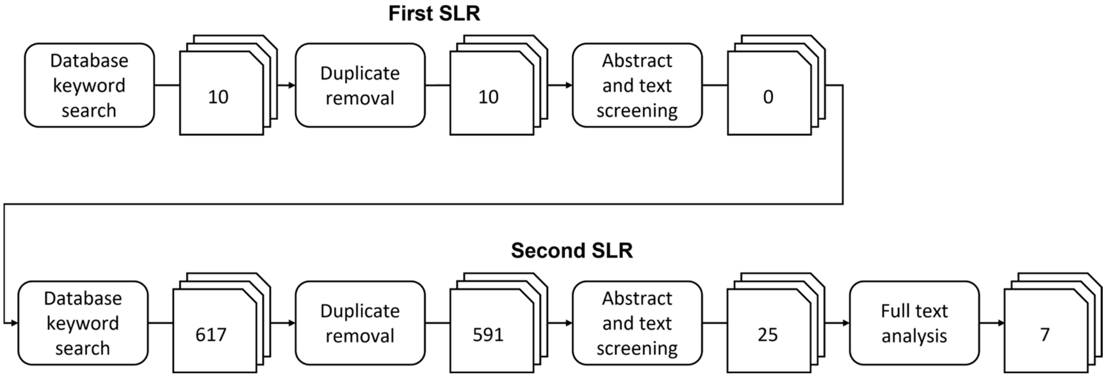


Figure 1: Process of the conducted SLRs

4.2. Experiment

Since our findings show that literature on GAI in dici there is no literature on GAI in digital negotiations is sparse, it is important to widen our explorative analysis through a practical experimental study. In particular, we will perform three experimental evaluations to assess GAI's abilities to support human negotiators and/or to replace them. The GPT model is the GAI model that is most often used in the literature we found. For the experiments we will, therefore, use ChatGPT 4, as it is the latest GPT version. The experiments will answer the following questions:

1. How can ChatGPT support a negotiator during a digital negotiation?
2. Can ChatGPT conduct negotiations with human negotiators?
3. Can two instances of ChatGPT negotiate with each other?

The first experimental setting will entail an explorative evaluation of recommendations provided by ChatGPT 4 in human-to-human digital negotiations. ChatGPT 4 will be presented with a series of negotiation scenarios to provide communication and decision support in the form of recommendations.

The second explorative experimental setting will use ChatGPT 4 as a negotiator interacting with a human negotiation partner. The negotiation will be conducted based on a complex negotiation case study with 30 Bachelor students studying Information Systems, Business and Economics, Digital Business Management, and Education for Business and Economics. The students will negotiate with a prompted ChatGPT 4 instance. The initial prompt used to prepare the ChatGPT 4 instance for the negotiation is constructed, tested, and iteratively improved prior to the experimental study.

The third experimental setting will use two instances of ChatGPT 4 to negotiate among themselves. In this setting, an existing case study will be selected that was used in a previous negotiation experiments using via Negoisst. This case study is integrated into the initial prompt for the two ChatGPT 4 instances. Subsequently, the two instances will engage in negotiations with each other multiple times. The average of the results of those negotiations will be compared to the results of the human negotiations.

5. Results

This section presents the results of the second systematic literature review (SLR) and discusses their implications.

5.1. State-of-the-Art

In the following section, we present the results of the second SLR, which focuses on decision support and communication support. Firstly, we investigate the capabilities of GAI in enhancing communication support, and then proceed to explore its potential in decision support.

The analysed literature focuses on GAI in general as well as on specific GAI applications such as ChatGPT 4. In the following we will use the term GAI for both.

The use of GAI in communication support is diverse, addressing different aspects from patient communication to language translation. In the context of the medical field, Garcia Valencia et al. (2023) highlight the effectiveness of GAI in patient communication, particularly in the sensitive context of kidney transplantation. This illustrates the potential of GAI to convey complex medical information. They also note GAI's ability to provide language translation and to generate culturally appropriate wordings to ensure respectful and effective communication. Eysenbach (2023) investigates the use of GAI in developing medical education scenarios and patient communication examples, emphasising its value as a useful tool in medical training. In the area of online dispute resolution Westermann et al. (2023) propose an online platform for dispute resolution using GAI. It can reformulate user messages to de-escalate the dispute between two parties and can suggest messages to mediators. They suggest that GAI could be used to classify exchanged messages as either "inflammatory" or "non-inflammatory", in order to alert users or mediators. In educational research, Kovačević (2023) examines how GAI can be fine-tuned to generate domain-specific text. This adaptability of GAI to specialised content demonstrates its potential in various application fields. The study by Bandi et al. (2023) about the power of GAI discusses GAI for the role of a conversational agent. This is also mentioned by Ooi et al. (2023) who look at potentials of GAI across disciplines, highlighting GAI's ability to generate structured responses that are nuanced and comparable to human communication. This demonstrates the ability of GAI to mimic human conversational patterns.

In decision support, GAI is used primarily for generating suggestions or alternative scenarios as well as for summarisation. In the area of healthcare, Liu et al. (2023) explore the potential of GAI to improve clinical decision support systems by generating relevant and comprehensible suggestions for supplementing clinical decision-making. In the judicial domain, Galli and Sartor (2023) investigate the use of GAI in decision support for the justice system, specifically in predicting and assisting with complex legal decision-making processes, including summarising legal texts. Bandi et al. (2023) show how well GAI can summarise input, emphasizing its adaptability in processing various types of content, including code and text. Additionally, Ooi et al. (2023) demonstrate how GAI supports decision-making by generating alternative scenarios and evaluating probable outcomes. Furthermore, the context of human-AI collaboration is a topic in Ferguson et al. (2023) who discuss the effectiveness of GAI in ambiguous situations by revealing new perspectives, rationales, and insights, highlighting GAI's ability to deal with complex and uncertain decision-making environments. Finally, Saka et al. (2024) suggest that GAI models can provide valuable recommendations based on a thorough data analysis to enable data-driven procurement or project management decisions.

5.2. Implications

Based on the literature, the following potential of GAI for digital negotiations can be identified.

(Re-)Formulation and improvement of negotiation messages: As GAI can generate domain-specific texts as well as human-like responses and suggestions, it could be used to formulate and improve messages in digital negotiations.

Language translation for multi-lingual communication: GAI's ability to provide language translation could be used in international digital negotiations to bridge language barriers between different negotiators.

Summarising ongoing negotiations: Due to GAI's summarisation capability, it could be used to summarise negotiations in real-time which helps negotiators to maintain an overview of the exchanged offers and arguments in complex negotiations.

Post-negotiation summary: GAI's summarisation capability could also be used for summarising negotiations after they ended. This can provide insights into negotiation dynamics and negotiation processes which in turn provide lessons learned for negotiators.

Autonomous Negotiator: Due to GAI's abilities to generate human-like domain-specific responses and arguments, to act as a conversational agent as well as its ability to deal with complex and uncertain decision-making environments, GAI could potentially be used as an autonomous digital negotiator.

Not all of GAI's potential extracted though the SLR is applicable to the digital negotiation context. This is due to the dynamic nature of negotiations and the combination of textual communication data in form of messages and numeric data in form of utility and preference values (Kaya & Schoop, 2023).

Considering these implications, it is conceivable that GAI could support digital negotiators and even replace human negotiators in complex digital negotiation scenarios. This could have a significant impact on the landscape of digital negotiations and of negotiation support.

6. Discussion and Conclusion

Generative AI has already and will continue to change the academic landscape and organisational practice. It is not yet clear where this will lead to. Due to the rapid development of technologies such as the GPT model from OpenAI, the technical capabilities explored in the reviewed literature may already be outdated. For instance, while writing this paper, Google introduced Gemini, a new large language model that comprehends multimodal information across millions of contextual tokens (Gemini Team, 2024; Google, 2024).

While GAI provides potential applications in the context of digital negotiations, it is important to acknowledge its limitations and potential risks.

Hallucination: Because GAI models are trained on large amounts of data, they are prone to hallucinations, which means they confidently fabricate information (Bandi et al., 2023; Cao et al., 2023; Li et al., 2024; Liu et al., 2023; Mosaiyebzadeh et al., 2023; Saka et al., 2024). This is especially dangerous in digital negotiations where a certain set of issues which the negotiators agreed upon are negotiated (Schoop, 2021b). For instance, adding new issues or altering the negotiation setting would render an autonomous negotiation agent using GAI useless.

Trust: Due to the risk of hallucination, the trust in GAI models to support decision making and communication can be limited (Mosaiyebzadeh et al., 2023; Saka et al., 2024). Generative models are also intransparent on their own (Bandi et al., 2023). This lack of transparency further diminishes trust in the system.

Skills in prompt engineering: To use GAI models effectively, the user or the system needs to interact with the model in an effective way. This requires prompt engineering, which is a skill that needs to be developed (OpenAI, 2024b; Saka et al., 2024).

During our SLR, a limited number of databases was selected. To address this limitation, Google Scholar was utilised, as it provides a large corpus of literature from various sources. However, it is important to note that Google Scholar's search algorithm is intransparent and difficult to understand. There is a considerable amount of literature on the potential of GAI beyond communication and decision support, which could also be relevant to digital negotiations. For example, literature on the potential in the areas of document management and mediation will be considered in a next step.

The quality of the initial prompt plays an important role in determining the outcome of the experiments conducted with ChatGPT 4. Therefore, it is crucial to assess the effectiveness of different prompts and techniques in future studies.

References

Bandi, A., Adapa, P. V. S. R., & Kuchi, Y. E. V. P. K. (2023). The Power of Generative AI: A Review of Requirements, Models, Input–Output Formats, Evaluation Metrics, and Challenges. Future Internet, 15(8), 260. https://doi.org/10.3390/fi15080260

Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P. S., & Sun, L. (2023, July 3). A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT. http://arxiv.org/pdf/2303.04226v1

Dannenmann, A., & Schoop, M. (2011). Conflict Resolution Support in Electronic Negotiations, Wirtschaftsinformatik Proceedings 2011, Article 31.

Eysenbach, G. (2023). The Role of ChatGPT, Generative Language Models, and Artificial Intelligence in Medical Education: A Conversation With ChatGPT and a Call for Papers. JMIR Medical Education, 9. https://doi.org/10.2196/46885

Ferguson, S. A., Aoyagui, P. A., & Kuzminykh, A. (2023). Something Borrowed: Exploring the Influence of AI-Generated Explanation Text on the Composition of Human Explanations. In A. Schmidt, K. Väänänen, T. Goyal, P. O. Kristensson, & A. Peters (Eds.), Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems. ACM. https://doi.org/10.1145/3544549.3585727

Galli, F., & Sartor, G. (2023). AI Approaches to Predictive Justice: A Critical Assessment. Humanities and Rights Global Network Journal, 5(2). https://www.humanitiesandrights.com/journal/index.php/har/article/view/118

Garcia Valencia, O. A., Thongprayoon, C., Jadlowiec, C. C., Mao, S. A., Miao, J., & Cheungpasitporn, W. (2023). Enhancing Kidney Transplant Care through the Integration of Chatbot. Healthcare (Basel, Switzerland), 11(18). https://doi.org/10.3390/healthcare11182518

Gemini Team, G. (2024). Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. Retrieved March 12, 2024, from https://storage.googleapis.com/deepmind-media/gemini/gemini_v1_5_report.pdf

Google. (2024). Gemini. Retrieved March 12, 2024, from https://deepmind.google/technologies/gemini/#introduction

Hancock, J. T., Naaman, M., & Levy, K. (2020). AI-Mediated Communication: Definition, Research Agenda, and Ethical Considerations. Journal of Computer-Mediated Communication, 25(1), 89–100. https://doi.org/10.1093/jcmc/zmz022

Katz, D. M., Bommarito, M. J., Gao, S., & Arredondo, P. (2023). GPT-4 Passes the Bar Exam. SSRN Electronic Journal. Advance online publication. https://doi.org/10.2139/ssrn.4389233

Kaya, M.-F., & Schoop, M. (2023). The More the Merrier? A Machine Learning Analysis of Information Growth in Negotiation Processes. In Y. Maemura, M. Horita, L. Fang, & P. Zaraté (Eds.), Lecture Notes in Business Information Processing. Group Decision and Negotiation in the Era of Multimodal Interactions (Vol. 478, pp. 20–34). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-33780-2_2

Kmiecik, M. (2023). ChatGPT in third-party logistics – The game-changer or a step into the unknown? Journal of Open Innovation: Technology, Market, and Complexity, 9(4), 100174. https://doi.org/10.1016/j.joitmc.2023.100174

Kovačević, D. (2023). Use of ChatGPT in ESP Teaching Process. In 22nd International Symposium INFOTEH-JAHORINA 3152023 (pp. 1–5). https://doi.org/10.1109/INFOTEH57020.2023.10094133

Li, J., Dada, A., Puladi, B., Kleesiek, J., & Egger, J. (2024). Chatgpt in healthcare: A taxonomy and systematic review. Computer Methods and Programs in Biomedicine, 245. https://doi.org/10.1016/j.cmpb.2024.108013

Liu, S., Wright, A. P., Patterson, B. L., Wanderer, J. P., Turer, R. W., Nelson, S. D., McCoy, A. B., Sittig, D. F., & Wright, A. (2023). Using AI-generated suggestions from ChatGPT to optimize clinical decision support. Journal of the American Medical Informatics Association, 30(7), 1237–1245. https://doi.org/10.1093/jamia/ocad072

Mosaiyebzadeh, F., Pouriyeh, S., Parizi, R., Dehbozorgi, N., Dorodchi, M., & Macêdo Batista, D. (2023). Exploring the Role of ChatGPT in Education: Applications and Challenges. In Y. Xie, B. Rutherfoord, & H. Park (Eds.), The 24th Annual Conference on Information Technology Education (pp. 84–89). ACM. https://doi.org/10.1145/3585059.3611445

Ooi, K.-B., Tan, G. W.-H., Al-Emran, M., Al-Sharafi, M. A., Capatina, A., Chakraborty, A., Dwivedi, Y. K., Huang, T.-L., Kar, A. K., Lee, V.-H., Loh, X.-M., Micu, A., Mikalef, P., Mogaji, E., Pandey, N., Raman, R., Rana, N. P., Sarker, P., Sharma, A., . . . Wong, L.-W. (2023). The Potential of Generative Artificial Intelligence Across Disciplines: Perspectives and Future Directions. Journal of Computer Information Systems, 1–32. https://doi.org/10.1080/08874417.2023.2261010

OpenAI. (2022). Introducing ChatGPT. Retrieved March 12, 2024, from https://openai.com/blog/chatgpt

OpenAI. (2023). DALL·E 3. Retrieved March 12, 2024, from https://openai.com/dall-e-3

OpenAI. (2024a). ChatGPT. Retrieved March 12, 2024, from https://openai.com/chatgpt

OpenAI. (2024b). Prompt engineering. Retrieved March 12, 2024, from https://platform.openai.com/docs/guides/prompt-engineering

Saka, A., Taiwo, R., Saka, N., Salami, B. A., Ajayi, S., Akande, K., & Kazemi, H. (2024). GPT models in construction industry: Opportunities, limitations, and a use case validation. Developments in the Built Environment, 17. https://doi.org/10.1016/j.dibe.2023.100300

Schoop, M. (2010). Support of Complex Electronic Negotiations. Kilgour, D., Eden, C. (Eds) Handbook of Group Decision and Negotiation, 409–423. https://doi.org/10.1007/978-90-481-9097-3_24

Schoop, M. (2021a). Negoisst: Complex Digital Negotiation Support. In D. M. Kilgour & C. Eden (Eds.), Handbook of Group Decision and Negotiation (2nd ed., pp. 1149–1167). Springer Cham. https://doi.org/10.1007/978-3-030-49629-6_24

Schoop, M. (2021b). Negotiation communication revisited. Central European Journal of Operations Research. Advance online publication. https://doi.org/10.1007/s10100-020-00730-5

Schoop, M., Köhne, F., & Staskiewicz, D. (2004). Decision Support in Electronic Negotiation Systems - New challenges. Journal of Decision Systems, 13(4), 375–398. https://doi.org/10.3166/jds.13.375-398

vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., & Cleven, A. (2009). Reconstructing the Giant - On the Importance of Rigour in Documenting the Literature Search Process.doc. ECIS 2009 Proceedings, Article 161.

Weigand, H., Schoop, M., Moor, A. de, & Dignum, F. (2003). B2B Negotiation Support: The Need for a Communication Perspective. Group Decision and Negotiation.

Westermann, H., Savelka, J., & Benyekhlef, K. (2023). LLMediator: GPT-4 Assisted Online Dispute Resolution. Workshop on Artificial Intelligence for Access to Justice. http://arxiv.org/pdf/2307.16732v1

Xiao, Y., & Watson, M. (2019). Guidance on Conducting a Systematic Literature Review. Journal of Planning Education and Research, 39(1), 93–112. https://doi.org/10.1177/0739456X17723971

24th International Conference on Group Decision and Negotiation &
10th International Conference on Decision Support System Technology
Human-Centric Decision and Negotiation Support for Societal Transitions

Predictive Detection of Negotiation Progress with Machine Learning

Muhammed-Fatih Kaya[1*] and Mareike Schoop[1]
[1] University of Hohenheim, Stuttgart, Germany
[*] Muhammed-Fatih.Kaya@uni-hohenheim.de

Abstract

Negotiators perform different strategic and tactical actions to achieve their desired negotiation goals. During this dynamic process, it is always an open question for the negotiators what negotiation progress has been made as a result of previous actional and reactional efforts. Additional procedural information on the current progress could help negotiators to assess the process better. The use of Machine Learning (ML) has been shown to provide an essential benefit in the generation of such procedural predictions by considering negotiation data, for example from Negotiation Support Systems. To this end, various renowned and modern ML methods have to be evaluated regarding their applicability as well as their prediction performance in a customised framework. This paper presents a methodological progress detection framework based on ML which can be used to train and evaluate ML models for the prediction of negotiation progresses. The results show that numerous data-model-split combinations must be analysed in detail to maximise the prediction accuracy of this ML problem and to ensure the real-word applicability of such kind of predictive models.

Keywords: Machine Learning, Negotiation Progress, State Detection, Classification, Negotiation Support Systems

1. Motivation

Negotiations are complex iterative communication and decision-making processes (Bichler et al., 2003). They are used in a wide variety of application areas and can contribute to value-adding solutions in interdependent conflicts of interest (Agndal, 2007). From a procedural perspective, negotiations can be very dynamic and take different courses (Olekalns & Weingart, 2008). Whilst distributive or integrative bilateral negotiations with a high conflict level can take long before an agreement is reached, negotiation progress can be achieved much faster in other negotiations with the same negotiation scenario by taking the right choice of strategies (Carnevale & Pruitt, 1992).

Procedural information about the state of the negotiation process is an important information for negotiators (Lewicki et al., 2020). It is essential to know what progress has been made in the negotiation process and what kind of effort is still required to complete the negotiation (Koeszegi & Vetschera, 2010). For example, is the negotiation in the early stages or is it nearing its end? In this paper, sections are used as negotiation progress which are to be predicted by processing negotiation data. Sections represent information regarding the question to what extent a communication or utility-based interaction takes place in a certain part of the negotiation (e.g. in section 3 of 4 for 75% of negotiation progress). The usage of negotiation data to predict the information in which negotiation section a person is currently in (as a dependent variable) could enable negotiators to use and adapt their individual strategies and tactics more effectively. Alternatively, such section-based progress information could also be used for simple process monitoring. In any case, this type of predictive forecasting requires the entire negotiation process to be analysed in a systematic way across various cycles (Kaya & Schoop, 2023). This requires a large amount of negotiation data which can be generated by negotiators using Negotiation Support Systems (NSSs).

Manual analysis cannot deal with such large amount of data from NSSs. In contrast, Machine Learning (ML) methods can be used to process large volumes of data for precisely these kind of predictive analysis purposes. They can process both structured and unstructured electronic negotiation data and derive value-adding (descriptive and predictive) knowledge for the negotiation process (Kaya, 2022; Vetschera et al., 2021). We will present a methodological framework considering a variety of ML approaches to evaluate their application potential for the prediction of negotiation progress. Thus, the research question is:

Can the progress of e-negotiations be predicted using predictive machine learning methods?

After an in-depth motivation of the research topic, the following two chapters describe the different courses of negotiation processes (chapter 2) and the characteristics of the NSS Negoisst (chapter 3). Chapter 4 analyses the prediction potential of ML-techniques and categorises different methods according to their functionality. Chapter 5 then presents the framework for progress detection which is used to analyse the application potential of selected ML techniques for predicting the progress of negotiations. Finally, chapter 6 summarises the key implications and limitations of this work and shows how the implementation and evaluation process will be carried out in further steps.

2. Negotiation processes and the effect of negotiation strategies

For many years, negotiation researchers, particularly in the context of electronic negotiations, have been investigating certain behavioural patterns in negotiations to control the strategic orientation of the negotiation process and ensure successful negotiation management (Adair & Brett, 2005; Weingart et al., 2004; Koeszegi et al., 2006). Bilateral negotiations are described as a sequence of actions and reactions in which these kind of repetitive negotiation patterns develop as individual negotiation sequences (Putnam & Jones, 1982; Putnam, 1983).

Reciprocal negotiation sequences describe interactions where the recipient reacts with the same behaviour than the sender of an offer (Weingart et al., 1990; Olekalns & Smith, 2000). Reciprocal sequences therefore consist of a series of similarly alternating strategic orientations (Fells, 2010; Weingart et al., 1990). In contrast, non-reciprocal sequences can occur as complementary or structural sequences (Olekalns & Smith, 2000). Complementary sequences indicate a response behaviour that is characterised by the same strategic orientation, but is fulfilled by different strategies of the same orientation. Olekalns and Smith (2000) describe the example that threats by one negotiation partner are followed by demands of the other partner. Structural sequences occur when the behaviour of the negotiation partners is based on different strategic orientations (such as integrative vs. distributive strategies). For example, a negotiator's integrative-cooperative behaviour is followed by the partner's threatening demand as a competitive strategy (Adair & Brett, 2005; Brett et al., 2003). As can be seen from the examples, there are various strategies in different sequences. As will be described in the following section, these strategies can have different effects on the progress of the negotiation process depending on their individual usage (Koeszegi & Vetschera, 2010).

Negotiation literature has shown an initial offer to have an anchoring effect on the negotiation process, irrespective of the various strategic orientations. Hence, all offers and counteroffers are orientated towards the initial offer, particularly in the first phase of the negotiation (Liebert et al., 1968; Ritov, 1996). An unrealistic initial offer combined with minimal concessions can prolong the negotiation process before an agreement is reached. On the other hand, a more realistic initial offer with more substantial concessions can lead to a significant acceleration of the negotiation process (Carnevale & Pruitt, 1992).

It has been shown that distributive strategies can generally have a negative influence on satisfaction with the negotiation process (Schoop et al. 2014). The strict reciprocal use of distributive strategies does not only reduce the probability of reaching an agreement (Pruitt & Levis, 1975, Koeszegi et al., 2006), but can also lead to an escalation of the conflict and thus to the failure of the negotiation in extreme cases (Olekalns & Weingart, 2008; Weingart et al., 1990). In the best-case scenario, the use of hard tactics such as insisting on one's own negotiating position leads to significantly longer negotiation processes (Filzmoser & Vetschera, 2008). Surprisingly, on the other hand, previous studies have shown that the use of time pressure as a distributive instrument can shorten the duration of negotiations (de Dreu, 2003; Stuhlmacher & Champagne, 2000).

Integrative strategies can also have different effects on negotiation processes. For example, the reciprocal use of concessions as a problem-solving and compromise-orientated approach can lead to a positive relationship and trust dynamic between the negotiators (Lewicki & Polin, 2013). However, the use of concession strategies tends to prolong the negotiation process, even if the high reciprocal commitment can significantly increase the agreement rate (Filzmoser & Vetschera, 2008). The logrolling strategy exhibits similar behaviour. It enables pareto-optimal agreements to be reached (Lopes & Coelho, 2010; Qu & Cheung, 2012), although it also tends to lead to extended negotiation processes. In addition to a realistic anchor that is not too high, the targeted and transparent disclosure of information on preferences, interests, needs, etc. (also referred to as integrative information exchange) can have a positive effect on the negotiation process and shorten the negotiation process required to reach an agreement (Thompson et al., 2010).

3. The NSS Negoisst

The support of electronic negotiations in NSSs is directly affected by process-orientated influences of negotiation sequences and strategies on the negotiation process. NSSs enable the electronic conduct of negotiations and provide additional benefit through information and communication technology (Schoop, 2021). The main requirements for an NSS thus include decision support, communication support and support for document management (Elsler & Schoop, 2012). The Negoisst system is one renowned NSS-system. It enables bilateral e-negotiations to be conducted via an asynchronous exchange of messages and offers by reflecting the individual strategies of the negotiators (Schoop et al., 2003). While communication support enables the structured and comprehensible exchange of negotiation messages between sender and receiver in the form of natural language (Schoop et al., 2010), decision support enables the efficient exchange of offers by representing the issues and attributes through multi-attributive utility functions (Schoop et al., 2014). Metric utility values (i.e. individual utility of both negotiators, joint utility and contract imbalance) and the corresponding negotiation message (as unstructured text) are stored for each offer exchange during the negotiation process (Kaya & Schoop, 2019; Kaya & Schoop, 2020). This exchange continues until a negotiation is either accepted or rejected. In summary, both data dimensions (utility as well as communication) represent a valuable database for predicting the progress of negotiations using machine learning.

4. Machine Learning for prediction problems

ML approaches have so far been used to tackle various tasks in the application field of electronic negotiations. To provide first insights into existing studies, ML has been used e.g., to predict the behaviour of negotiating partners (Carbonneau et al., 2008), to predict the probability of negotiation success in sense of contract acceptance and rejection (Lewis et al., 2017), to determine first behavioural patterns in negotiations (Mashayekhy et al, 2006), to recognise syntactic communication patterns (Kaya & Schoop, 2022; Sokolova & Szpakowicz 2007) and to analyse the effects of sentiment usage in negotiation language (Filzmoser et al. 2016; Körner & Schoop, 2014; Laubert & Parlamis, 2019). Previous studies have not explicitly focused on the prediction of negotiation progress by considering the communication and utility data of an NSS. Consequently, the first fundamental milestones for prediction of negotiation progress are provided with this paper.

The use of ML is particularly suitable for processing predictive analysis tasks. ML refers to methods that can be used to discover patterns in existing data sets (in our case negotiation data from the NSS Negoisst) and to generate predictions (Murphy, 2012). To achieve this goal, the negotiation data used must first be divided into a training data set and a test data set. In the training phase, the selected ML-method is trained using the training data to generate a generalisable model (Nasteski, 2017). The resulting model can be described as a simplified representation of a real-world problem (Dahmen et al., 2019). This model has learned the pattern structures hidden in the training data and can now be applied to unknown data (test data) in the test phase in the subsequent step (Singh et al., 2016).

Classification represents one of the best-known predictive approaches in ML, which is one of the supervised learning methods. In the supervised learning approach, all training data must be labelled with a corresponding class label. The model to be created has to be trained with the aim of predicting these specific labels (Castelli et al., 2019; Han et al., 2023). Based on the training data, a prediction function with a series of input and output pairs (x,y) is created (Kotsiantis et al., 2007). x in our case represents the negotiation data and y represents the negotiation progress to be predicted.

ML literature reveals that there are different classification approaches (Kowsari et al., 2019; Soofi & Awan, 2017). In particular, the literature distinguishes between renowned approaches and the more modern deep learning-based approaches. Renowned approaches are divided into non-parametric techniques (e.g., Support Vector Machine (SVM) (Cervantes et al., 2020) and k-nearest neighbour (kNN) (Kramer, 2013)), tree-based techniques (e.g., decision trees (DT) (Singh & Gupta, 2014), random forest (RF) (Parmar et al., 2019) and gradient boosting decision trees (GBT) (Bentéjac et al., 2021)), regression-based techniques (such as logistic regression (LogReg) (Hosmer et al., 2013)) and the Bayesian approach (Naïve Bayes (NB) (Jadhav & Channe, 2016)) (Kowsari et al., 2019). Deep learning methods are based on a wide variety of constellations of deep neural networks and are among the more modern approaches (Ivanovic & Radovanovic, 2015). As process data from Negoisst is to be analysed as part of our work, two methods that have proven to be working well on processes will be our focus, namely long short-term memory (LSTM (Sherstinsky, 2020)) and the transformer architecture of Bidirectional Encoder Representations (BERT) (Ganesh et al., 2021). Both have recently achieved promising results in different application areas.

Furthermore, classifiers are categorised into white-box and black-box algorithms. Black-box algorithms are characterised by the fact that their derived patterns are not comprehensible in most cases due to their complexity (Castelvecchi, 2016). Only the integrated data set and the prediction result are visible in black box approaches. The patterns by which the algorithm came to a particular conclusion remain a secret (Carabantes, 2020). Most ML methods are based on black-box algorithms. White-box algorithms, on the other hand, allow an insight into the processing structure of the algorithms. This means that not only the integrated data and the derived results can be seen. Decision-based patterns that lead to the prediction result can be interpreted (Loyola-Gonzales, 2019; Thampi, 2022). This allows classification decisions to be tracked in a transparent manner. Tree-based techniques are an example of white-box algorithms.

5. Progress Detection Framework

Considering the overarching research question, a methodological framework is presented based on ML. ML models can be trained and evaluated based on this framework for the prediction of negotiation progress.

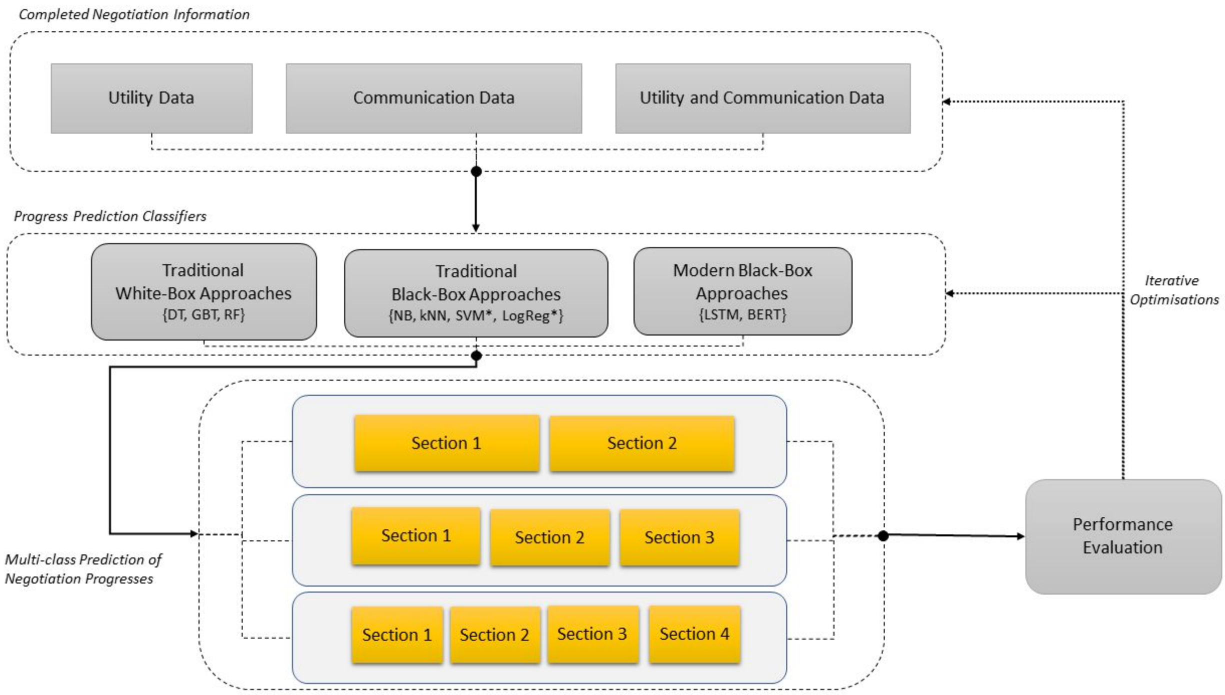


Figure 1: Progress Detection Framework

Negotiation data from the NSS Negoisst is divided into structured utility data and unstructured communication data. While the utility data can be fed directly into machine learning procedures, communication data requires special treatment. It must be processed in a complex manner in order to derive structured numerical matrices from the originally unstructured, textual negotiation messages (Kaya & Schoop, 2020). Natural language processing opens up various possibilities and enables the processing of such kind of data by ML methods (Christian et al., 2016; Li & Yang, 2018). Due to the different characteristics of these two data types, communication data and utility data are considered separately in a first step. Consequently, solely communication messages and subsequently solely utiltiy data are considered when predicting the progress of negotiations. However, since these dimensions are interwoven and the negotiators formulate the communication messages corresponding to the offers, these two data types are then combined in a joint data set (communication data and utility data) (see Figure 1).

Hence, a total of three different data sets is available which depict the entire negotiation process from a total of 668 negotiations with 7026 negotiation messages from different perspectives. On average, the number of messages exchanged in these negotiations is 10.5 with a standard deviation of 4.5. At least two negotiation messages must be exchanged in a section to be able form one section. This condition is fulfilled anyway in our case as in most data scenarios more than two messages are exchanged per section in view of the average value and standard deviation. As part of the data preparation process and after carrying out a descriptive analysis, the negotiations are divided into sections of equal length. We analyse (1) two sections with (section 1 with the first half of negotiations and section 2 with the second half of negotiations), (2) three sections (section 1 with the first third of negotiations, section 2 with the second third of the negotiations and section 3 with the last third of the negotiation) and (3) four sections (section 1 with the first quarter of negotiations; section 2 with the second quarter; section 3 with the third quarter and section 4 with the last quarter of negotiations). Since a division with more than four sections leads to a significant loss of data and to a violation of the precondition defined above, four sections are set as the maximum section-split. All offer-based utility values and corresponding messages are now labelled with one of the corresponding section information they belong to (as a dependent variable) in order to be able to train the ML models.

The training process of various ML approaches can now be carried out for each of the three section divisions. For this purpose, the negotiation data is used together with the section-based label information to derive value-adding patterns for the predictive classification of section-based negotiation progress. Nine different ML techniques are thus evaluated across all groups for the evaluation of the prediction potential of the negotiation progress in the form of sections. These were presented in the previous chapter and are summarised in Figure 1 as the traditional white-box approaches, traditional black-box approaches and modern black-box approaches. Since the prediction problem is a multi-class prediction problem, especially when divided into three and four sections, the techniques marked with * must be extended with an algorithmic extension (Bisong, 2019; Tomar & Agarwal, 2015) as these techniques were originally developed for binary classification problems (Musa, 2013).

For the overall evaluation of the prediction potential, nine different ML methods are available for each of the three underlying data sets. These data-model constellations (i.e. the usage of one of the three underlying negotiation data types for one of the nine ML-methods) can now be used to predict the negotiation progress via the two, three, or four negotiation sections. This approach makes it possible to measure to what degree the section-based negotiation progress can be predicted by considering the underlying negotiation data. Across all section splits, 27 performance results would be saved for each data-method combination. Through an iterative optimisation process, further data-method constellations can be tested in the subsequent steps in addition to the optimisation of the model parameters. A completed optimisation loop would result in 243 performance results across the entire Progress Detection Framework from Figure 1 which can be subjected to an analytical comparison to determine the best model constellation. This analytical approach enables an in-depth analysis of the predictive potential of negotiation progress.

6. Conclusion and Future Research Potential

The presented framework clearly shows that a substantial data-centred and methodological evaluation is required to evaluate the application potential of ML for progress detection in the field of electronic negotiations. Without this in-depth analysis, it would not be possible to discuss under which conditions ML methods can perform progress detection in a valid way. There are no universal ML methods that provide the best solution for all forecasting scenarios (Adam et al., 2019). Therefore, an intensive analytical and iterative process is necessary which was presented in this abstract.

In the next research steps to be presented at the conference, the application potential of the developed methods will be assessed using data of the Negoisst system. The reactions of the individual ML methods will be investigated based on different NSS data types. It is to be analysed to which extent the methodological constellations can cope with the defined prediction tasks of progress detection and to what degree challenges exist. The optimisation and expansion process of these approaches will be documented transparently in further steps. For practitioners and researchers, these results provide a wide range of added value, as they show the application potential of various ML methods for procedural business data and in particular for negotiations using NSSs.

References

W. L. Adair, J. M. Brett, The Negotiation Dance: Time, Culture, and Behavioral Sequences in Negotiation. *Organization Science*, *16*(1), 33-51, 2005.

S. P. Adam, S. A. N. Alexandropoulos, P. M. Pardalos, M. N. Vrahatis, No free lunch theorem: A review. *Approximation and optimization: Algorithms, complexity and applications* (pp. 57-82), 2019.

H. Agndal, Current Trends in Business Negotiation Research. *SSE/EFI Working Paper Series in Business Administration* (pp. 1-55), 2007.

C. Bentéjac, A. Csörgő, G. Martínez-Muñoz, A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review* (pp. 1937-1967), 2021.

M. Bichler, G. Kersten, S. Strecker, Towards a Structured Design of Electronic Negotiations. *Group Decision and Negotiation*, *12*(4), 311-335, 2003.

E. Bisong, Logistic Regression. In E. Bisong (Ed.) *Building Machine Learning and Deep Learning Models on Google Cloud Platform*. Apress, Berkeley, CA, 2019.

J. Brett, L. Weingart, M. Olekalns, Baubles, Bangles, and Beads: Modeling the Evolution of Negotiating Groups over Time. *Research on managing groups and teams: Time in Group* (pp. 1–44), 2003.

M. Carabantes, Black-box artificial intelligence: an epistemological and critical analysis. *AI & Society*, *35*(2), 309-317, 2020.

R. Carbonneau, G. E. Kersten, R. Vahidov, Predicting opponent's moves in electronic negotiations using neural networks. *Expert Systems with Applications*, *34*(2), 1266-1273, 2008.

P. J. Carnevale, D. G. Pruitt, Negotiation and Mediation. *Annual Review of Psychology*, *43*(1), 531-582, 1992.

M. Castelli, L. Vanneschi, Á. R. Largo, Supervised learning: Classification. In S. Ranganathan, M. Gribskov, K. Nakai, & C. Schönbach (Eds.), *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics* (pp. 342-349). Elsevier, 2019.

D. Castelvecchi, Can we open the black box of AI?. *Nature News* (vol. 538, pp. 20-23), 2016.

J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, A. Lopez, A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing* (vol. 408, pp. 189-215), 2020.

H. Christian, M. P. Agus, D. Suhartono, Single document automatic text summarization using term frequency-inverse document frequency (TF-IDF). *ComTech: Computer, Mathematics and Engineering Applications*, *7*(4), 285-294, 2016.

T. Dahmen, P. Trampert, F. Boughorbel, J. Sprenger, M. Klusch, K. Fischer, P. Slusallek, Digital reality: a model-based approach to supervised learning from synthetic data. *AI Perspectives*, *1*(1), 1-12, 2019.

C. K. W. de Dreu, Time pressure and closing of the mind in negotiation. *Organizational Behavior and Human Decision Processes*, *91*(2), 280-295, 2003.

R. Elsler, M. Schoop, Konzept eines Präferenzmodells basierend auf der GRIP Methodologie im Kontext elektronischer Verhandlungen. *Multikonferenz Wirtschaftsinformatik 2012*: *Tagungsband der MKWI 2012*, 2012.

R. E. Fells, *Effective Negotiation: From Research to Results.* (1st ed.). New York: Cambridge University Press, 2010.

M. Filzmoser, R. Vetschera, A Classification of Bargaining Steps and their Impact on Negotiation Outcomes. *Group Decision and Negotiation* (vol. 17, pp. 421-443), 2008.

M. Filzmoser, S. T. Koeszegi, G. Pfeffer, What Computers Can Tell Us About Emotions – Classification of Affective Communication in Electronic Negotiations by Supervised Machine Learning. In D. Bajwa, S. Koeszegi, R. Vetschera (Eds.) *Group Decision and Negotiation. Theory, Empirical Evidence, and Application*. *GDN 2016. Lecture Notes in Business Information Processing*, (vol. 274, pp. 113-123). Springer, Cham., 2017.

P. Ganesh, Y. Chen, X. Lou, M. A. Khan, Y. Yang, H. Sajjad, M. Winslett, Compressing large-scale transformer-based models: A case study on bert. *Transactions of the Association for Computational Linguistics* (vol. 9, 1061-1080), 2021.

M. Han, H. Wu, Z. Chen, M. Li, X. Zhang, A survey of multi-label classification based on supervised and semi-supervised learning. *International Journal of Machine Learning and Cybernetics*, *14*(3), 697-724, 2023.

D. W. Hosmer Jr, S. Lemeshow, R. X. Sturdivant. Applied logistic regression (vol. 398). John Wiley & Sons, 2013.

M. Ivanovic, M. Radovanovic, Modern machine learning techniques and their applications. *International Conference on Electronics, Communications and Networks*, 2015.

S. D. Jadhav, H. P. Channe, Comparative study of K-NN, naive Bayes and decision tree classification techniques. *International Journal of Science and Research (IJSR)*, *5*(1), 1842-1845, 2016.

M. F. Kaya, Pattern Labelling of Business Communication Data. *Group Decision and Negotiation*, *31*(6), 1203-1234, 2022.

M. F. Kaya, M. Schoop, Application of data mining methods for pattern recognition in negotiation support systems. In *Group Decision and Negotiation: Behavior, Models, and Support: 19th International Conference (GDN 2019)*, *Loughborough, UK, June 11–15*, 2019, *Proceedings 19* (pp. 223-237). Springer International Publishing, 2019.

M. F. Kaya, M. Schoop, Maintenance of Data Richness in Business Communication Data. In *European Conference of Information Systems*, *A Virtual Conference*, 2020.

M. F. Kaya, M. Schoop, Analytical comparison of clustering techniques for the recognition of communication patterns. *Group Decision and Negotiation*, *31*(3), 555-589, 2022.

M.F. Kaya, M. Schoop, The More the Merrier? A Machine Learning Analysis of Information Growth in Negotiation Processes. In *International Conference on Group Decision and Negotiation* (pp. 20-34). Cham: Springer Nature Switzerland, 2023.

S. T. Koeszegi, R. Vetschera. Analysis of negotiation processes. *Handbook of group decision and negotiation* (pp. 121-138), 2010.

S. T. Koeszegi, K. J. Srnka, E. M. Pesendorfer, Electronic Negotiations – A Comparison of Different Support Systems. *Die Betriebswirtschaft*, *66*(4), 441-463, 2006.

S. B. Kotsiantis, I. Zaharakis, P. Pintelas, Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, *160*(1), 3-24, 2007.

K. Kowsari, K. Jafari Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, D. Brown, Text classification algorithms: A survey. *Information*, *10*(4), 150, 2019.

M. Koerner, M. Schoop, Sentiment-Based Assessment Of Electronic Mixed-Motive Communication - A Comparison Of Approaches. *UK Academy for Information Systems Conference Proceedings*, 25, 2014.

O. Kramer, *K-Nearest Neighbors*. In *Dimensionality Reduction with Unsupervised Nearest Neighbors*. Intelligent Systems Reference Library (vol 51). Springer, Berlin, Heidelberg, 2013.

C. Laubert, J. Parlamis, Are you angry (happy, sad) or aren't you? Emotion detection difficulty in email negotiation. *Group Decision and Negotiation* (vol. 28, pp. 377-413), 2019.

R. J. Lewicki, B. Barry, D. M. Saunders, *Negotiation*. McGraw-Hill, 2020.

R. J. Lewicki, B. Polin, The role of trust in negotiation processes. In R. Bachmann, A. Zaheer (Eds.). *Handbook of advances in trust research* (pp. 29-54). Cheltenham: Edward Elgar Publishing, 2013.

M. Lewis, D. Yarats, Y. N. Dauphin, D. Parikh, D. Batra, Deal or no deal? end-to-end learning for negotiation dialogues. *arXiv preprint arXiv:1706.05125*, 2017.

Y. Li, T. Yang, Word embedding for understanding natural language: a survey. *Guide to big data applications* (pp. 83-104), 2018.

R. M. Liebert, W. P. Smith, J. Hill, M. Keiffer, The effects of information and magnitude of initial offer on interpersonal negotiation. *Journal of Experimental Social Psychology*, *4*(4), 431-441, 1968.

F. Lopes, H. Coelho, Strategic and Tactical Behaviour in Automated Negotiation. *International Journal of Artificial Intelligence* (vol. 4, 35-63), 2010.

O. Loyola-Gonzalez, Black-box vs. white-box: Understanding their advantages and weaknesses from a practical point of view. *IEEE access* (vol. 7, pp. 154096-154113), 2019.

L. Mashayekhy, M. A. Nematbakhsh, B. T. Ladani, E-negotiation model based on data mining. *Proceedings of the IADIS e-Commerce 2006 International Conference*, pp. 369-373, 2006.

K. P. Murphy, *Machine learning: A probabilistic perspective*. Cambridge: The MIT Press, 2012.

A. B. Musa, Comparative study on classification performance between support vector machine and logistic regression. *International Journal of Machine Learning and Cybernetics* (vol. 4, pp. 13-24), 2013.

V. Nasteski, An overview of the supervised machine learning methods. *Horizons* (vol. 4, pp. 51-62), 2017.

M. Olekalns, P. L. Smith, Understanding optimal outcomes. The role of strategy sequences in competitive negotiations. *Human Communication Research*, *26*(4), pp. 527-557, 2000.

M. Olekalns, L. R. Weingart. Emergent negotiations: Stability and shifts in negotiation dynamics. *Negotiation and Conflict Management Research*, *1*(2), 135-160, 2008.

A. Parmar, R. Katariya, V. Patel, A review on random forest: An ensemble classifier. In *International conference on intelligent data communication technologies and internet of things (ICICI)* (pp. 758-763). Springer International Publishing, 2019.

D. G. Pruitt, S. A. Lewis, Development of integrative solutions in bilateral negotiation. *Journal of Personality and Social Psychology*, *31*(4), 621–633, 1975.

L. L. Putnam, Small group work climates: A Lag-Sequential Analysis of Group Interaction. *Small Group Behavior*, *14*(4), 465-494, 1983.

L. L. Putnam, T. S. Jones, The role of communication in bargaining. *Human Communication Research*, *8*(3), 262-280, 1982.

Y. Qu, S. O. Cheung, Logrolling “win–win” settlement in construction dispute mediation. *Automation in Construction* (vol. 24, pp. 61-71), 2012.

I. Ritov, Anchoring in Simulated Competitive Market Negotiation. *Organizational Behavior and Human Decision Processes*, *67*(1), 16-25, 1996.

M. Schoop, F. Koehne, K. Ostertag, Communication quality in business negotiations. *Group Decision Negotiation*, *19*(2), 193-209, 2010.

M. Schoop, Negoisst: complex digital negotiation support. *Handbook of group decision and negotiation* (pp. 1149-1167), 2021.

M. Schoop, Negotiation communication revisited. *Central European Journal of Operations Research*, *29*(1), 163-176, 2021.

M. Schoop, A. Jertila, T. List, Negoisst: a negotiation support system for electronic business-to-business negotiations in e-commerce. *Data & Knowledge Engineering*, *47*(3), 371-401, 2003.

M. Schoop, M. van Amelsvoort, J. Gettinger, M. Koerner, S. T. Koeszegi, P. van der Wijst, The interplay of communication and decisions in electronic negotiations: communicative decisions or decisive communication? *Group Decision and Negotiation* (vol. 23, pp. 167-192), 2014.

A. Sherstinsky, Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D: Nonlinear Phenomena* (vol. 404, p. 132306), 2020.

A. Singh, N. Thakur, A. Sharma, A review of supervised machine learning algorithms. In *2016 3rd international conference on computing for sustainable global development (INDIACom)* (pp. 1310-1315). IEEE, 2016.

S. Singh, P. Gupta, Comparative study ID3, cart and C4. 5 decision tree algorithm: a survey. *International Journal of Advanced Information Science and Technology (IJAIST)*, *27*(27), 97-103, 2014.

M. Sokolova, S. Szpakowicz, Strategies and language trends in learning success and failure of negotiation. *Group Decision and Negotiation*, (vol. 16, pp. 469-484), 2007.

A. A. Soofi, A. Awan, Classification techniques in machine learning: applications and issues. *Journal of Basic*

& Applied Sciences, *13*(1), 459-465, 2017.

M. Ströbel, C. Weinhardt, The Montreal Taxonomy for Electronic Negotiations. *Group Decision and Negotiation*, *12*(2), 143-164, 2003.

A. F. Stuhlmacher, M. V. Champagne, The Impact of Time Pressure and Information on Negotiation Process and Decisions. *Group Decision and Negotiation* (vol. 9, pp. 471-491), 2000.

A. Thampi, *Interpretable AI: Building explainable machine learning systems*. Simon and Schuster, 2022.

L. L. Thompson, J. Wang, B. C. Gunia, Negotiation. *Annual review of psychology* (vol. 61, pp. 491-515), 2010.

D. Tomar, S. Agarwal, A comparison on multi-class classification methods based on least squares twin support vector machine. *Knowledge-Based Systems* (vol. 81, pp. 131-147), 2015.

R. Vetschera, S. T. Koeszegi, M. Filzmoser, Methods to Analyze Negotiation Processes. *Handbook of Group Decision and Negotiation* (pp. 39-60), 2021.

L. R. Weingart, L. L. Thompson, M. H. Bazerman, J. S. Carroll, Tactical behavior and negotiation outcomes. *International Journal of Conflict Management*, *1*(1), 7-31, 1990.

L. R. Weingart, P. Smith, M. Olekalns, Quantitative coding of negotiation behavior. *International negotiation*, *9*(3), 441-456, 2004.

24th International Conference on Group Decision and Negotiation &
10th International Conference on Decision Support System Technology
Human-Centric Decision and Negotiation Support for Societal Transitions

Towards a Zero-Trust Architecture for Negotiation Support Systems

Dominik Schoop [1*] and Mareike Schoop [2]

[1] Esslingen University of Applied Sciences, Esslingen, Germany
[2] University of Hohenheim, Stuttgart, Germany
[*] dominik.schoop@hs-esslingen.de

Abstract

Negotiation Support Systems (NSSs) provide many functionalities to negotiators, ranging from communication support to decision support, conflict management, and contract management. To offer such powerful support, shared negotiation data and private negotiation data is required. For negotiators to use an NSS, they have to trust the system and the system operator. To meet this trust precondition, NSSs are implemented as Trusted Third Parties. We will present an architecture that does not require trust in the NSS operator which makes NSSs more widely applicable. Such zero trust architecture is realised through homomorphic encryption. We implement a proof of concept to identify further technological challenges, estimate the performance of the solution and prepare for an operational implementation.

Keywords: negotiation support systems; homomorphic encryption; zero trust; proof of concept

1. Introduction

Negotiation support systems (NSSs) such as the well-known NSS Negoisst are established digital systems for supporting human negotiators in complex digital negotiation processes (Schoop, 2021, 2010; Schoop et al., 2003).

All NSSs are implemented as web-based client-server systems with the server containing the system and the client containing the user interface. All negotiations consist of shared data and private data. Whilst the shared data can be seen and used by all negotiation partners involved, the private data is only available for the negotiator who own that data. Examples of private data are preferences, goals, BATNA. The utility function for each negotiator is computed based on shared data and public data; the function itself is private for each negotiator (Vetschera, 2015).

Private data is a concept that is implemented in an NSS. Whilst the NSS prevents a negotiator from seeing the partner's private data, nothing prevents a malicious NSS operator to misuse the private data of negotiators. Thus, the implemented Trusted Third Party (TTP) approach only works if all users trust the system and technology as well the NSS operator (Schoop and Schoop, 2023).

To avoid such precondition of trust, we introduced a system architecture for privacy-preserving negotiations providing an advantage regarding data security and privacy over traditional NSSs (Schoop and Schoop, 2023). The data of the negotiation partners in the NSS is protected against disclosure by attacks on the NSS as well as against malicious disclosure of the NSS operator. In addition, the privacy protection laws, such as the European GDPR might be easier to fulfil in international negotiations and with international NSS providers.

The ultimate goal of the suggested technical approach is a zero-trust architecture for an NSS where negotiators need to trust in the technology used, as they do today as well, but do not need to trust the NSS operator and their potentially unknown sub-contractors. The zero-trust architecture has to fulfil the following requirements (Schoop and Schoop, 2023):

R1. Communication between negotiators and the NSS is cryptographically secured so that adversaries cannot get access to sensitive data during data transmission.

R2. The NSS is protected against unauthorised access. Each negotiator has access to their negotiation traces (i.e. all messages containing offers, counter-offers, questions, clarifications, acceptance, rejection etc.) only. Internet users who are not negotiators have no access at all.

R3. A negotiator does not know which other negotiators and organisations are using the NSS, except for those negotiators and organisation they are negotiating with.

R4. A negotiator does not know the private data of other negotiators.

R5. The NSS operator does not know the content of negotiation traces.

R6. The NSS operator does not know the private data of the negotiators.

Requirements R1 to R4 need to be fulfilled by any secure NSS which maintains the privacy of the negotiators. R1 is often fulfilled for internet systems by a transport encryption such as TLS, which, if configured correctly, is considered sufficiently secure. Requirements R2, R3 and R4 are usually implemented via secure authentication and authorisation. Negotiators are allowed only access to the necessary data. Authentication and authorisation will not be effective for requirements R5 and R6 since the NSS operator will have authorisation and the technical opportunity to access any data in the NSS for operational purposes. Therefore, requirements R5 and R6 ensure that a malicious NSS operator will not succeed as an adversary. The technology we have selected to implement the two requirements is homomorphic encryption. Trust in the technology and the correct implementation of the technology can be supported by a commercial institution in cooperation with a government body, who vouch for the compliance of the technology and its implementation to the requirements. Such approach is taken, for example, in the safety domain.

The contribution of the paper is to show in a proof of concept that functions of an NSS can be implemented with homomorphic encryption without too much computational overhead.

The following section introduces the technique and challenges of homomorphic encryption. Section 3 describes the proof of concept by implementing one function of an NSS with homomorphic encryption. We summarise the results in Section 4.

2. Homomorphic Encryption

As motivated in the introduction, an NSS uses a centralised architecture where negotiators use a central data storage and central computing resources to document and support their negotiations. Using centralised computing resources without disclosing the content of the data computed is the central application of homomorphic encryption.

Under a correct homomorphic computation we understand that for the plaintexts m_1 and m_2 there are operations $\oplus$ and $\oplus_{HE}$ such that $Dec\big(Enc(m_1) \oplus_{HE}\ Enc(m_2)\big) = m_1 \oplus m_2$ where *Enc* is the encryption function and *Dec* is the decryption function of some homomorphic encryption scheme. With homomorphic encryption a computation that can be carried out on plaintexts can also carried out on ciphertexts leading to the same result (Armknecht et al., 2015).

The common intuition about encrypted data seems to forbid correct homomorphic computation on encrypted data. This is true for encryptions which rely solely on seemingly random permutations and transpositions on the plaintext as it is done, for example, in the widely used Advanced Encryption Standard (AES). However, when plaintext and ciphertext become objects of a mathematical structure the homomorphic properties can be a by-product. For example, the classical RSA encryption uses modular exponentiation of the plaintext m with the key $pk = (n, e)$ to obtain the ciphertext c: $Enc_{pk}(m) = c = m^e \; mod \; n$ where n is the product of two large distinct primes. Consequently, given $Enc_{pk}(m_1) = c_1 = m_1^e \ mod \ n$ and $Enc_{pk}(m_2) = c_2 = m_2^e \ mod \ n$, $Enc_{pk}(m_1) \cdot Enc_{pk}(m_2) = c_1 \cdot c_2 \ mod \ n = (m_1 \cdot m_2)^e \ mod \ n = Enc_{pk}(m_1 \cdot m_2)$. However, such encryption schemes are only partially homomorphic in the sense that only specific mathematical operations can be carried out on the encrypted data.

Gentry (2009) showed that homomorphic encryption schemes can be constructed which are not constrained in the types of mathematical operation that can be carried out. However, in a somewhat homomorphic encryption scheme, the number of times an operation can be carried out on a ciphertext is limited while a fully homomorphic encryption scheme supports any number of operations on ciphertexts.

Many of the current homomorphic encryption schemes, e.g. BGV, BFV, CKKS, rely on the hardness assumption of the Learning with Error problem (LWE) or its ring variant (RLWE) (Cheon et al., 2017). The encryption of the plaintext "hides" the ciphertext in noise. Operations on the ciphertexts increase the noise. If the amount of noise gets too large, the plaintext will not be recoverable from the ciphertext any more. In a levelled homomorphic encryption scheme, the number of operations is restricted not to reach the maximal noise level. A fully homomorphic encryption scheme uses a procedure called "bootstrapping" to reduce the noise or to increase the maximal level so that more operations can be applied. The bootstrapping procedure is usually computationally costly which makes it necessary to select an appropriate homomorphic encryption scheme for the current purpose.

The reluctance in the application of fully homomorphic encryption schemes were their bad computational performance and lack of expressiveness. The expressiveness of a number-based scheme (in contrast to a gate-based scheme) is constrained by the mathematical operations it supports. CKKS, for example, provides addition, subtraction and multiplication only (Cheon et al. 2017). Since CKKS works on fixed-point numbers, the division of a plaintext m_1 by some plaintext m_2 can be homomorphically approximated by a multiplication $Enc(m_1) \cdot Enc(\frac{1}{m_2})$. While the square of a ciphertext can easily be computed, the square root cannot and needs to be approximated if necessary. Consequently, any complex mathematical function to be carried out on encrypted data must be expressed or approximated by the operations available.

The multiplications will increase the noise considerably more than addition requiring more frequent applications of the bootstrapping procedure, which is computationally costly. The mathematical functions need to be optimised regarding their use of the operations, e.g. calculating $\mathrm{a} = \mathrm{x}^2$, $\mathrm{b} = \mathrm{a}^2$, $\mathrm{c} = \mathrm{x} \cdot \mathrm{b}$, $\mathrm{d} = \mathrm{a} \cdot \mathrm{b} \cdot \mathrm{c} = \mathrm{x}^{11}$ requires five multiplications less than explicitly calculating $\mathrm{x}^{11} = \mathrm{x} \cdot \ldots \cdot \mathrm{x}$. A comparison of the performance of implemented homomorphic encryption schemes shows CKKS to outperform other schemes in some applications [Fawaz et al., 2021]. CKKS operates on fixed-point numbers while other schemes operate on integers or binary data. Therefore, for the present proof of concept, we employ CKKS. In CKKS the result of a homomorphic operation leads to an approximate result. However, since the size of the error of the approximate result can be controlled, the scheme can be configured to result in irrelevant errors.

3. Proof of Concept

We implement a proof of concept to identify further technological challenges, to estimate the performance of the solution and to prepare for an operational implementation. The generic system architecture for a bilateral negotiation is shown in Figure 1. Negotiators A (NA) and B (NB) use clients to interact with the NSS server, which interacts with the data storage which contains the private and shared negotiation data. To keep things simple, we assume that NA and NB share a common pair of asymmetric cryptographic keys, a secret key sk and a public key pk. Consequently, an attacker who has gained access to the NSS server is not able to read the negotiation data, unless the attacker is one of NA and NB. This architecture also still assumes the NSS operator to be trusted and not to forward NA's private encrypted data to NB, who could decrypt the private data. Future versions of the system will consider an architecture, where NA and NB will use distinct keys and thereby protect encrypted data better against attacks from each other.

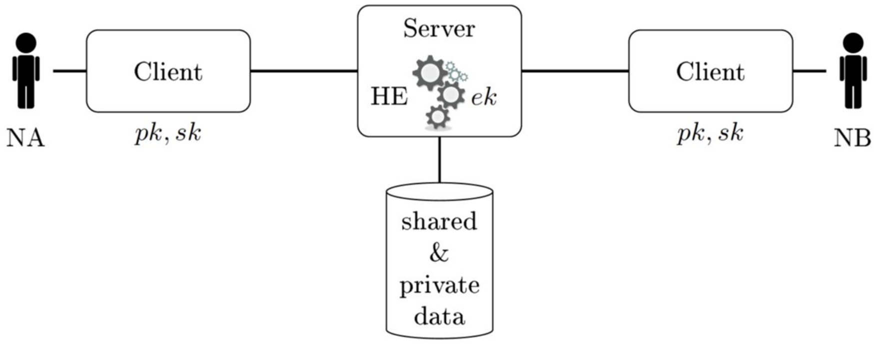


Figure 1: System architecture for a bilateral negotiation with homomorphic encryption

The client implements the communication with the NSS server and provides encryption and decryption functionalities. The NSS server provides the privacy-preserving computation on the shared negotiation data that cannot be assigned to one of the negotiators for privacy reasons. The NSS server has to compute functions over the encrypted private data of NA and NB using an evaluation key ek. The secure distribution of the keys is out of scope of the proof of concept and topic of future research.

In our proof of concept implementation, we consider the utility function

$$u(\mathrm{x}_1, \ldots, \mathrm{x}_\mathrm{n}) = \sum_{k=0}^{n} \mathrm{w}_\mathrm{k} \cdot u_k(x_k)$$

where w_k are the weights and $u_k(\cdot)$ are the single-attribute utility functions. We assume the single-attribute utility functions to be linear on the attribute value ranges, i.e. given the acceptable range $min_k, \ldots, max_k$ of the attribute k the single-attribute utility function is defined as $u_k(x_k) = (x_k - \mathrm{min_k})/(max_k - min_k)$.

We implement the function in a stand-alone proof-of-concept program that implements the homomorphic capabilities of the NSS with its clients and servers. The program does not provide an NSS nor is it integrated in an existing NSS at present. It solely serves the purpose to show that the function can be computed homomorphically and to give an indication whether the computational performance does not prohibit a usable system.

We employ the capabilities of the open-source fully homomorphic encryption library OpenFHE (https://www.openfhe.org/) (Badawi et al., 2022) for the homomorphic computation of the function.

Since CKKS does not provide homomorphic division the utility function can be calculated by multiplying $(x_k - min_k)$ with the encrypted value $1/(max_k - min_k)$. Since this value cannot be computed by the NSS server from the encrypted values max_k and min_k, the negotiator can provide the encrypted value $1/(max_k - min_k)$ once at the beginning of the negotiation. Consequently, the NSS can calculate the utility function using the encrypted values which are provided by the negotiators.

The encrypted calculation of the utility function takes far longer than an unencrypted calculation. However, the calculation time of several utility functions with a standard PC is still in an acceptable range and, therefore, it is feasible for an NSS. The calculation time certainly can be further improved by sensible choice of parameters for the cryptographic scheme and tuning of algorithms.

4. Conclusions

Negotiation Support Systems provide various forms of important support for complex digital negotiations. To do so, data on each step of the (sometimes lengthy) negotiation processes is required. Data must be protected from unauthorised access and modification. Whilst data can be shared between the negotiation partners involved in one negotiation, entities that are not involved in the process must not be able to access the data. Private data is only available for one particular negotiation partner. All other partners and all other entities must be prevented from accessing the data. These access rights are implemented in a Trusted Third Party (TTP) architecture. The assumption is that the NSS operator can be trusted and will not misuse the data they can access. The NSS operator is never malicious in a TTP approach.

This assumption can be a significant obstacle for NSS users. To remove this precondition, we introduced a different approach that does not require the users to trust the operator. The so-called zero trust architecture uses homomorphic encryption to implement the approach. We have provided a proof of concept showing that the functionality is not affected by the zero trust architecture. An NSS based on that architecture can still offer the same support but ensures that the system operator has no access to any of the negotiators' data. Even a malicious operator is thus prevented from misusing the data.

The next steps include research in the key distribution in the architecture, a multi-key approach to protect the data better, the complete implementation of the architecture in Negoisst and empirical investigations into its usability.

References

F. Armknecht, C. Boyd, Ch. Carr, K. Gjøsteen, A. Jäschke, Ch. Reuter, M. Strand, A Guide to Fully Homomorphic Encryption, Cryptology ePrint Archive, Paper 2015/1192, 2015

A. A. Badawi, A. Alexandru, J. Bates, F. Bergamaschi, D. B. Cousins, S. Erabelli, N. Genise, S. Halevi, H. Hunt, A. Kim, Y. Lee, Z. Liu, D. Micciancio, C. Pascoe, Y. Polyakov, I. Quah, K. Rohloff, J. Saylor, D. Suponitsky, M. Triplett, V. Vaikuntanathan, V. Zucca, OpenFHE: Open-Source Fully Homomorphic Encryption Library, Cryptology ePrint Archive, Paper 2022/915, 2022

J. H. Cheon, A. Kim, M. Kim, Y. S. Song, Homomorphic Encryption for Arithmetic of Approximate Numbers

2017, *Advances in Cryptology - ASIACRYPT 2017 - 23rd International Conference on the Theory and Applications of Cryptology and Information Security*, Vol. 10624, Lecture Notes in Computer Science, Springer, 409-437, 2017

S. M. Fawaz, N. Belal, A. ElRefaey, M. W. Fakhr, A Comparative Study of Homomorphic Encryption Schemes Using Microsoft SEAL, ????????????????

Journal of Physics: Conference Series, Volume 2128, 6th International conference on Advanced Technology and Applied Sciences (ICaTAS 2021) 12-14 October 2021, Cairo, Egypt Citation Shereen Mohamed Fawaz et al 2021 J. Phys.: Conf. Ser. 2128 012021

C. Gentry, Fully homomorphic encryption using ideal lattices, *Proceedings of the 41st Annual ACM Symposium on Theory of Computing (STOC 2009),* 169-178, 2009.

D. Schoop, M. Schoop, Privacy-preserving electronic negotiations. *Proceedings of 23rd International Conference on Group Decision and Negotiation (GDN 2023)*, 2023.

M. Schoop, Support of complex electronic negotiations. *Handbook of group decision and negotiation* (pp. 409-423), Springer, 2010.

M. Schoop, Negoisst: complex digital negotiation support. *Handbook of group decision and negotiation* (pp. 1149-1167), Springer, 2021.

M. Schoop, A. Jertila, T. List, Negoisst: a negotiation support system for electronic business-to-business negotiations in e-commerce. *Data & Knowledge Engineering*, *47*(3), 371-401, 2003.

R. Vetschera, Preference structures and negotiator behavior in electronic negotiations, *Decision Support Systems*, 44(1), 135-146, 2007.

Group Decision and Negotiation
- Preference Modeling for Group Decision and Negotiation -

Conference papers and extended abstracts

Research on E-commerce Dual Channel Supply Chain Pricing Decision under the Context of Live Streaming

Chu Yanfeng [1*], Wang Rui [1]

[1] Nanjing University of Aeronautics and Astronautics, China
[2] Affiliation, City, Country
[*] yanfengc@nuaa.edu.cn

Abstract

This study builds e-commerce dual-channel supply chain models focused on brand-operated and influencer-operated live streaming, further categorizing influencer live streaming into markup cooperation and commission cooperation modes. The research compares optimal decisions in these models and analyzes the influence of traditional sales channel market share, consumer channel preference, influencer traffic generation, and commissions on equilibrium solutions. The findings reveal that brands should opt for influencer live streaming with markup cooperation under specific conditions related to influencer effectiveness and traditional sales channel market share. Additionally, a positive relationship is observed between optimal brand pricing, influencer traffic generation, and channel cross-price elasticity.

Keywords: Live streaming; E-commerce; Dual-channel Supply Chain; Stackelberg game; Nash equilibrium.

1. Introduction

Relying on the media technology revolution, live streaming e-commerce has established a new model for product management and mass consumption in the internet realm. Since its explosive growth in 2019, China's live streaming e-commerce industry has rapidly developed, with increasingly fierce competition. According to the "2023 (First Half) China Live Streaming E-commerce Market Data Report," the transaction volume in the first half of 2023 was about 1.9916 trillion yuan, with an expected annual transaction volume reaching 4.9157 trillion yuan, an increase of 30.44% compared to the previous year; the user base was about 520 million, expected to reach 540 million by the end of the year, growing by 14.16%, with an increasing growth rate [1].

Live streaming e-commerce can be divided into enterprise live streaming and influencer live streaming based on the source of products. Enterprise live streaming relies on self-operated channels of own brands, ensuring control over product attributes and the professionalism of the presenters, achieving refined customer operation in private traffic pools. Influencer live streaming, depending on personal fan bases, covers a broader marketing range and can direct traffic to other sales channels for brands, but faces challenges like multiple competing brands in the same streaming room, limited time for individual product explanation, and significant commission rates. For brands, choosing the appropriate form of live streaming and how to collab-orate with influencer hosts is a pressing issue.

Current domestic and international research on live streaming e-commerce mainly focuses on consumer purchasing behavior and live streaming marketing strategies. Research on consumer behavior predominantly revolves around the influence of host characteristics and live stream room interactions on consumer purchasing intentions. Many scholars have found that factors like influencer homogeneity, host influence, professionalism, and entertainment value all impact consumer purchasing intentions [2 – 5]. From the perspective of interactivity, Ma Linye and others[6] discovered through research on gender differences among consumers that male consumers are more satisfied with interactivity than female consumers. Social interaction is one of the motivations for consumer satisfaction [7], and many scholars have found that host interactivity and platform interactivity significantly impact consumers' psycho-logical experiences [8, 9]; Yuan Haixia and others[10] constructed an influence mechanism model of live streaming interaction based on social influence theory and the resource-based view.

In terms of live streaming marketing strategies, Gong Hanyu and others [11] studied the live streaming strategies of online retailers under multi-channel sales, proposing different live streaming marketing strategies for standardized and personalized products. Lin Zhibing and others [12] found that both manufacturers and retailers prefer live streaming marketing strategies when influencer hosts' level of social responsibility and the spillover effect coefficient of live streaming marketing are high, through comparing traditional marketing strategies and live streaming marketing strategies. Wang Chenyu and others [13] discussed manufacturers' sales mode choices and live streaming marketing strategies under different sales modes by constructing a game model and analyzing equilibrium results, finding that both platform-operated store distribution and flagship store direct sales live streaming marketing increase retail prices and profits for manufacturers and e-commerce platforms; the higher the pro-portion of customers watching live streams or product similarity, the higher the in-vestment in live streaming, retail prices, and profits for supply chain members.

Through summarizing the above literature, this paper, in the context of integrating live streaming, establishes three live streaming supply chain models: brand-operated live streaming, influencer live streaming with markup cooperation, and influencer live streaming with commission cooperation. It uses Stackelberg game and Nash equilibrium to solve and compare the optimal decisions and profits of these three supply chains, analyzing the influence of traditional sales channel market share, consumer channel preference, influencer traffic generation, and commission on the equilibrium solutions of the supply chains, and studies how brands should choose live streaming models and cooperate with influencers.

2. Descriptions and Assumptions of Three Supply Chain Models

2.1. Description of Models

In influencer live-streaming sales, factors such as the professional skills of the influencer host, their emotional influence, and their celebrity effect can enhance fans' affection for a brand. However, due to the delayed effect of influencer advertising and the randomness of live-stream content and timing, consumers cannot always enter the live room to make purchases. This creates a spillover effect θ on other sales channels of the brand through the influencer live-stream channel. Moreover, considering that the main forms of collaboration between influencers and brands are either markup cooperation or commission cooperation, this paper further subdivides the supply chain model under influencer live-streaming into two modes: influencer live-streaming with markup cooperation and influencer live-streaming with commission cooperation. Figure 1 describes the operation of the supply chain in these three scenarios.

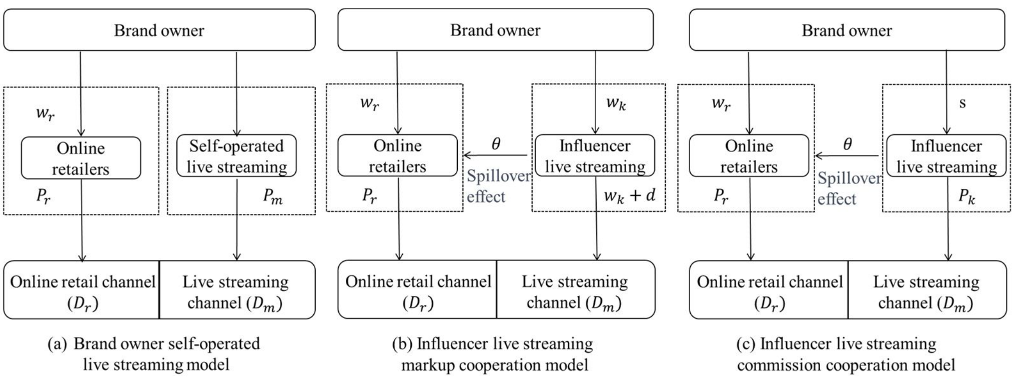


Figure. 1. Structure of the dual-channel e-commerce supply chain model under three scenarios.

In the brand's self-operating model, the brand sells its products to online retailers at a wholesale price of w_r through the direct online sales channel, and the online retailer then sells them to consumers at a price of P_r. Meanwhile, in the live-streaming channel, the brand sets the live-streaming price P_m, at which consumers purchase the product. In the model involving hiring influencer hosts, the supply chain includes a brand, an online retailer, and an influencer live-streaming room. The direct online sales channel is the same as in the self-operating model, but in the live-streaming channel, the cooperation model between the brand and the influencer host differs, necessitating the discussion of two scenarios:

1. Influencer Markup Model: The brand first wholesales the product to the influencer host at a price of w_m. The influencer host then adds a markup *d* to sell the product to consumers, earning a profit from the markup.

2. Commission Model: The brand sells the product at a price of P_k through the influencer's live-streaming room and pays the influencer host a certain percentage (*s*) of commission and a slot fee *F*.

2.2. Assumptions

To facilitate subsequent research, this paper proposes the following assumptions.

Assumption 1. The demand for the product is influenced by the product's price and the competitive prices across channels. The demand functions for each channel under the three models are as follows:

Self-operated Live Streaming Model (SL). The demand functions for the direct online sales channel and the live streaming channel are respectively:

$$D_r = \alpha A - \beta P_r + \mu P_m,\ \ D_m = (1-\alpha)\mathrm{A} - \beta P_m + \mu P_r$$

Influencer Live Streaming with Markup Cooperation Model (IM). The demand func-tions for the direct online sales channel and the influencer live streaming channel are respectively:

$$D_r = \alpha A - \beta P_r + \mu(w_k + d) + \theta,\ \ D_k = (1-\alpha)\mathrm{A} - \beta(w_k + d) + \mu P_r$$

Influencer Live Streaming with Commission Cooperation Model (IC). The demand functions for the direct online sales channel and the influencer live streaming channel are respectively:

$$D_r = \alpha A - \beta P_r + \mu P_k + \theta,\ \ D_k = (1-\alpha)A - \beta P_k + \mu P_r$$

Where, D_i represents the market demand for each sales channel, with i∈{r,m,k}, where r is the online retail channel, m is the brand's self-operated live streaming channel, and k is the influencer live streaming channel; A represents the potential market capacity of the product; β is the coefficient of the effect of the unit retail price of the product on its own demand; μ is the cross-price elasticity coefficient between online sales channels; θ is the spillover effect of influencer live streaming on traditional online direct sales channels; P_i is the product selling price for each sales channel, i∈{r,m,k}; w_i is the wholesale price of the product for each channel, i∈{r,m,k}; d is the markup portion of the product selling price in the influencer live streaming channel under the markup cooperation model.

Assumption 2. As the research object of this paper is the live streaming model and the cooperation model with influencer hosts, it is assumed that the brand sells only one product, without considering the production cost of the product and the effort cost of the influencer host, and it is assumed that the brand holds a dominant position in this supply chain.

Assumption 3. This paper represents the market share of traditional online sales channels with α. Since live streaming is still an auxiliary method to traditional sales channels, the market share of the live streaming channel will not exceed that of traditional online sales channels in the short term. Therefore, it is assumed that 0.5<α≤1. In fact, according to related statistical analysis, the market share of traditional online retail channels is around 0.8.

Assumption 4. Assume that consumers' sensitivity to the product's own price is greater than their sensitivity to price competition between channels, that is, β>μ.

3. Optimal Pricing and Profit Solutions for the Three Supply Chain Models

3.1. Brand's Self-Operated Live Streaming Model

In the self-operated live streaming model, the optimization problems for the online retailer and the brand can be represented as follows:

$$\max_{P_r} \pi_R^{SL} = (P_r - w_r)(\alpha A - \beta P_r + \mu P_m)$$

$$\max_{w_r,\ P_m} \pi_M^{SL} = w_r(\alpha A - \beta P_r + \mu P_m) + P_m((1-\alpha)A - \beta P_m + \mu P_r)$$

Proposition 3.1. The profit π_R of the online retailer is a strictly concave function of P_r, and the profit π_M of the brand is a strictly concave function of w_r and P_m. By solving backward according to the aforementioned game sequence, the optimal pricing for the product by the brand can be determined as:

$$w_r = \frac{A(\mu + \beta\alpha - \alpha\mu)}{2(\beta^2 - \mu^2)}$$

$$P_m = \frac{A(\beta - \beta\alpha + \alpha\mu)}{2(\beta^2 - \mu^2)}$$

The optimal pricing for the retailer is:

$$P_r = \frac{A(2\beta\mu + \alpha(\beta - \mu)(3\beta + \mu))}{4\beta(\beta^2 - \mu^2)}$$

By substituting w_r, P_m, and P_r into the respective profit functions of the online retailer and the brand, the optimal profits can be obtained as:

$$\pi_R^{SL} = \frac{\alpha^2 A^2}{16\beta}$$

$$\pi_M^{SL} = \frac{A^2(2\beta^2 - 4\alpha\beta^2 + 3\alpha^2\beta^2 + 4\alpha\beta\mu - 4\alpha^2\beta\mu + \alpha^2\mu^2)}{8(\beta^3 - \beta\mu^2)}$$

3.2. Influencer Live Streaming with Markup Cooperation Model

Under the influencer live streaming with markup cooperation model, the optimization problems for the online retailer, the brand, and the influencer can be respectively represented as:

$$\max_{P_r} \pi_R^{IM} = (P_r - w_r)(\alpha A - \beta P_r + \mu(w_k + d) + \theta)$$

$$\max_{w_r,\ w_k} \pi_M^{IM} = w_r(\alpha A - \beta P_r + \mu(w_k + d) + \theta) + w_k((1-\alpha)A - \beta(w_k + d) + \mu P_r)$$

$$\max_{d} \pi_K^{IM} = d((1-\alpha)A - \beta(w_k + d) + \mu P_r)$$

Proposition 3.2. The profit π_R of the online retailer is a strictly concave function of P_r, the profit π_M of the brand is a strictly concave function of w_r and w_k, and the profit π_K of the influencer is a strictly concave function of d. By solving backward according to the aforementioned game sequence, the optimal pricing for the product by the brand can be determined as:

$$w_r = \frac{A\alpha\beta + \beta\theta + A\mu - A\alpha\mu}{2(\beta^2 - \mu^2)}$$

$$w_k = \frac{A\beta - A\alpha\beta + A\alpha\mu + \theta\mu}{2(\beta^2 - \mu^2)}$$

The optimal pricing for the retailer is:

$$P_r = \frac{6\beta^3(A\alpha + \theta) - 5A(-1 + \alpha)\beta^2\mu - 3\beta(A\alpha + \theta)\mu^2 + 2A(-1 + \alpha)\mu^3}{2(4\beta^4 - 5\beta^2\mu^2 + \mu^4)}$$

The optimal pricing for the influencer host is:

$$d = \frac{-2A(-1 + \alpha)\beta + A\alpha\mu + \theta\mu}{8\beta^2 - 2\mu^2}$$

By substituting w_r, w_k, P_r, and d into the respective profit functions of the online retailer, the brand, and the influencer, the optimal profits can be obtained as:

$$\pi_R^{IC} = \frac{\beta(2\beta\theta + A(2\alpha\beta + \mu - \alpha\mu))^2}{4(4\beta^2 - \mu^2)^2}$$

$$\pi_M^{IC} = \frac{\beta\begin{pmatrix}2\beta^2(A^2(1 + 2(\alpha - 1)\alpha) + 2A\alpha\theta + \theta^2) - 6A(\alpha - 1)\beta(A\alpha + \theta)\mu \\ +(A^2(1 + 2(\alpha - 1)\alpha) + 2A\alpha\theta + \theta^2)\mu^2\end{pmatrix}}{4(4\beta^4 - 5\beta^2\mu^2 + \beta^4)}$$

$$\pi_K^{IC} = \frac{\beta(-2A(\alpha - 1)\beta + A\alpha\mu + \theta\mu)^2}{4(4\beta^2 - \mu^2)^2}$$

3.3. Influencer Live Streaming with Commission Cooperation Model

In the influencer live streaming commission model, the optimization problems for the online retailer, the brand, and the influencer host can be respectively represented as:

$$\max_{P_r} \pi_R^{IC} = (P_r - w_r)(\alpha A - \beta P_r + \mu P_k + \theta)$$

$$\max_{w_r,\ P_k} \pi_M^{IC} = w_r(\alpha A - \beta P_r + \mu P_k + \theta) + (1 - s)P_k\big((1 - \alpha)A - \beta P_k + \mu P_r\big) - F$$

$$\max \pi_K^{IC} = sP_k((1 - \alpha)A - \beta P_k + \mu P_r) + F$$

Proposition 3.3. The profit π_R of the online retailer is a strictly concave function of P_r, and the profit π_M of the brand is a strictly concave function of w_r and P_k. By solving backward according to the aforementioned game sequence, the optimal pricing for the product by the brand can be determined as:

$$w_r = \frac{(s - 1)(4\beta^2(A\alpha + \theta) + 2A(s - 2)(\alpha - 1)\beta\mu - s(A\alpha + \theta)\mu^2)}{8(s - 1)\beta^3 + (8 + (s - 8)s)\beta\mu^2}$$

$$P_k = \frac{-4A(s - 1)(\alpha - 1)\beta + (3s - 4)(A\alpha + \theta)\mu}{8(s - 1)\beta^2 + (8 + (s - 8)s)\mu^2}$$

The optimal pricing for the retailer is:

$$P_r = \frac{(s-1)(6\beta^2(A\alpha+\theta)+A(s-4)(\alpha-1)\beta\mu-2(A\alpha+\theta)\mu^2)}{8(s-1)\beta^3+(8+(s-8)s)\beta\mu^2}$$

By substituting w_r, P_r, and P_k into the respective profit functions of the online retailer, the brand, and the influencer, the optimal profits can be obtained as:

$$\pi_R^{IC} = \frac{(s-1)^2(2\beta^2(A\alpha+\theta)-As(\alpha-1)\beta\mu+(s-2)(A\alpha+\theta)\mu^2)^2}{\beta(8(s-1)\beta^2+(8+(s-8)s)\mu^2)^2}$$

$$\pi_M^{IC} = \frac{\begin{gathered}(1-s)(\beta^2(A^2(-2+2s(\alpha-1)^2+4\alpha-3\alpha^2)-2A\alpha\theta-\theta^2)\\ -A(3s-4)(\alpha-1)\beta(A\alpha+\theta)\mu+(s-1)(A\alpha+\theta)^2\mu^2)\end{gathered}}{8(s-1)\beta^3+(8+(s-8)s)\beta\mu^2} - F$$

$$\pi_K^{IC} = \frac{\begin{gathered}(s(4A(s-1)(\alpha-1)\beta-(3s-4)(A\alpha+\theta)\mu)(4A(s-1)(\alpha-1)\beta^3-\\ (3s-2)\beta^2(A\alpha+\theta)\mu-A(3s-4)(\alpha-1)\beta\mu^2+2(s-1)(A\alpha+\theta)\mu^3))\end{gathered}}{\beta(8(s-1)\beta^2+(8+(s-8)s)\mu^2)^2} + F$$

3.4. Comparison and Key Parameter Analysis of Optimal Decisions under Three Live Streaming Models

Proposition 3.4. To ensure practical feasibility, the commission rate s must satisfy $0 < s < 4 + \frac{2(-2\beta^2+\sqrt{2}\sqrt{2\beta^4-3\beta^2\mu^2+\mu^4})}{\mu^2}$.

Proof: (1) $P_k > 0$

$$\begin{cases} P_k = \dfrac{-4A(s-1)(\alpha-1)\beta+(3s-4)(A\alpha+\theta)\mu}{8(s-1)\beta^2+(8+(s-8)s)\mu^2} \\ S.t\ \beta > \mu > 0;\ 0 < s < 1;\ 0 < \alpha < 1; \end{cases}$$

$$\text{Let} \begin{cases} X_1 = (s-1)(-4A(\alpha-1)\beta+3(A\alpha+\theta)\mu)-(A\alpha+\theta)\mu \\ X_2 = 8(s-1)\beta^2+(8+(s-8)s)\mu^2 \end{cases},$$

Given these constraints, it can be deduced that $X_1 < 0$, Therefore, to ensure $P_k > 0$, it is required that $X_2 < 0$. $X_2 = 8(s-1)\beta^2+(8+(s-8)s)\mu^2 = \mu^2 s^2 + 8(\beta^2-\mu^2)s - 8(\beta^2-\mu^2)$, which means X_2 is symmetrical around $s = -4(\beta^2-\mu^2)$. Setting $X_2 = 0$ and solving for s yields:

$$s = \begin{cases} 4 - \frac{2(2\beta^2+\sqrt{2}\sqrt{2\beta^4-3\beta^2\mu^2+\mu^4})}{\mu^2} < 0(\text{discard}) \\ 4 + \frac{2(-2\beta^2+\sqrt{2}\sqrt{2\beta^4-3\beta^2\mu^2+\mu^4})}{\mu^2} > 0 \end{cases},$$

When $s = 1, X_2 = \mu^2 > 0$, and since X_2 increases monotonically when $s > -4(\beta^2-\mu^2)$, it can be proven that: $4 + \frac{2(-2\beta^2+\sqrt{2}\sqrt{2\beta^4-3\beta^2\mu^2+\mu^4})}{\mu^2} < 1$. In summary, to ensure $P_k > 0$, it is required that $0 < s < 4 + \frac{2(-2\beta^2+\sqrt{2}\sqrt{2\beta^4-3\beta^2\mu^2+\mu^4})}{\mu^2}$.

Similarly, it can be shown that $P_r > 0$and $w_r > 0$ only when $0 < s < 4 + \frac{2(-2\beta^2+\sqrt{2}\sqrt{2\beta^4-3\beta^2\mu^2+\mu^4})}{\mu^2}$.

Proposition 3.5. In three modes, the relationship between the optimal pricing and profit of each member and the cross-price elasticity coefficient μ exhibits the following characteristics:

(1) The optimal wholesale price of the brand owner, the optimal wholesale price of the retailer, and the selling price of products in live streaming rooms all increase with the increase of the cross-price elasticity coefficient μ.

(2) In all three modes, the profits of the brand owner and the influencer increase with the increase of the cross-price elasticity coefficient μ. Only in the influencer-led sales model, the profit of the retailer increases with the increase of μ. In the merchant self-operating live broadcast model, the retailer's profit is unrelated to μ.

Proof:

Proof of Proposition 3.5 (1) (taking the influencer live streaming commission model as an example.

$$\frac{\partial w_r}{\partial \mu}=\frac{2(s-2)(s-1)\begin{pmatrix}8A(s-1)(\alpha-1)\beta^2-4(3s-4)\beta(A\alpha+\theta)\mu\\-A(8+(s-8)s)(\alpha-1)\mu^2\end{pmatrix}}{(8(s-1)\beta^2+(8+(s-8)s)\mu^2)^2}>0$$

Similarly, it can be proven that$\frac{\partial P_r}{\partial \mu}>0, \frac{\partial P_k}{\partial \mu}>0$.

Proof of Proposition 3.5 (2) (taking the influencer live streaming commission model as an example).

$$\frac{\partial \pi_M^{IC}}{\partial \mu}=\frac{X_6(X_7+X_8)(X_9-X_{10})}{(8(s-1)\beta^2+(8+(s-8)s)\mu^2)^2}$$

Because $X_6=(s-1)<0$, $X_7=2\beta(3s-4)(A\alpha+\theta)<0$, $X_8=A\mu(8+(s-8)s)(\alpha-1)<0$, $X_9=4A\beta(s-1)(\alpha-1)>0$, $X_{10}=(3s-4)(A\alpha+\theta)\mu<0$, so $X_6(X_7+X_8)(X_9-X_{10})>0$, that is $\frac{\partial \pi_M^{IC}}{\partial \mu}>0$.

$$\frac{\partial \pi_R^{IC}}{\partial \mu}=-\frac{X_{11}(X_{12}+X_{13}+X_{14})(X_{15})}{(8(s-1)\beta^2+(8+(s-8)s)\mu^2)^3}$$

According to Proposition 3.4, it is known that 8(s-1)β^2+(8+(s-8)s)μ^2<0, therefore, $(8(s-1)\beta^2+(8+(s-8)s)\mu^2)^3<0$. And because $X_{11}=2s(s-1)^2>0$, $X_{12}=-8A(s-1)(\alpha-1)\beta^2<0$, $X_{13}=4(3s-4)\beta(A\alpha+\theta)\mu<0$, $X_{14}=A\mu^2(8+(s-8)s)(\alpha-1)<0$, $X_{15}=-2\beta^2(A\alpha+\theta)+As(\alpha-1)\beta\mu-(s-2)(A\alpha+\theta)\mu^2=\alpha(\beta-\mu)(-2A\beta+A\mu(s-2))-2\beta^2\theta-As\beta\mu+2\theta\mu^2-s\theta\mu^2<-As\mu^2-s\theta\mu^2<0$, therefore, $\frac{\partial \pi_R^{IC}}{\partial \mu}>0$.

Similarly, it can be proven that $\frac{\partial \pi_K^{IC}}{\partial \mu}>0$.

In the brand self-operating model, the profit of the brand owner $\pi_R^{SL}=\frac{\alpha^2A^2}{16\beta}$, thus it is unrelated to μ.

Proposition 3.6. In the influencer live streaming sales model, the relationship between the optimal pricing and profits of each member and the influencer's traffic generation effect θ exhibits the following characteristics:

(1) The optimal wholesale price of the brand owner, the optimal wholesale price of the retailer, and the selling price of products in the influencer's live streaming room all increase with the increase of the influencer's traffic generation effect θ.

(2) The profits of the brand owner, the retailer, and the influencer increase with the increase of the influencer's traffic generation effect θ.

Proof:

Proof of Proposition 3.6 (1) (taking the influencer live streaming commission model as an example).

$$\frac{\partial w_r}{\partial \theta}=\frac{(s-1)(4\beta^2-s\mu^2)}{\beta(8(s-1)\beta^2+(8+(s-8)s)\mu^2)}>0$$

$$\frac{\partial P_r}{\partial \theta}=\frac{(s-1)(6\beta^2-2\mu^2)}{\beta(8(s-1)\beta^2+(8+(s-8)s)\mu^2)}>0$$

$$\frac{\partial P_r}{\partial \theta}=\frac{(3s-4)\mu}{\beta(8(s-1)\beta^2+(8+(s-8)s)\mu^2)}>0$$

Proof of Proposition 3.6 (2) (taking the influencer live streaming commission model as an example).

$$\frac{\partial \pi_M^{IC}}{\partial \theta}=\frac{(s-1)(2\beta^2(A\alpha+\theta)+A(3s-4)(\alpha-1)\beta\mu-2(s-1)(A\alpha+\theta)\mu^2)}{\beta(8(s-1)\beta^2+(8+(s-8)s)\mu^2)}$$

Because $(s-1)(2\beta^2(A\alpha+\theta)+A(3s-4)(\alpha-1)\beta\mu-2(s-1)(A\alpha+\theta)\mu^2)=(s-1)((A\alpha+\theta)(2\beta^2+2(1-s)\mu^2)+A(3s-4)(\alpha-1)\beta\mu)<0$,therefore $\frac{\partial \pi_M^{IC}}{\partial \theta}>0$.

Similarly, it can be proven that $\frac{\partial \pi_R^{IC}}{\partial \theta}>0$, $\frac{\partial \pi_K^{IC}}{\partial \theta}>0$.

4. Numerical Analysis

In the following, a numerical analysis is conducted to analyze the effects of key factors such as traditional online sales channel market share, cross-price elasticity coefficient, channel spillover effect, and commission rate on equilibrium decisions and supply chain performance. Assume the base parameters are A=500, β=1.5, and conduct the numerical analysis based on this.

4.1. Impact of Traditional Online Sales Channel Market Share on Equilibrium Decisions and Profit

Set μ=0.4, θ=30, s=0.1, F=2000, and take α as the independent variable, with α∈[0,1]. The optimal decision and profit variation with the change in cross-price elasticity coefficient are shown in Figure 2. From the graph, it can be seen that the profits of the brand under the three live streaming models first decrease and then increase with the expansion of the market share of the traditional channel. Moreover, as the market share of the live streaming channel increases, the profit from the brand's self-operated live streaming gradually surpasses that of the influencer live streaming sales model.

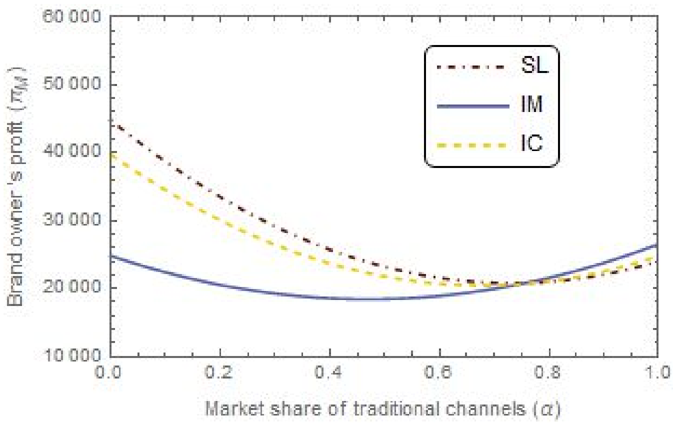
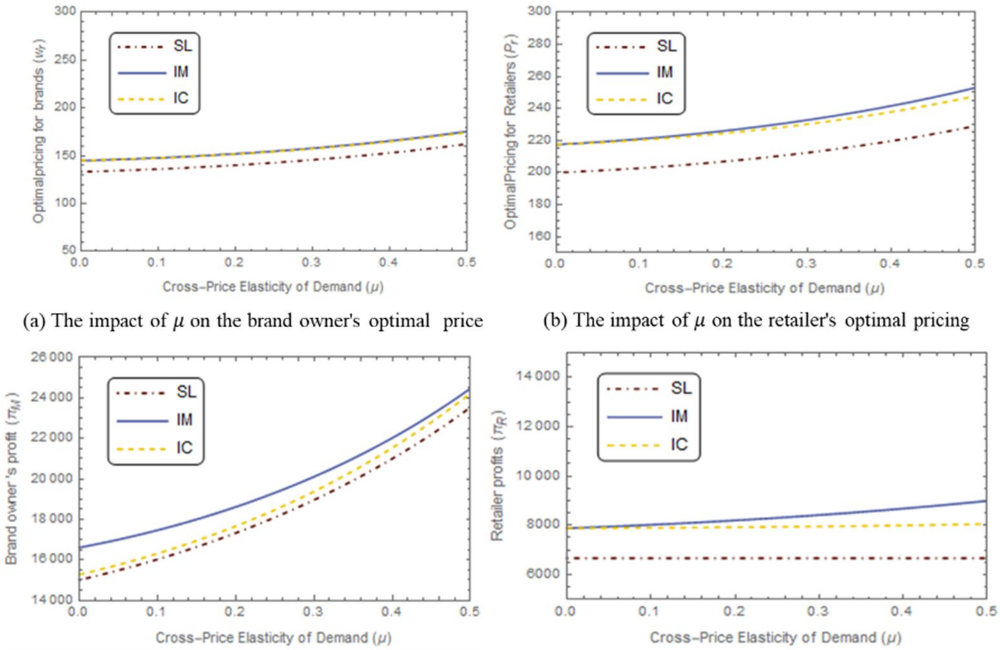


Figure 2. Impact of Traditional Online Sales Channel Market Share on Equilibrium Decisions and Profit

4.2. The Impact of Channel Cross-Price Elasticity on Equilibrium Decisions and Profits

Let α = 0.8, θ = 30, s = 0.1, F = 2000, taking μ as the variable, where μ ∈ [0,0.5], the optimal decisions and profits with the change of channel cross-price elasticity are shown in Figure 3.



(a) The impact of μ on the brand owner's optimal price (b) The impact of μ on the retailer's optimal pricing

(c) The impact of μ on the brand owner's profit (d) The impact of μ on retailer's profit

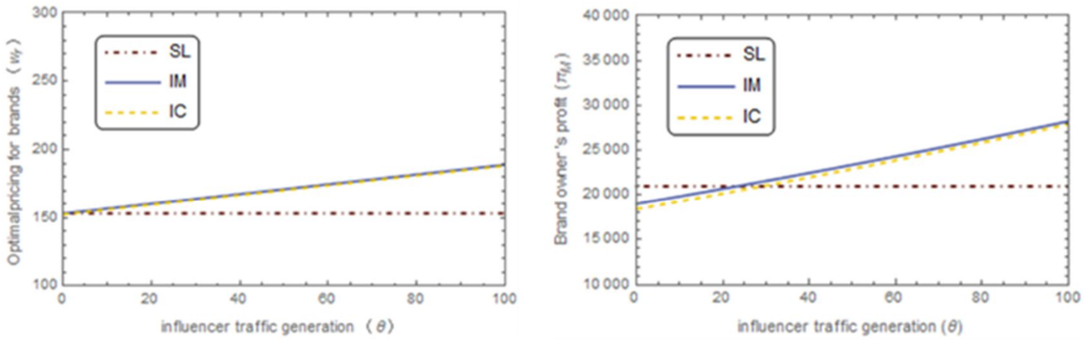
Figure 2. The Impact of Channel Cross-Price Elasticity on Equilibrium Decisions and Profits

From Figures 3(a) and 3(b), it can be seen that in all three scenarios, the brand's op-timal wholesale price has a positive relationship with the cross-price elasticity. The price difference between the influencer live streaming model and the brand's self-operated live streaming model is consistently greater in retail price difference than in wholesale price difference. This indicates that compared to the brand, the retailer's pricing behavior is more sensitive to the brand's introduction of the influencer live streaming model.

From Figures 3(c) and 3(d), it can be seen that in all three scenarios, the brand's profits increase with the increase of the channel cross-price elasticity. Under the brand's self-operated live streaming model, the retailer's profit is unrelated to the channel cross-price elasticity, while in the influencer live streaming model, the retail-er's profit increases with the increase of the channel cross-price elasticity. As the cross-price elasticity increases, the profit differences among the three scenarios for the brand gradually narrow.

4.3. The Impact of Channel Traffic Generation on Equilibrium Decisions and Profits

Set α = 0.8, μ = 0.4, s = 0.1, F = 2000, take θ as the variable, where θ ∈ [0,100], the optimal decisions and profits with the change of channel traffic generation are shown in Figure 4.



(a) The impact of θ on the brand owner's optimal price (b)The impact of θ on the brand owner's profit

Figure 3. The Impact of Channel Traffic Generation on Equilibrium Decisions and Profits

When the influencer's traffic generation is 0, meaning the influencer cannot bring traffic support to the brand, the brand's wholesale price remains consistent across all three scenarios. As the influencer's traffic generation increases, in the influencer-led sales model, the brand's optimal wholesale price also rises.

Under the influencer live streaming sales model, both the brand's and retailer's optimal profits increase as the influencer's traffic generation increases. When the influencer's traffic generation is less than 24.57, the brand's optimal profit is highest under the brand's self-operated live streaming model. However, when the influencer's traffic generation exceeds 24.57, as θ increases, the profit under the influencer live streaming sales model far exceeds that of the brand's self-operated live streaming model.

4.4. The Impact of Commission Rate on Equilibrium Decisions and Profits

Set $\alpha = 0.8$, $\mu = 0.4$, $\theta = 30$, $F = 2000$, take s as the variable, where $s \in [0,1]$, the optimal decisions and profits with the change of commission rate are shown in Figure 5.

From Figure 5, it can be seen that within a certain range, as the commission rate for the influencer increases, the profits for both the influencer and the retailer rise, while the brand's profit decreases. When the commission rate exceeds a certain range, i.e., when it is too high, the demand for the influencer live streaming sales channel will be completely lost. Therefore, for influencers, an excessively high commission rate does not bring higher earnings.

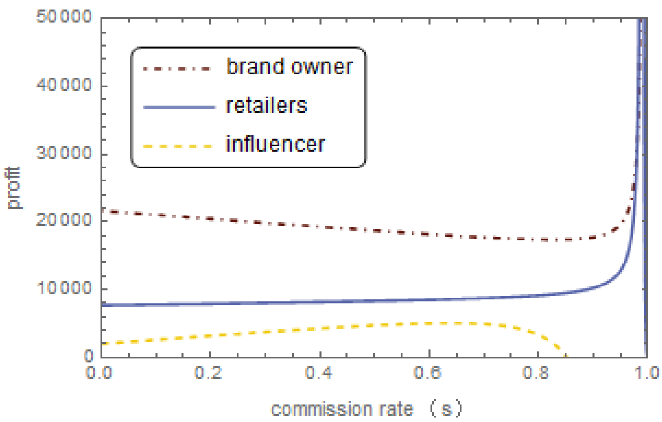


Figure 5. The Impact of Commission Rate on Equilibrium Decisions and Profits

5. Conclusion

The research results indicate that in practical situations, the optimal choice for brands is to introduce the influencer live streaming and reject ineffective low price competition. When the impact of channel traffic generation reaches a certain level and the market share of traditional sales channels are high, the markup model is more advantageous for the brand. The optimal pricing of the brand is positively correlated with the impact of channel traffic generation and consumer channel preferences.

The following managerial implications can be drawn from the analysis.

(1) Prioritize opening live streaming channels for internet celebrities and quickly expand the market. It can be seen that when traditional sales channels occupy 80% of the market share and the impact of channel traffic generation achieve a certain degree, the brand's profit under the situation of influencer live streaming sales is higher than the brand's self-operated live streaming model. Therefore, influencer live streaming sales can expand the market in the short term and increase the profit of brand merchants.

(2) Prioritize the mode of markup cooperation with influencers. The previous research indicates that in the current situation, priority should be given to cooperating with influencers through a markup model. However, under the markup cooperation model of influencers live streaming, the profit of influencers will be lower than that of commission cooperation model, which makes brands less willing to negotiate with influencers for the markup cooperation model, and cooperation may be hindered.

(3) Highlight the advantages of various channels and reject ineffective low price competition. It can be seen that when the cross price elasticity coefficient and the impact of channel traffic generation are constant, the profit of brands in the influencer live streaming mode is greater than in the self-operated mode, and the pricing of brands also shows that the online celebrity live streaming mode is greater than the self-operated live streaming mode, indicating that high profits do not necessarily require low price promotions. Therefore, for brands, there is no need to blindly pursue low prices to make customers more sensitive to prices. Instead, they should deeply explore the demand scenarios of customers shopping in various channels, maximize the advantages of each channel, weaken customer sensitivity to prices, and maintain stable and harmonious relationships among multiple channels.

Acknowledgments. This study was funded by Humanities and Social Science Research Plan-ning Project of the Ministry of Education of the People's Republic of China, "Research on the Resilience Measurement and Improvement Strategy of High-tech Manufacturing Industry Chain under Carbon Emission Reduction Policy"(grant number 23YJA630016).

References

1. Wangjing Society & China General Chamber of Commerce Live Broadcasting E-commerce Committee: "China Live Broadcasting E-commerce Market Data Report (First Half of 2023)" published, https://www.100ec.cn/zt/2023zbdsscsjbg/.
2. Zhou Rong, Tong Lei. "A Study on the Influencing Factors of Consumers' Purchase Intention During Livestreaming e-Commerce: The Mediating Effect of Emotion." Journal, 2022.
3. Liu Dawei, Yu Jinyang. "Impact of Perceived Diagnosticity on Live Streams and Consumer Purchase Intention: Streamer Type, Product Type, and Brand Awareness as Moderators." Information Technology Management, 2022, pp. 11-14.
4. Li Guangming, Yan Jiayue. "Market Literacy of Live-streaming Hosts and Its Impact on Consumer Purchase Intention." Journal of Harbin University of Commerce (Social Science Edition), 2023, Issue 04, pp. 59-70.
5. Xu Mengyang. "Research on the Relationship between Internet Celebrity Live Broadcasting, Presence, and Consumer Purchase Intention." Commercial Economy Research, 2023, Issue 13, pp. 67-70.
6. Ma Linye, Gao Shuqing, Zhang Xiaoyan. "How to Use Live Streaming to Improve Consumer Purchase Intentions: Evidence from China." Sustainability, 2022, Vol. 14, No. 2, pp. 1045-1045.

7. Yingying M. "To Shop or Not: Understanding Chinese Consumers' Live-stream Shopping Intentions from the Perspectives of Uses and Gratifications, Perceived Network Size, Perceptions of Digital Celebrities, and Shopping Orientations." Telematics and Informatics, 2021, Issue 59 (prepublish).
8. Wang Huijuan. "Interaction in Agricultural Product Live Broadcasting, Consumer Psychological Experience, and Purchase Intention." Commercial Economy Research, 2023, Issue 22, pp. 80-83.
9. Wu Ruihua. "The Impact of Barrage Interaction in E-commerce Live Broadcasting on Consumer Purchase Decisions—Based on the Moderating Effect of Flow Experience." Commercial Economy Research, 2023, Issue 17, pp. 74-77.
10. Yuan Haixia, Huang Liwen. "Research on the Impact of E-commerce Live Broadcasting Interaction Models on Consumer Purchase Intention." Journal of Harbin University of Commerce (Social Science Edition), 2022, Issue 06, pp. 19-30.
11. Gong Hanyu, Zhao Meng, Ren Jiali, Hao Zhihua. "Live Streaming Strategy under Multi-channel Sales of the Online Retailer." Electronic Commerce Research and Applications, 2022, Issue 55.
12. Lin Zhibing, Li Yuwen, Chen Mofan. "Research on Live Marketing Strategies in Dual-Channel Supply Chains." Systems Science and Mathematics, 2023, Vol. 43, No. 10, pp. 2615-2629.
13. Wang Chenyu, Sun Jingchun, Shi Siyu. "Research on Sales Model Selection and Live Broadcasting Marketing Strategies in E-commerce Platforms." Journal of Industrial Engineering, 2023, Vol. 37, No. 05, pp. 190-199.

joint conference group and negotiation DECISION support systems technology Porto 2024

24th International Conference on Group Decision and Negotiation &
10th International Conference on Decision Support System Technology
Human-Centric Decision and Negotiation Support for Societal Transitions

Strategy-proofness with detectable lies

Gilbert LAFFOND[1], Jean LAINE[1,2,*] and M. Remzi SANVER[3]

[1] Murat Sertel Center for Advanced Economic Studies, Istanbul Bilgi University, Turkey
[2] Lirsa, Conservatoire National des Arts et Métiers, Paris, France
[3] LAMSADE, Université Paris-Dauphine PSL, France
[*] corresponding author: jeanlainelirsa@gmail.com

Abstract

We introduce a new concept of manipulation of a social choice function, based on the possibility for some individuals to detect and report preference misrepresentation. More precisely, we assume that each individual having some preference p (1) can detect whether any other individual lies or not when claiming that p is also her preference, and (2) will report a lie that is not beneficial to herself. A social choice function is called strategy-proof with detectable lies (SPDL) if every potential manipulation is reported. We characterize the class of onto and SPDL social choice functions. We also consider weaker versions of SPDL, for which a preference misrepresentation using p can be detected by individuals whose preference agrees with p only on subsets of alternatives.

Keywords: strategy-proofness; manipulation; information; social choice function

1. Introduction

One of the most well-known results in social choice theory is the Gibbard-Satterthwaite theorem (Gibbard 1973, Satterthwaite 1975) which states that if the preference of each individual can be any linear order over three alternatives or more, every social choice function, which maps every profile of orders to a single alternative, is exposed to manipulation, unless it is a dictatorship or it violates ontoness (that is, there is an alternative that is never chosen, regardless individual preferences). Manipulation means the possibility for an individual to get a better outcome by misrepresenting her preferences over outcomes. Ways to overcome this negative result are investigated in a vast and still growing literature, most of it considering restrictions upon individual preferences (for a survey, see Barberà 2011), or the choice of sets as irresolute outcomes (Taylor 2005). Another stream of papers considers possible weakenings of strategy-proofness while keeping the assumption that any preference is possible. At least three different approaches can be retained along this line of research. First, one may assume that individuals have limited available information about other preferences (Nurmi 1987, Conitzer, Walsh, Xia 2011, Reinjgoud and Endriss 2012, Campbell, Kelly and Qi 2018, Gori 2021). Indeed, manipulation in the Gibbard-Satterthwaite sense implicitly assumes that every individual has a perfect knowledge of all preferences. If this knowledge becomes limited, manipulation may also be limited. For instance, Gori (2021) assumes that individual information about the preferences of the others is limited to the knowledge, for every pair of alternatives, of the number of people preferring the first alternative to the second one. Gori (2021) shows that non-dictatorial and Pareto efficient choice functions become non-manipulable if manipulation is defined in the following conservative way: an individual reports false preferences if, for every combination of preferences of the others consistent with her information, false preferences cannot make her worse off and, for at least one combination, make her better off. A second approach consist of limiting manipulations to those allowing for a gain in satisfaction that is large enough (Campbell and Kelly 2009, Reffgen 2011)[1]. A third approach studies the implications of limiting the options

[1] One difficulty with this approach is that in an ordinal framework, utility ranks can hardly be used to measure gains in satisfaction.

for preference misrepresentation (Sato, 2013, Muto and Sato, 2017). For instance, Sato (2013) assumes that saying that manipulation occur only through preferences adjacent to the sincere one, and shows that this does not overcome the Gibbard-Satterthwaite impossibility result[2].

In this paper, we introduce the notion of strategy-proofness with detectable lies. This notion combines restrictions upon the range of possible preferences for manipulation with the nature of available information about others. We assume that each individual correctly partitions the individual set into the subset of all individuals who share her preference, and the subset of individuals who do not. Moreover, we assume that a detected manipulation will be reported only it harms the well-being of the detector, and that no individual will manipulate if the manipulation is reported. Hence, a social choice function is called strategy-proof with detectable lies (SPDL) if every possible manipulation is reported.

SPDL naturally fits with several collective decision-making frameworks, such as representative democracy, group identification and transferable voting rights:

- *Representative democracy:* In many countries, citizens elect delegates to act on their behalf by voting on laws, policies, and other matters of government. Hence, the group of voters (the Parliament members) is a subgroup of the citizens. Suppose that the society is large enough for every possible preference to be the one of some citizen. If a representative with preference p (interpreted as a political platform) announces another preference q, all citizens with preference q are likely to identify in this move an attempt to manipulate, and they will denounce it unless they gain from keeping silent. As being depicted as a betrayer, or as a politician turning round, puts her future career at risk, the delegate will drop any move that may be denounced. Hence, she will change her opinion only if this change favors not only her own group but also the one she pretends to join.
- *Group identification:* Suppose that the society is partitioned into several groups, each being associated with a specific preference that is publicly known. Moreover, suppose the society is large enough for all possible groups to exist (that is, for all possible preferences to be represented). The fact that all members of a group share the same preference can be regarded as the existence of a common social (or political) identity. Moreover, while the preference of each group is common knowledge, the social identity of each individual is a partially private information: only members of a same group can identify whether newcomers actually share its social identity. Under this interpretation, potential manipulation prevails only by reporting a 'false' social identity. However, by pretending having some social identity, an individual may harm all individuals having this identity. In this case, one anticipates that all current members of the group will denounce the newcomer. Hence, manipulation actually occurs only if taking a false social identity is beneficial not only to an individual, but also to the entire group having this identity.
- *Transferable voting rights:* Another interpretation is based on the existence of vote delegation. Assume the society is divided into two groups. The insiders have the right to vote, while the outsiders do not. However, an insider can choose to transfer her right to an outsider, under the condition that the outsider accepts it. Clearly, delegation is accepted if the outsider gets better o¤ by exerting her new right rather than staying out. Hence, interpreting manipulation as a proposal to transfer rights naturally leads to the concept of safe manipulation, and therefore to SPDL.

One may also relate SPDL to some ethical motivations, by which individuals may be reluctant to 'lie' unless this lie does not contradict the true opinion regarding the truth-telling and the post-lie outcomes. Put differently, individuals may find more justified from a moral viewpoint to misrepresented their opinions involving "irrelevant alternatives"[3].

[2] Restricting 'false' preferences to those adjecent to the 'true' one may be justified by the existence of limited computational skills, or by ethical reasons which favor 'small' lies.

[3] For instance, if truth-telling brings a right-wing winner a, one may argue that a left-wing ideological citizen will be reluctant to behave strategically (and allows the victory of a left-wing candidate b) by pretending she prefers a to b. In contrast, she might be ready to misrepresent her opinion regarding b and another left-wing candidate c who has no chance to get elected.

Clearly, SPDL is a property weaker than strategy-proofness (which holds if no manipulation is possible) and therefore we may expect finding non-dictatorial, onto, and SPDL social choice functions. Our main result (Theorem 1) states that if there are at least three outcomes, and if all preferences are admissible, ontoness combined with SPDL imply either dictatorship or anti-dictatorship. Hence, weakening SP to SPDL does not yield any new appealing solution. As a corollary of Theorem 1, we get that the well-known property of Independence of Irrelevant alternatives (IIA) for social choice functions is actually equivalent to a weaker version of IIA defined as follows: if a is chosen in some situation and if exactly one individual changes her opinions about any pair of outcomes but the pair {a,b}, then b cannot be chosen in the new situation. The main reason for this equivalence is that SPDL turns to be a property equivalent to weak IIA.

We also generalize the notion of detectable lie to various informational contents, where manipulation with some order q can be detected by individuals with a preference similar enough to q. A social choice function F is called k-SPDL if any attempt to manipulate by announcing some preference q is detected and reported by at least one individual whose preference agrees with q about the k-best alternatives. Clearly, F is SPDL if and only if F is (m-1)-SPDL where m is the number of alternatives. For every $k \in \{1,\dots,m\text{-}1\}$, we establish the existence of a social choice function that is k-SPDL and neither a dictatorship nor an antidictatorship. However, such a function necessarily fails a simple monotonicity property, and as such, does not appear as a very appealing collective choice method.

2. Definition of strategy-proofness with detectable lies

We consider a finite set $N = \{1,\dots,n\}$ of individuals confronting a finite set A of m social alternatives, where $n \geq 2$ and $m \geq 3$. The set of linear orders (i.e. reflexive, anti-symmetric, complete and transitive binary relations) over A is denoted by L. Each individual i has preferences over alternatives represented by an element p_i of L. We define the position (or rank) of alternative a in p_i as $pos(a,p_i) = 1+|\{b : (b,a) \in p_i\}|$. A profile is an element $P = (p_i)_{i\in N}$ of L^n. Given a profile P together with $q \in L$ and $i \in N$, (P_{-i},q) stands for the profile obtained from P by changing i's preference from p_i to q. A social choice function (henceforth, SCF) is a function $F: L^n \to A$. Moreover, F is onto if $F(L^n) = A$. Given a profile P together with an SCF F, we write $(i, P\to a, q\to b)$ for a situation where $F(P) = a$ and $F(P_{-i},q) = b \neq a$. Such a situation is called a pivotal situation. We formalize below the key-concept of strategy-proofness with detectable lies.

DEFINITION 1 A voter i safely manipulates F at profile P if there exists a pivotal situation $(i, P\to a, q\to b)$ such that $(b,a) \in p_i \cap q$.

Definition 1 refers to the possibility for some individual to detect and report a preference misrepresentation by another individual. Suppose that some individual i can change the outcome at some profile P from a to b by announcing order q instead of her true order p_i, and suppose that such a change is beneficial, that is $(b,a) \in p_i$. If every individual having q as order can detect that i manipulates, she will report this manipulation unless she benefits herself from it, that is if $(b,a) \in q$. Hence, individual i's manipulation is safe, i.e. immune to reporting, if $(b,a) \in p_i \cap q$. We call strategy-proof with detectable lies (SPDL) an SCF that is safely manipulated at no profile. Equivalently, an SCF F is SPDL if any pivotal situation $(i, P\to a, q\to b)$ is such that either $(a,b) \in p_i$ or $(a,b) \in q$. Clearly, SPDL is a property weaker than strategy-proofness (SP) à la Gibbard and Satterthwaite[4].

2. Main Result

An SCF F is onto if $\forall a \in A$, there exists $P \in L^n$ such that $F(P) = a$: every alternative is the outcome of F at some profile. Moreover, F is a dictatorship (resp. anti-dictatorship) if there exists an individual i^* such that $\forall a \in A, \forall P \in L^n, F(P) = a \Leftrightarrow pos(a,p_{i^*}) = 1$ (resp. $pos(a,p_{i^*}) = m$). It turns out that a very narrow class of non-appealing SCFs simultaneously satisfy ontoness and SPDL.

[4] An SCF is strategy-proof à la Gibbard and Satterthwaite if any pivotal situation $(i, P\to a, q\to b)$ is such that $(a,b) \in p_i$.

THEOREM 1 An SCF is onto and SPDL if and only if it is either a dictatorship or an anti-dictatorship.

The Gibbard-Satterthwaite theorem (Gibbard 1973, Satterthwaite 1975) states that only a dictatorship is onto and SP. Under ontoness, weakening SP to SPDL adds only anti-dictatorship as a possible SCF. Hence, this weakening does not allow to escape from an impossibility result.

An interesting corollary of Theorem 1 relates to the well-known arrowian property of Independence to Irrelevant Alternatives (IIA) for SCFs. Recall that an SCF F satisfies IIA if $\forall a,b \in A$, $\forall P,P' \in L^n$, $[P|_{\{a,b\}} = P'|_{\{a,b\}}] \Rightarrow \neg[F(P) = a \text{ and } F(P') = b]$. With words, if a is chosen by F at some profile P and if P and another profile P′ agree on how to compare a and another alternative b, then F cannot choose b at profile P′. We define weak IIA by restricting this requirement to profiles P′ that agree with P on all orders but one. Formally, F satisfies weak IIA if $\forall a,b \in A$, $\forall P \in L^n$, $\forall q \in L$, $\forall i \in N$, $[p_i|_{\{a,b\}} = q|_{\{a,b\}}] \Rightarrow \neg[F(P)=a \text{ and } F(P_{-i}, q) = b]$: If exactly one individual changes her preference without reversing her opinion about the current outcome a and another alternative b, then b cannot become the new outcome. We show that SPDL and weak IIA are actually equivalent properties. Moreover, it is already known (Yu, 2015, claim 2) that an SCF satisfies ontoness and IIA if and only if it is either a dictatorship or an anti-dictatorship. Combining this result with Theorem 1 shows

COROLLARY 1 IIA and weak IIA are equivalent properties.

3. Generalizing strategy-proofness with detectable lies

Pick any pivotal situation (i, P→a, q→b) where $(b,a) \in p_i$. One may assume that individuals may detect i′s manipulation if their preference is 'close enough' to q. Alternative meanings can be given to closeness. We consider a specific one, parametrized by $k \in \{1,\dots,m-1\}$, for which two orders are k-close if they agree on the first k-best choices. Hence, given k, every individual who agrees with the k-best alternatives in q can detect a manipulation using q, and will report it unless it is beneficial to her. This motivates the following definition of safe k-manipulation.

DEFINITION 2 Given $k \in \{1,\dots,m-1\}$, a voter i safely k-manipulates F at profile P if there exists a pivotal situation (i, P→a, q→b) such that $(b,a) \in p_i \cap q$ and $pos(b,q) \leq k$.

To fully understand definition 2 , pick a pivotal situation (i, P→a, q→b) such that $(b,a) \in p_i \cap q$ and consider the two cases below:

Case 1: $pos(b,q) > k$

If $pos(a,q) \leq k$, every voter with a preference q′ that agree with q on the first k alternatives has an incentive to report the lie. If $pos(a,q) > k$, at least one voter with a preference q′ that agree with q on the first k alternatives has an incentive to report the lie.

Case 2: $pos(b,q) \leq k$

If $(a,b) \in q$, every voter with a preference q′ that agree with q on the first k alternatives has an incentive to report the lie. In contrast, no such voter will report the liar if $(b,a) \in q$. This last situation is the only one where manipulation is safe, hence the definition of safe k-manipulation.

DEFINITION 3 Take any $k \in \{1,\dots,m-1\}$. An SCF F is k-strategy-proof with detectable lies (k-SPDL) if there is no profile at which some voter can safely k-manipulate F.

Equivalently, F is k-SPDL iff for any pivotal situation (i, P→a, q→b), either pos(b,q) > k or [pos(b,q) ≤ k and (a,b) ∈ q].

Observe that SPDL and (m-1)-SPDL are actually equivalent. Moreover, it is rather straightforward to show that ∀k,k′ ∈{1,…,m-1} with k > k′, k-SPDL implies k′-SPDL. Intuitively, more individuals can detect a manipulation if k decreases, and therefore the number of SCFs that are k-SPDL potentially increases. This is actually true, even if one discard autocracy, as stated in the next proposition. Define as autocratic an SCF for which all pivotal situations involve the same individual.

PROPOSITION 1 For each k ∈ {1,…,m-2}, there exists an SCF that is onto, k-SPDL, not (k+1)-SPDL and not autocratic.

However, there is no hope to find appealing SCFs that are k-SPDL even for k=1. At least if appealing means satisfying simple monotonicity, which is defined as follows. Given a profile P together with an alternative a, say that profile P′ is a lift for a in P if ∀i ∈ N, we have (1) ∀c ≠ a, (a,c) ∈ p_i ⇒ (a,c) ∈ p'_i, and (2) ∀c ≠ a, ∀d ≠ c, (c,d) ∈ p_i ⇔ (c,d) ∈ p'_i. An SCF F is simply monotonic if for any profile P and any alternative a, F(P) = a implies F(P′) = a where P′ is a lift for a w.r.t P.[5]

PROPOSITION 2 An SCF is onto, simply monotonic and 1-SPDL if and only if it is a dictatorship.

4. Further comments

Our main contribution is to show that considering strategy-proofness with detectable lies does not allow to escape from the conclusions of the Gibbard-Satterthwaite theorem. Considering detectable lies is a rather natural weakening of strategy-proofness in the mathematical sense. Indeed, an SCF is strategy-proof if (a,b) ∈ p_i and (b,a) ∈ q at every pivotal situation (i, P→a, q→b), while it is strategy-proof with detectable lies if (a,b) ∈ p_i or (a,b) ∈ q. Another weakening of strategy-proofness, called one-way monotonicity, is introduced in Sanver and Zwicker (2009). An SCF is one-way monotonic if (a,b) ∈ p_i or (b,a) ∈ q at every pivotal situation (i, P→a, q→b).[6]

Even if they are logically independent, it is worth noting that SPDL and one-way monotonicity yield very different results. Indeed, one-way monotonicity is satisfied by several well-known SCFs while being violated by others, whereas SPDL does not discriminate among appealing SCFs. However, one-way monotonicity is not easy to motivate, while SPDL naturally relates to the type of knowledge individuals have about others and as such completes previous studies on manipulation with limited information.

Several questions may be addressed in further research. First, the class of onto and k-SPDL social choice functions remains to be characterized. One may also search for restrictions upon individual preferences that allow for the existence of appealing SCFs that are onto and SPDL. Finally, considering the notion of detectable lies in the case where social outcomes are sets of alternatives rather than single alternatives remains to be done.

References

S. Barberà, Strategy-proof social choice. In: Handbook of Social Choice and Welfare, Volume 2 (eds. Arrow, K.J., Sen, A., and K. Suzumura), Netherlands North-Holland: 731-831, 2011.

D.E. Campbell, J.S. Kelly J.S., Gain from manipulating social choice rules. Economic Theory, 40:349-371, 2009.

[5] Most well-known and widely used SCFs, such as scoring rules or functions based on majority tournament solutions, are simply monotonic.

[6] One-way monotonicity holds whenever a pivotal situation (i,P→a,q→b) is beneficial to individual i, the "mirror" pivotal situation (i, (P_{-i},q)→b, p_i→a) is not beneficial to i.

D.E. Campbell, J.S. Kelly, S. Qi, A stability property in social choice theory. International Journal of Economic Theory, 14:85-95, 2018.

V. Conitzer, T. Walsh, L. Xia, Dominating manipulations in voting with partial information. In: Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, 2011.

A. Gibbard, Manipulation of voting schemes: a general result. Econometrica, 41:587-601, 1973.

M. Gori, Manipulation of social choice functions under incomplete information, Games and Economic Behavior 129:350-369, 2021.

N. Muto, S. Sato, An impossibility under bounded response of social choice functions. Games and Economic Behavior, 106:1-15, 2017.

H. Nurmi, Comparing Voting Systems. D. Reidel Publishing Company: Dordrecht, 1987.

A. Reffgen, Generalizing the Gibbard-Satterthwaite theorem: partial preferences, the degree of manipulation, and multi-valuedness. Social Choice and Welfare, 37:39-59, 2011.

A. Reijngoud, U. Endriss, Voter response to iterated poll information. In: Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems, 2012.

M.R. Sanver, W.S. Zwicker, One-way monotonicity as a form of strategy-proofness. International Journal of Game Theory, 38:553-574, 2009.

S. Sato, A sufficient condition for the equivalence of strategy-proofness and nonmanipulability by preferences adjacent to the sincere one. Journal of Economic Theory, 148:259-278, 2013.

M.A. Satterthwaite, Strategy-proofness and Arrow's conditions: existence and correspondence theorems for voting procedures and social welfare functions. Journal of Economic Theory, 10:187-217, 1975.

A.D. Taylor, Social choice and the mathematics of manipulation, Cambridge University Press, 2005.

N. Yu, A quest for fundamental theorems of social choice. Social Choice and Welfare, 44:3 533-548, 2015

Combining the Borda count with approval and disapproval voting

José Luis García-Lapresta [1], Miguel Martínez-Panero [1*]

[1] IMUVA, PRESAD Research Group, Departamento de Economía Aplicada, Universidad de Valladolid, Valladolid, Spain
[*] miguel.mpanero@uva.es

Abstract

In this contribution, we extend preference-approvals to a more general situation, where voters can sort the alternatives in three disjoint classes instead of two (for instance, acceptable, neutral and unacceptable). We propose a parameterized family of voting systems related to the Borda count, where positive (negative) decreasing individual scores are assigned to acceptable (unacceptable) alternatives, and neutral alternatives obtain null scores.

Keywords: voting systems; Borda count; approval voting; preference-approvals.

1. Introduction

Voting systems aggregate individual opinions on a set of alternatives to generate an outcome, usually a single winning alternative, a subset of winning alternatives or a weak order on the set of alternatives (some surveys on voting systems can be found in Nurmi, 1983; and Brams & Fishburn, 2002).

Individual opinions can be of different nature: the best alternative, as in plurality rule; the worst alternative, as in antiplurality rule; the best and worst alternatives, as in best-worst voting systems (see García-Lapresta et al., 2010); a subset of acceptable alternatives, as in approval voting (see Brams & Fishburn, 1978); a weak or linear order on the set of alternatives, as in the Borda rule; an evaluation of each alternative through an ordinal scale, as in Majority Judgment (see Balinski & Laraki, 2007, 2011); etc.

Preference-approval structures combine preferences over the alternatives, through a weak order, and which alternatives are acceptable (see Brams, 2008, Chapter 3; Brams and Sanver, 2009; Sanver, 2011; and Erdamar et al. 2014). This kind of structures increase the expressivity of voters with respect to approvals and weak orders (see Albano et al., 2023, Table 1).

A higher expressivity is possible when voters provide a weak order on the set of alternatives and, additionally, classify them in three classes, arranged from the best to the worst (for instance, acceptable, neutral and unacceptable). On this, see Felsenthal (1989), Yilmaz (1999) and Alcantud & Laruelle (2014), among others.

In this contribution, we assume that, from a behavioral point of view, voters first sort the alternatives in the mentioned three classes and, secondly, they rank order the alternatives in each class. If voters only focus on the first and third classes (say acceptable and unacceptable subsets of alternatives, respectively), they do not need to rank order the alternatives of the second class and they are indifferent between them. This threefold approach is an extension of the inputs managed by the best-worst voting systems, where voters only show their best and worst alternatives (see García-Lapresta et al., 2010).

In this scenario, we propose a parameterized family of voting systems where, considering the individual preferences, decreasing positive scores are assigned to the alternatives in the first class; a null score is assigned to each alternative in the second class; and decreasing negative scores are assigned to the alternatives in the third class. These scores are inspired by Black's second Borda count in the setting of weak orders (see Black (1976)), yet adapted to our new scenario. It would be interesting to find the values of the parameters ensuring good properties to the proposed voting systems.

Although the Borda count and the approval voting procedures deal with different informational bases from the agents when considering the alternatives (pairwise comparisons and acceptance assessments one by one, respectively), García-Lapresta & Martínez-Panero (2002) outlined a fuzzy approach between these distinct points of view. On the other hand, it is interesting to note that, more recently, Barokas & Sprumont (2022) have proposed what they called a "broken Borda rule" in a preference-approval context, appearing here as a particular case in our more general setting (see Section 3 for more details).

In this contribution, we provide a compelling and comprehensive way to deal with scorings and approval/disapproval components of individual preferences (also allowing the voters to stay neutral in front of some alternatives). We also analyze the suitability of the parameters' values appearing in our family of voting systems.

2. The procedure

Consider a set of voters $V = \{1, 2, \dots, m\}$, with $m \geq 2$, that provide their opinions on a set of alternatives $X = \{x_1, x_2, \dots, x_n\}$, with $n \geq 2$, by sorting them in a partition of *X* composed of three classes *A*, *N* and *U*. Since this classification depends on each voter, A_v, N_v and U_v denote the three classes of voter $v \in V$. Since they are a partition of *X*, we have $A_v \cup N_v \cup U_v = X$ and $A_v \cap N_v = A_v \cap U_v = N_v \cap U_v = \emptyset$.

We also assume that each voter $v \in V$ ranks order the alternatives by means of a weak order on *X*, R_v. With P_v and I_v we denote the asymmetric and symmetric parts of R_v, respectively.

The following consistency conditions are required: if either $x_i \in A_v$ and $x_j \in (N_v \cup U_v)$, or $x_i \in N_v$ and $x_j \in U_v$, then $x_i \, P_v \, x_j$. This means that all voters prefer the alternatives of each class to those in the subsequent class. We also assume that all the neutral alternatives are indifferent to each other, i.e., if $x_i, x_j \in N_v$, then $x_i \, I_v \, x_j$.

With $\mathcal{T}(X)$ we denote the set of 4-tuples (R, A, N, U), where *R* is a weak order on *X* and $A, N, U \subseteq X$ satisfy the conditions mentioned above. A profile is a vector of $\mathcal{T}(X)^m$.

The following individual scores are assigned to each alternative $x_i \in X$, where $\alpha \geq 0$ and $\beta, \varepsilon > 0$, for each voter $v \in V$:

$$B_v(x_i) = \#\{x_j \in A_v \mid x_i \, P_v \, x_j\} + \tfrac{1}{2} \cdot \#\{x_j \in A_v \setminus \{x_i\} \mid x_j \, I_v \, x_i\} + \alpha \cdot \#N_v + \beta \cdot \#U_v + \varepsilon, \text{ if } x_i \in A_v,$$

$$B_v(x_i) = 0, \text{ if } x_i \in N_v,$$

$$B_v(x_i) = -\#\{x_j \in U_v \mid x_j \, P_v \, x_i\} - \tfrac{1}{2} \cdot \#\{x_j \in U_v \setminus \{x_i\} \mid x_j \, I_v \, x_i\} - \alpha \cdot \#N_v - \beta \cdot \#A_v - \varepsilon, \text{ if } x_i \in U_v.$$

Note that $B_v(x_i)$ coincides with the score given by the Borda count in the setting of weak orders in the case of $x_i \in A_v$ whenever $\alpha = \beta = 1$ and $\varepsilon = 0$. With the same values of the parameters $\alpha, \beta, \varepsilon$, $B_v(x_i)$ coincides with the opposite of the score given by the Borda count in the setting of weak orders in the case of $x_i \in U_v$.

Note that $x_i \in A_v \Leftrightarrow B_v(x_i) > 0$, $x_i \in N_v \Leftrightarrow B_v(x_i) = 0$, and $x_i \in U_v \Leftrightarrow B_v(x_i) < 0$. This is ensured by the fact that $\varepsilon > 0$. Otherwise, it would possible $B_v(x_i) = 0$ when:

- $x_i \in A_v$, if x_i is the only worst alternative of A_v and $N_v = U_v = \emptyset$; or just $U_v = \emptyset$ and $\alpha = 0$.
- $x_i \in U_v$, if x_i is the only best alternative of U_v and $A_v = N_v = \emptyset$; or just $A_v = \emptyset$ and $\alpha = 0$.

As an example, consider $(R_v, A_v, N_v, U_v) \in \mathcal{T}(\{x_1, x_2, \dots, x_9\})$ represented by

$$\begin{array}{c} x_2 \\ x_1 \quad x_5 \\ \hline x_3 \quad x_6 \\ \hline x_8 \\ x_4 \quad x_7 \quad x_9 \end{array}$$

This means that $A_v = \{x_1, x_2, x_5\}$, $N_v = \{x_3, x_6\}$ and $U_v = \{x_4, x_7, x_8, x_9\}$.

Then, the obtained scores arranged in a decreasing way are

$$B_v(x_2) = 2 + 2\alpha + 4\beta + \varepsilon$$
$$B_v(x_1) = B_v(x_5) = 0.5 + 2\alpha + 4\beta + \varepsilon$$
$$B_v(x_3) = B_v(x_6) = 0$$
$$B_v(x_8) = -2\alpha - 3\beta - \varepsilon$$
$$B_v(x_4) = B_v(x_7) = B_v(x_9) = -2 - 2\alpha - 3\beta - \varepsilon.$$

A jump can be observed between the lowest positive and highest negative scores:

$$0.5 + 2\alpha + 4\beta + \varepsilon - (-2\alpha - 3\beta - \varepsilon) = 0.5 + 4\alpha + 7\beta + 2\varepsilon.$$

The following result quantifies the extent of the aforementioned gap in the general case.

Proposition: If both A_v, $U_v \neq \emptyset$ for voter v, then the gap between v's scores of the worst acceptable alternative(s) and the best unacceptable alternative(s) is

$$g_v = \frac{t}{2} + (2\alpha - \beta) \cdot \#N_v + \beta n + 2\varepsilon\,,$$

where $t = \#\{x_i \in A_v \mid \forall\, x_j \in A_v \;\; x_j\, R_v\, x_i\} + \#\{x_i \in U_v \mid \forall x_j \in U_v \;\; x_i\, R_v\, x_j\} - 2.$

(bottom acceptable alternatives) (top unacceptable alternatives)

In particular, if such sets with minimum positive and maximum negative scores are singletons, then

$$g_v = (2\alpha - \beta) \cdot \#N_v + \beta n + 2\varepsilon.$$

Additionally, if $\alpha = \beta$, then $g_v = \alpha \cdot (\#N_v + n) + 2\varepsilon$.

The global score of the alternative $x_i \in X$ is defined as the sum of individual scores:

$$B(x_i) = B_1(x_i) + B_2(x_i) + \cdots + B_m(x_i).$$

Then, the alternatives of X are ordered according to their global scores:

$$x_i \geq x_j \Leftrightarrow B(x_i) \geq B(x_j).$$

We note that the role of the parameters $\alpha, \beta, \varepsilon$ is crucial.

3. Discussion

Our aim is to determine suitable values of the parameters $\alpha, \beta, \varepsilon$ so that our proposal of the Borda count combining approval and disapproval voting will be endowed with good features, as the Borda rule does.

In this way, an interesting property shared by both procedures is the invulnerability to the inversion paradox:

if a candidate is the unique winner, and all the voters invert their preferences, then such candidate must not be elected (see Saari & Barney (2003) and Felsenthal (2012)). This is a consequence of the reversal symmetry property satisfied by the classic Borda rule (in fact, it is trivial taking into account that the second Borda count considered by Black (1976) is equivalent to the usual one). This is also true for our rule, due to the definition or the threefold Borda count: for all possible allowed values of $\alpha, \beta, \varepsilon,$ the individual scores obtained by the alternatives in $(R_v, A_v, N_v, U_v) \in \mathcal{T}(X)$ are just the opposite of those in $(R_v, A_v, N_v, U_v)^{-1} = (R_v^{-1}, U_v, N_v, A_v) \in \mathcal{T}(X)$, where both the order and the approval/disapproval are inverted. Hence, the symmetric design of the classic Borda count is preserved in this extended setting, but suitable values of some parameters should still be adjusted to obtain appropriate features.

We note that when $N_v = \emptyset$ and $\beta = 1$, the value $\varepsilon = (mn - m - n + 1)/2$ (obtained by Barokas & Sprumont (2022)) provides a good compromise. Significatively, our approach also converges to the same value in this particular case.

In order to adjust the parameters α and β in the process leading to determine the individual scores, one can ask the following questions.

1. What would happen if an alternative x_i declared as neutral by a voter becomes acceptable? It seems reasonable to think that if, in the new situation, the moving alternative x_i:

    a) Is below a prefixed acceptable alternative x_j

    b) Is indifferent to a prefixed acceptable alternative x_j

    then the score of x_j should increase in both cases, because new pairwise comparisons have appeared between non-neutral alternatives, and the feeling might be more intense.

    c) The same also applies to advocate that the score of any unacceptable alternative should increase.

2. Similarly, if an alternative x_i declared as neutral by a voter becomes unacceptable, it seems compelling to argue that if, in the new situation, the moving alternative x_i:

    a) Is above a prefixed unacceptable alternative x_j

    b) Is indifferent to a prefixed unacceptable alternative x_j

    then the score of x_j should decrease in both cases, by the same aforementioned reasons.

    c) The same also applies to maintain that the score of any acceptable alternative should decrease.

    After some computations, we obtain:

    - 1.a and 2.a entail $\alpha < 1$.
    - 1.b and 2.b entail $\alpha < 0.5$.
    - 1.c and 2.c entail $\beta < \alpha$.

3. On the other hand, let's think about an alternative x_i declared acceptable for a voter which becomes unacceptable. If x_i

    a) Was originally below a prefixed acceptable alternative x_j

    b) Was formerly indifferent to a prefixed acceptable alternative x_j

    c) Becomes above a prefixed unacceptable alternative x_j

    d) Becomes indifferent to a prefixed unacceptable alternative x_j

    then the score of x_j should decrease in all the cases, because the new comparisons between x_j and

x_i might become less significative due to the loss of importance of the last one.

4. Conversely, suppose that an alternative x_i declared as unacceptable for a voter becomes acceptable. If in the new situation x_i

    a) Becomes below a prefixed acceptable alternative x_j

    b) Becomes indifferent to a prefixed acceptable alternative x_j

    c) Stayed originally above a prefixed unacceptable alternative x_j

    d) Was formerly indifferent to a prefixed acceptable alternative x_j

    then the score of x_j should increase in all the cases, because the new pairwise comparison between x_j and x_i would become more significative due to the rise of the last one.

    After some computations, we obtain:

    - 3.a, 3.c, 4.a and 4.c entail $\beta < 1$.
    - 3.b, 3.d, 4.b and 4.d entail $\beta < 0.5$.

5. Another possible movement is that an alternative x_i declared as acceptable for a voter becomes neutral. If in the former situation x_i

    a) Was below a prefixed acceptable alternative x_j

    b) Was indifferent to a prefixed acceptable alternative x_j

    then then the score of x_j should decrease after the movement (as in 3).

    c) The same also applies to advocate that the score of any unacceptable alternative should decrease.

6. Finally, an unacceptable alternative x_i for a voter becomes neutral. If in the original situation x_i

    a) Was above a prefixed unacceptable alternative x_j

    b) Was indifferent to a prefixed unacceptable x_j

    then the score of x_j should increase after the movement in all the cases, due to the better new status (as in 4.).

    c) The same also applies to advocate that the score of any acceptable alternative should increase.

After some computations, we obtain:

- 5.a and 6.a entail $\alpha < 1$.
- 5.b and 6.b entail $\alpha < 0.5$.
- 5.c and 6.c entail $\beta < \alpha$.

All in all, our analysis leads to the following restrictions of the parameters: $\alpha, \beta < 0.5$ and $\beta < \alpha$. Nonetheless, further bounds to these values might still be necessary in order to obtain good social choice features.

It would be also interesting to analyze the compatibility between preference and approval/disapproval components of the inputs in connection with the social outcome, answering questions such: Should an alternative disapproved by a majority of voters be elected? Which values of $\alpha, \beta, \varepsilon$ would prevent this situation, if any? What does it mean collective approval or disapproval? Should prevail collective approval against merely highest scores?

References

Albano, A., García-Lapresta, J. L., Plaia, A., & Sciandra, M. (2023). A family of distances for preference-approvals. *Annals of Operations Research*, 323, 1-29.

Alcantud, J. C. R., & Laruelle, A. (2014). Dis&approval voting: a characterization. *Social Choice and Welfare*, 43, 1-10.

Balinski, M., & Laraki, R. (2007). A theory of measuring, electing and ranking. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 8720-8725.

Balinski, M., & Laraki, R. (2011). *Majority Judgment: Measuring, Ranking, and Electing*. The MIT Press.

Barokas, G., & Sprumont, Y. (2022). The broken Borda rule and other refinements of approval ranking. *Social Choice and Welfare*, 58, 187-199.

Black, D. (1976): Partial justification of the Borda count. *Public Choice*, 28, 1-16.

Brams, S. J. (2008). *Mathematics and Democracy: Designing Better Voting and Fair-Division Procedures*, Princeton University Press.

Brams, S. J., & Fishburn, P. C. (1978). Approval voting. *American Political Science Review*, 72, 831-857.

Brams, S. J., & Fishburn, P. C. (2002). Voting procedures. In K. J Arrow, A. K. Sen, & K. Suzumura (Eds.), *Handbook of Social Choice and Welfare* (Vol. 1, pp. 173-236), Elsevier.

Brams, S. J., & Sanver, M. R. (2009). Voting systems that combine approval and preference. In S. J. Brams, W. V. Gherlein, & F. S Roberts (Eds.), *The Mathematics of Preference, Choice and Order: Essays in Honour of Peter C. Fishburn*, Studies in Choice and Welfare (pp. 215-237), Springer-Verlag.

Erdamar, B., García-Lapresta, J. L., Pérez-Román, D., & Sanver, M. R. (2014). Measuring consensus in a preference-approval context. *Information Fusion*, 17, 14-21.

Felsenthal, D. S. (1989). On combining approval with disapproval voting. *Behavioral Science*, 34, 56-70.

Felsenthal, D. S. (2012): Review of paradoxes afflicting procedures for electing a single candidate. In D. S. Felsenthal, & M. Machover (Eds.), *Electoral Systems: Paradoxes, Assumptions, and Procedures* (pp. 19-91), Springer.

García-Lapresta, J. L., Marley, A. A. J., & Martínez-Panero, M. (2010). Characterizing best-worst voting systems in the scoring context. *Social Choice and Welfare*, 34, 487-496.

García-Lapresta, J. L., & Martínez-Panero, M. (2002). Borda count *versus* approval voting: A fuzzy approach. *Public Choice*, 112, 167-184.

Nurmi, H. (1983). Voting procedures: A summary analysis. *British Journal of Political Science*, 13, 181-208.

Saari, D. G. & Barney, S. (2003): Consequences of reversing preferences. *The Mathematical Intelligencer*, 25, 17-31.

Sanver, M. R. (2011). Approval as an intrinsic part of preference. In J. F. Laslier, & M. R. Sanver (Eds.), *Handbook on Approval Voting* (pp. 469-481), Springer-Verlag.

Yilmaz, M. R. (1999). Can we improve upon approval voting? *European Journal of Political Economy*, 15, 89-100.

Analyzing the perceived importance of a multicriteria decision support method in a group decision-making process

Marcelo Rodrigo da Silva [1*] and Alexandre Bevilacqua Leoneti [1**]

[1] Research Group in Decision Sciences, School of Economics, Business Administration and Accounting at Ribeirão Preto, University of São Paulo, Brazil
[*] marcelo.rodrigo@usp.br ** ableoneti@usp.br

Abstract

An experiment was used for analyzing the perceived importance of a decision support method in negotiation and group decision-making. The qualitative data gathered from the experiment with 100 students were processed by a software of textual data analysis. The results shown that the use of the method was perceived as a support tool to the negotiations' skills and as a descriptive or a perspective tool either. Additionally, students demonstrated a high satisfaction level with the approach, which demonstrated the level of importance given to the method.

Keywords: group decision; decision making process; negotiation; perceived importance; textual data analysis

1. Introduction

Decision-making is considered the core of management process and the ability to make an effective decision is one of the main qualities of organizational managers (de Almeida, 2013; Hammond et al., 1998; Sofo et al., 2013). However, managers should be familiar with the social environment they work and understand how decision-making process operates (Simon, 1987), since the success and effectiveness of strategic decisions are also related to the use of a structured process to support managerial decision-making (de Almeida, 2013).

Especially at the strategic level, decision problems have multiple conflicting objectives that need to be addressed in an interrelated way (de Almeida, 2013; de Almeida et al., 2015). As several criteria must be considered to evaluate a set of alternatives, this type of problem is categorized as a multicriteria decision problem. Multicriteria decision analysis is indispensable when it is not possible to represent all objectives of a problem using a single measurement scale (de Almeida et al., 2015).

The application of decision support tools allows the decision-maker to think about the problem in a structured way and, consequently, improve the quality of the outcomes that result from a decision and avoid questions regarding the validity of the analysis, as well as bringing consistency to the decision-making process (Baker et al., 2001; Clemen & Reilly, 2013; Ragsdale, 2019; Render et al., 2018).

Structuring a model can be employed to support the decision-making process because it provides insights, helps to understand, and analyze the problem investigated, assists the manager in making choices and is the scientific approach that has been applied by operational research (Clemen & Reilly, 2013; Ragsdale, 2019; Render et al., 2018). Eventually, an alternative strategy is to apply social choice theory, as voting procedures have demonstrated to be efficient and are usually implemented to support multicriteria group decision-making in organizations, usually preceded by a negotiation process between the decision-makers (de Almeida et al., 2019).

Combining multicriteria decision support tools and voting procedures is also a recommended approach to support group decisions, as this methodology can increase both the group members' acceptability of the decision and their commitment to implement the chosen alternative (Leoneti & de Sessa, 2017; Leoneti, 2016; Ziotti, 2018). Ziotti & Leoneti (2020) also verified that sense of justice and satisfaction towards group decision problems are higher when the agreement is supported by mathematical methods. Besides, Ziotti & Leoneti (2020) validated that commitment to agreements in a group decision is higher when the process is supported by such methods. Another intrinsic aspect that should be considered when evaluating a decision support tool is the users' perception about the method's usefulness, since this metric is associated with the degree to which users believe that this methodology can improve their problem-solving performance and with user's acceptance (Porat; Delaney & Kostopoulou, 2017; Shibl, Lawley & Debuse, 2013).

Game theory is also a suitable tool for mathematically modelling strategic interactions. In these situations, decisions made by individuals influence each other in a way that the results obtained by one decision-maker depend not only on their choices, but also on the strategies taken by others (Fiani, 2009). For example, as an alternative to the existing decision support methods in the literature, Leoneti (2016) proposed a utility function to model multicriteria group decision problems as games, whose main advantage is the possibility of evaluating conflicts between decision-makers from a strategic approach.

Therefore, decision-making in organizations has become an increasingly complex task for managers, elucidating the need to build systems and/or procedures that help choose the most appropriate multi-criteria analysis methodology for the problem evaluated (Ferreira et al., 2018).

In this context, this paper aimed to analyse the application of a multicriteria group decision support method based on game theory, proposed by Leoneti (2016), according to methodology proposed by Leoneti et al. (2015), with new cases created by Ziotti & Leoneti (2020). The sequential group decision was evaluated in a negotiation experiment where a group of five participants was asked to make a group decision in three different cases. The results of the experiment's overall evaluation and the importance of the multicriteria method were collected to investigate the group's degree of acceptance and satisfaction using a software of textual data analysis.

2. Method

2.1. Group decision-making experiment and materials

An experiment in negotiation and group decision-making was chosen as the environment for data collection. The experiment took place in a meeting room of the School of Economics, Business Administration and Accounting at Ribeirão Preto (FEARP). After getting into the room, participants were arranged on the right side of a rectangular table and instructed that they could sit in any position they wanted. To standardize data collection, researchers identified the participants from decision-maker 1 to decision-maker 5, as illustrated in Figure 1. The arrangement of the participants at the table formed a semicircle to facilitate eye contact, interaction, and communication.

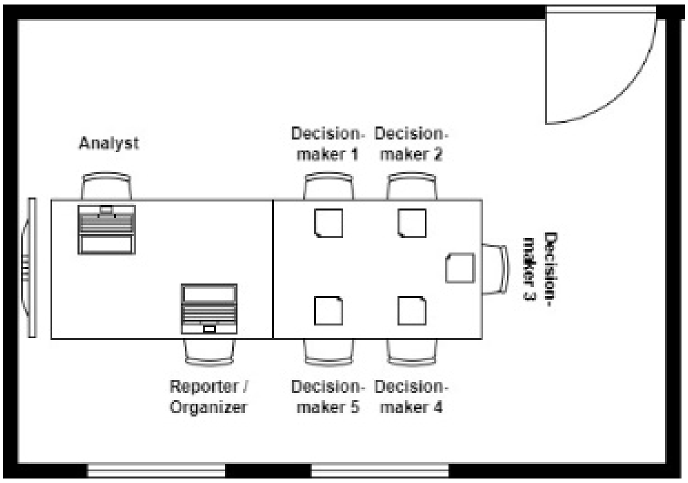


Figure 1: Layout of the meeting room and arrangement of participants

The researchers were arranged on the left side of the table, as also shown in Figure 1, to enable them to observe the negotiation and interact with the participants to answer questions and present the results. The researchers are illustrated in Figure 1 by the players of reporter/organizer and analyst. This layout was applied by Ziotti (2018) and was shown to be effective for conducting group decisions and negotiations. In general, the participants did not know each other previously, as they could not take part in the experiment more than once and they were not classmates. For this reason, participants received a sheet to fill in with their name. The participants were instructed to keep this sheet visible on the negotiation table, using a pencil hold-er, to facilitate communication between the parties during the experiment.

2.2. Study design, participants, and questionnaires

The negotiation experiment was planned based on the methodology presented by Leoneti et al. (2017) and Leoneti & de Sessa (2015), with two new cases created by Ziotti & Leoneti (2020). Ziotti & Leoneti (2020)'s purpose in structuring these two new cases was to provide a sequential negotiation experiment with decision environments that vary in level of complexity and time horizon, since it was considered a short-term decision (choosing a travel destination), a medium-term decision (choosing an English school) and a long-term one (choosing a CEO for a company).

The study's target population was made up of students from the Ribeirão Preto Campus of the University of São Paulo. The inclusion criterion was to be a regularly enrolled student at this campus. Students with any academic background and enrolled in any semester could participate. Participants were recruited using an online form, which was sent to undergraduate and graduate departments.

Students who interested in participating were arranged in groups of five according to their availability. They were arranged in groups of five because this methodology was presented by Leoneti & de Sessa (2017) and it was already applied on other studies (Ziotti & Leoneti, 2020; da Silva et al., 2020), demonstrating to be suitable to evaluate decision makers' behaviour in conflicting situations. Each application lasted around two hours and was performed at mutually agreed schedules between the participants and the researchers. There were 20 applications of the experiment, amounting to 100 participants.

The stages taken by the participants throughout the experiment are detailed in Table 1. After agreeing in participating, a contextualization presentation about the method, instructions and rules that had to be followed was presented. Then, students provided some personal information and, next, the simulation of each case started. Each case was solved upon in a sequence using three phases: (i) individual phase; (ii) group phase; and (iii) final phase.

Table 1. Experiment Flowchart

Experiment stages	Description
Informed Consent Form	Participants read the Informed Consent Form and agreed to participate in this research.
Application Presentation	This presentation was performed to contextualize the study's applicability and its goals. It was also presented some general information about multicriteria decision method, instructions, and rules of the experiment. Instructions: (i) Participants should interpret each case as realistic as possible; (ii) Participants should follow three different phases to solve each case; (iii) Participants would have to decide after analyzing a set of criteria and alternatives, which would be presented in each case through a decision matrix. Rules: (i) Coalitions were not allowed during the individual phase; (ii) During the individual phase, the participant could not share their choices and preferences; (iii) Participants could not veto during the negotiation; (iv) In the end of the negotiation, the group had to present an agreement; (v) After finishing the negotiation, participants would have to vote secretly to confirm or not the group decision.
Personal Identification	Participants filled out the Identification Form to provide personal data such as name, age, e-mail, undergraduate course, and semester. This Identification Form also introduced a summary of the topics discussed previously.
Individual Phase	Throughout this phase, participants read and analyzed both the case contextualization and decision matrix to rank criteria and alternatives according to their preferences. The decision matrix that was used in each case is presented in Table 2. Using a seven-level Likert scale, the participant was also asked about how much they believed their preferred alternative would be chosen by the group.
Group Phase	During this phase, participant had to persuade others to convince them that his most preferred alternative should be chosen as the solution. Participants had a period of 15 minutes to negotiate. In the end, the group had to present an agreement.
Final Phase	Individual voting happened during the final phase. This voting was secretly and if a breach of contract occurred, the alternative suggested by the method proposed by Leoneti (2016) would be selected as the final decision. Otherwise, the group decision would be maintained. Participants also filled in the evaluation form for the respective case, where they were asked about their satisfaction with the group's final decision, whether they considered this decision fair and what influence each participant had on their individual choice throughout the negotiation. All the answers were measured using a seven-level Likert scale. There was also an open question where the decision-makers were asked to assess the importance of the mathematical method in aiding the group's decision.

The simulation of cases proposed to this group decision-making experiment was performed using the following sequence: (1st) Choosing a Travel Destination (adapted from Leoneti & de Sessa (2017) by Ziotti & Leoneti (2020)); (2nd) Choosing an English School; and (3rd) Choosing a CEO for a Company (Ziotti & Leoneti (2020) created both from the structure of Leoneti & de Sessa (2017)). For each case, a brief contextualization was presented which involved the knowledge that the participant was going to engage in group decision-making which would be represented by a decision matrix made up of five alternatives. These alternatives were evaluated using eight benefit or cost criteria, measured using discrete or continuous scales, which were chosen based on the theoretical assumption of independence, as shown in Table 2. All decision matrices used in this research were previously structured and tested by Ziotti & Leoneti (2020).

Table 2. Decision Matrix for each case

Case 1 - Choosing a Travel Destination								
Alternatives	Criteria							
	Hotel Evaluation	Travelling Duration (hours)	Number of Nights (days)	Exchange Rate	Shopping	Cultural Attractions	Nature	Infrastructure
Destination A	2,5	8	4	R$ 0,9	5	3	7	8
Destination B	3,5	2,5	6	R$ 3,1	9	7	3	6
Destination C	3	4	7	R$ 4,7	4	5	9	7,5
Destination D	5	13	5	R$ 3,3	3	9	6	7
Destination E	4	16	8	R$ 1,1	6	8	5	5
Case 2 - Choosing an English School								
Alternatives	Criteria							
	Distance (Km)	Courseware	Class Size	Class Hours	Infrastructure	School Reputation	Extra Activities	Course Quality
School A	14	R$ 450	15	2	10	6	4	9
School B	7	R$ 650	12	3	9	8	5	8
School C	16	R$ 590	4	2,5	8	7	9	7
School D	6,5	R$ 570	8	6	5	8	6	8
School E	10	R$ 300	18	4	7	9	8	5
Case 3 - Choosing a CEO for a Company								
Alternatives	Criteria							
	Professional Qualification	Years of Company	Years in the Area	Leadership Positions	Ethic	Adaptation to Change	Commitment	Professional Influence
Candidate A	8	7	7	3	10	4	6	9
Candidate B	7	4	10	3	9	9	4	8
Candidate C	6	6	15	4	8	7	7	4
Candidate D	5	8	12	7	7	8	8	4
Candidate E	4	10	9	6	5	7	10	7

Source: Ziotti & Leoneti (2020)

When structuring the decision matrices presented by Table 2, Ziotti & Leoneti (2020) took into account hypothetical alternatives to avoid biased decisions and the possibility that participants could consider criteria that were not presented. Furthermore, alternatives' performance in each criteria was balanced to avoid that some alternative could be dominant, so all alternatives could be chosen by the group depending on its preferences since alternatives' weaknesses and strengths were compensated.

After the simulation of the three cases, each participant filled out a survey about their satisfaction with the experiment. Using a five-level Likert scale, the following items were evaluated: the subject addressed, the duration of the application, the realism of the travel destination case, the realism of the English school case, the realism of the CEO case and, finally, an overall evaluation of the experiment. There was also a space for additional comments. Furthermore, the satisfaction survey included the following question: "Would you recommend this application to a friend? Please indicate a score between 0 and 10". This question was used to calculate the Net Promoter Score (NPS), which provided the basis for measuring the reaction of the participants and the group's degree of acceptance with the decision support methodology presented.Study design, participants, and questionnaires

In this experiment, the following players' roles are particularly important: (i) decision-makers; (ii) reporter/organizer; and (iii) analyst.

The following sub-processes were performed by the analyst to conduct the analyses during the group phase: (i) inputting the preference ordering of the criteria to calculate the weighting vectors; (ii) calculating the possible payoffs and equilibriums for the game; and (iii) calculating the distances and ordering the best equilibriums for the game.

The following question provided on the experiment template was used to obtain the criteria preference order: "Order the criteria according to your preferences". Once the criteria had been ranked, the ROC method Edwards & Barron (1994) was used to elicit the respective weighting vector for each decision-maker.

Afterwards, a non-cooperative game was modeled by applying the utility function for multi-criteria decision problems, proposed by Leoneti (2016), whose calculation results in the players' payoffs for all alternative arrangements, which can be used to help decision-makers choose strategically the desirable solution when considering the choice made by the others involved. After obtaining the payoff tables, an equilibrium solution concept is applied and, in the case of finding multiple equilibria, a social welfare ordering function is used to select the equilibrium that will represent the so-called method's solution.

To calculate alternatives' utility, it was used the weighting vectors and the decision matrices presented previously. But firstly, criteria were normalized to standardize the descriptors of performance in the same direction. Based on this data, the alternatives would be evaluated according to the preferences of all decision makers. These calculations were not presented to the participants during negotiation, so they were free to evaluate the alternatives using their own method.

Finally, the last sub-process performed by the analyst during the group phase was to calculate the Euclidean distances and present an ordering of the best Nash equilibria, using a criterion based on the entropy-norm pair, as proposed by Leoneti & Prataviera (2020), where it is assumed that, when making group decisions, individuals simultaneously seek to maximize utility and avoid inequality, i.e. issues related to justice can also influence the decisions made by the group in the search for an agreement. The steps implemented by the analyst to calculate the norm and entropy to select the best Nash equilibria are described in Leoneti & Prataviera (2020).

The reporter/organizer implemented the following sub-process to conduct the analyses required during the group phase: (i) observing and documenting the negotiation. Throughout the group phase, the reporter/organizer followed the discussions to understand and document how each decision-maker interacted during the negotiations. The script developed by Ziotti (2018) was used as reference to support the observation and documentation.

2.3. The use of IRaMuTeQ© software and qualitative data analysis

Following, t this paper also aimed to demonstrate the use of IRaMuTeQ© (Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires) software, version 0.7 alpha 2 (2014), as a tool for analyzing the perceived importance of a mathematical multicriteria decision support method. IRaMuTeQ© is a free software that has an interface with the R software, allowing the processing of textual data and providing different possibilities for analysis based on lexicography (Camargo & Justo, 2013). IRaMuTeQ© can be used to carry out different forms of statistical analysis on textual corpus, a set of texts about a particular topic gathered in a single text file. The following analysis were performed: (i) classical textual statistics; (ii) word cloud; (iii) descending hierarchical classification (DHC); and (iv) similitude analysis.

Classical textual statistics enable formal aspects of a text to be explored. To do this, the software identifies the number of words, average frequency, and number of hapax (lexical forms that occur only once), reduces words based on their roots (lemmatization) and creates a dictionary of reduced forms, identifying active and supplementary forms according to grammatical classes. The word cloud is a simple but graphically interesting lexical analysis, as it allows an overview of the keywords in a corpus and gives an initial idea of its content. This technique clusters words and organizes them graphically according to their frequency, where the largest words are those with the highest frequency and are shown in the graphic's center. The DHC method classifies text segments according to their respective vocabularies, which are partitioned according to the co-occurrence of lexical forms and the frequency of reduced forms (words that have already been lemmatized). The set of text segments is divided into two classes based on the contrast between their vocabularies and this procedure is repeated until new stable classes are no longer produced. By using this method, it is possible to obtain text segment classes, which, at the same time, have a similar vocabulary to each other and a different vocabulary to the other classes. For this purpose, chi-square tests are applied to check the degree of association between the lexical forms and the classes. The result is usually presented as a dendrogram, which brings together the specific linguistic forms of each class. Similarity analysis is based on graph theory, which studies the relationships between objects in a given set. This analysis enables the identification of co-occurrences between words, indicates the connection between linguistic forms and develops graphical representations regarding the structural content of the textual corpus. In the graph generated, the words constitute the vertices, and the edges represent the relationship between them. It is also possible to combine the result with a cluster analysis, where the different colors in the generated graph represent the clusters identified (Camargo & Justo, 2013; Sousa, 2021).

3. Results and discussion

This textual corpus was formed from the participants' responses about the importance of the method used to support group decision-making. The answers were collected from the open question "Comment about the importance of the mathematical method's suggestion to support group decision-making", answered during the final phase of each case proposed for this experiment. When answering this question, participants considered the perceived importance of the decision-making support method described in sections 2.2 and 2.3.

The textual statistics and lexicographic analysis results showed 270 text segments with a total of 3,793 word occurrences, 582 forms (active and supplementary) present in the corpus, 287 hapax and an average of 14,05 word occurrences per text. It should be noted that, to identify the number of forms present in the corpus, the lemmatized corpus was considered in the analysis. Among the most frequent active forms in the responses are “method” (128 occurrences), “group” (93 occurrences), “decision” (70 occurrences), “choice” (55 occurrences) and “choose” (34 occurrences). This result was also identified from the word cloud, a resource that allows the graphic representation of the corpus where the size of each format is proportional to its frequency, as shown in Figure 2.

Figure 2: Word cloud of the corpus (words in Portuguese)

From the DHC, there was a retention of 208 text segments, that is, 77.04% of the total corpus, which is considered adequate for the application of the technique[1]. The DHC analysis produced four lexical classes divided into two subcorpus. In the first subcorpus, class 1 was obtained, which was formed by 45 (21.63%) text segments, and class 2, which contained 49 (23.56%) text segments. From the other subcorpus, class 3 and class 4 were obtained, formed by 76 (36.54%) and 38 (18.27%) text segments, respectively.

From the data obtained through the DHC analysis, it was possible to draw interpretations about the formations of each class, in addition to understanding the similarities and differences between the conceived classes. Regarding the four classes generated for the corpus under analysis, it was noticed that the words from classes 1 and 2, as they are in the same branch of the dendrogram, presented similarities between them and a smaller relationship between those from classes 3 and 4. For each class a list of words generated from the chi-square test was computed. The DHC with word phylogram favors the visualization of the main words that form each class constructed by the software, with the higher the list and the larger the word size, the greater the influence on the class.

Following, it was found that the main words (p-value)[2] that were related to class 1 were "point", "relevant", "result", "mathematical", "consideration" and "arrive". The following text segments illustrate the characteristic content of this class: "perhaps the mathematical method helped all participants to be somewhat satisfied with the final result" (Decision-maker 4; Group 4; Case 1), "the consideration that a mathematical method can be more assertive than a group discussion was of great importance to consider it relevant" (Decision-maker 3; Group 7; Case 1), "using the mathematical method it is possible to analyze each person's options and arrive at the fairest and most pleasant result" (Decision-maker 2; Group 14; Case 1), and "it took into account the main points raised by each person and arrives at the same decision as the group" (Decision-maker 4; Group 6; Case 1). Therefore, the class 1's content dealt mainly with the result achieved by the method and was named assertiveness.

Regarding class 2, it was found that the most characteristic words (p-value criterion) of this class were "decision", "make", "confirm", "recommendation" and "equal". The following text segments illustrate the characteristic content of this class: "the recommendation was the same as the group's decision helped to confirm our choice" (Decision-maker 4; Group 12; Case 1), "the method ensured that the decision made was in accordance with the group's opinion" (Decision-maker 4; Group 1; Case 2), "the method's recommendation was in accordance with the group's decision" (Decision-maker 3; Group 12; Case 1) and "helped to legitimize the decision made by the group" (Decision-maker 2; Group 4; Case 2). Therefore, the class 2's content dealt mainly with how the method helped the group's agreement and was named affinity. Therefore, the first subcorpus formed essentially addressed the usual characteristics of the negotiation process, which are affinity and assertiveness.

Regarding the second subcorpus, it was found that the main words (p-value criteroin) that were related to class 3 were "show" and "choice". The following text segments illustrate the characteristic content of this class: "it was good to show how much closer the other alternative was to justify the greater debate for the choice" (Decision-maker 5; Group 2; Case 2), "the method showed the group's optimal choice" (Decision-maker 3; Group 4; Case 1), "the method offers a better choice parameter and as we reached consensus, the method reinforced this" (Decision-maker 5; Group 1; Case 1), "since the group was more certain about the choice of the English school and the method reaffirmed it, there was also consensus in the voting" (Decision-maker 3; Group 7; Case 2) and "since the group didn't have a consensus it was important for a good choice" (Decision-maker 4; Group 9; Case 2). Therefore, the class 3's content dealt mainly with how the method can be useful as a predictive tool and was named as predictivity.

Regarding class 4, it was found that the most characteristic words (p-value criterion) of this class were "believe", "case", "count", "doubt", "discuss", "match" and "suggestion". The following text segments

[1] By implementing DHC, the most characteristic text segments of each class are identified. However, at the end of the procedure, not all text segments are classified, and the analysis is only considered satisfactory if the method retains at least 75%.
[2] It was considered in this research a criterion of p-value less than 0.0001.

illustrate the characteristic content of this class: “it was very different from what we discussed in this case I had no doubts when voting for the negotiated option” (Decision-maker 1; Group 14; Case 3), “the suggestion took into account other factors that were not discussed in the group, which makes it positive” (Decision-maker 5; Group 6; Case 3), “in this case the method was not taken into account as the option chosen did not benefit the majority” (Decision-maker 2; Group 17; Case 3), “in this case the method did not agree with most people's opinions but took into account what was best for everyone” (Decision-maker 1; Group 5; Case 3) and “I don't believe that in this case the method has reached a fair choice” (Decision-maker 4; Group 9; Case 2). Therefore, the class 4’s content mainly dealt with how the tool can be useful as a normative tool and was named normativity. Therefore, the second subcorpus formed essentially addressed the technical potential of a decision support tool.

Comparatively, Ziotti (2018) identified that in a group decision-making environment supported by a multi-criteria decision method, there are different perceptions regarding the importance of the method when it does or does not converge with the decision negotiated by the group. Thus, Ziotti (2018) noted that, in decisions where the decision-makers were compromised with the negotiated alternative and the method supported the decision, the method provided certainty to the decision-making process, which led to better levels of satisfaction and fairness. On the other hand, when there was no commitment and no support from the method in the decision, the method served to minimize the group's conflict and indecision. It can therefore be seen that the perception of importance is a relevant measure for analyzing the usefulness and acceptability of a method for supporting the decision-making process.

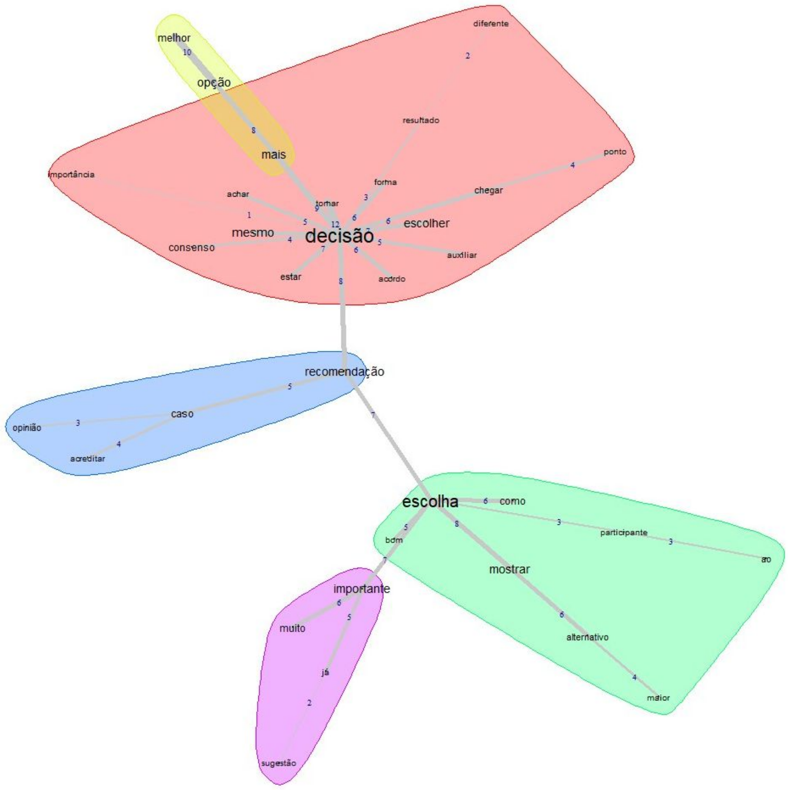


Figure 3: Formed structures and cores (words in Portuguese)

The similarity analysis made it possible to identify structures and cores in a textual corpus, in addition to also evaluating the interconnection between words and the level of relationship between them. The corpus similarity analysis graph regarding the importance of the method was formed by five cores, whose keywords for each core were: decision, choice, option, recommendation and important. In the graph obtained (Figure 3, in Portuguese), it was noticed that there was not a single prominent central word. There was similar emphasis on the words “decision” and “choice” so that they are not directly linked, as the word “recommendation” appears among them, which also includes “case”, “opinion” and “believe”. In relation to the word “decision”, the words “choose”, “same”, “consensus”, “result”, “agreement”, among others, are part of this nucleus. In relation to the word “choice”, this nucleus comprises the words “show”, “alternative”, “participant”, among others. Linked to the term “decision” there was also an attached set in which the words “option”, “better” and “more” are highlighted. Finally, there was a group attached to the term “choice” whose highlighted words were “important”, “very” and “suggestion”.

Consequently, the first subcorpus (classes 1 and 2) deals with aspects of the negotiation process and the use of the decision support method while the second subcorpus (classes 3 and 4) deals with the opinion of a technique as a tool and its potential predictive or normative aspects. It is important to highlight the importance of the analyst's role in identifying such characteristics of the decision-making process and evaluating what is the best way to conduct the decision-making process, in addition to which methodology is most effective in supporting decision-makers throughout this process.

Finally, the answer given by participants to the question “Would you recommend this application to a friend?” Indicate a score ranging between 0 and 10” in order to analyze the probability of participants indicating the proposed methodology to support the group’s decision. This analysis served as a basis for evaluating the participants' reaction and the group's degree of acceptance and satisfaction with the proposed methodology based on the NPS calculation. To do this, participants were divided into three groups based on the answer given: (i) promoters; (ii) passively satisfied; and (iii) detractors. Then, the groups were categorized according to the NPS percentage, with six groups with 100%, 7 with 80%, 4 with 60%, 2 with 50% and 1 with 40%. In general, it was noted that groups presented a high degree of satisfaction with the decision support methodology proposed for the experiment considering that, of the 20 groups that participated in this study, 19 presented an NPS with a percentage above 50%.

4. Conclusions

By means of an experimental study with the application of a visual and textual approach, it was possible to analyze the perceived importance of a decision support method in a group decision-making process. The main results have shown that there is, at least, two perspectives that participants identified. The first is the use of the method as a support to the negotiation skill, which demonstrated to have a positive perspective in the view of the participants. The second was related to the technical features of the method itself, which tended to be seen in non-positive view in the perspective of the participants. Concerning the last point, it should be noticed that although a critical view of the technical itself the participants demonstrated a high satisfaction level with the entire approach. These results are relevant for guiding managers in the decision-making process and structuring it more efficiently. Furthermore, conclusions from this research can also improve the methodology used to apply the experiment, including the use of the IRaMuTeQ©, which have not been applied in similar context. Additionally, evidences were found that its use facilitate and improve decision-making process.

Acknowledgments. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. To Iana Valéria Ferreira da Silva, for the aid with the organization of the experiments

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

Baker, D., Bridges, D., Hunter, R., Johnson, G., Krupa, J., Murphy, J., & Sorenson, K. (2001). *Guidebook to*

Decision-Making Methods, Department of Energy.

Camargo, B. V., & Justo, A. M. (2013). IRAMUTEQ: Um Software Gratuito para Análise de Dados Textuais. *Temas Psicol., 21*(2), 513-518. http://dx.doi.org/10.9788/TP2013.2-16.

Clemen, R. T., & Reilly, T. (2013). *Making Hard Decisions with DecisionTools®* (3rd ed.). Cengage Learning.

de Almeida, A. T. (2013). *Processo decisório nas organizações: construindo modelos de decisão multicritério*. Atlas.

de Almeida A. T., Cavalcante, C. A. V., Alencar, M. H., Ferreira, R. J. P., de Almeida-Filho, A. D., & Garcez, T. V. Multiobjective and Multicriteria Decision Processes and Methods. In: International Series in Operations Research & Management Science, pp. 23-87. Springer, Cham (2015)

de Almeida, A. T., Morais, D. C., & Nurmi, H. (2019). Systems, Procedures and Voting Rules in Context. Advances in Group Decision and Negotiation. Springer.

da Silva, M. R., Gimon, D., & Leoneti, A. B. (2020). Using a refined utility function for predicting sequential group decision making within a heterogeneous case solving process. In *International Conference on Group Decision and Negotiation* (pp. 35.1-35.12), Ryerson University, Toronto.

Edwards, W., & Barron, F. H. (1994). Improved Simple Methods for Multiattribute Utility Measurement. *Organizational Behavior and Human Decision Processes, 60*(3), 306-325. https://doi.org/10.1006/obhd.1994.1087

Ferreira, M. M. G., Ferreira Júnior, S. N., de Sousa, J. R., Eguchi, T. Y., da Silva, R. G., & Serra, C. M. V. (2018). Escolha dos métodos de multicritério a tomada de decisão com o auxílio de um sistema especialista formulado a partir de um fluxograma. *Braz. Ap. Sci. Rev., 2*(5), 1593-1608. https://doi.org/10.34115/basr.v2i5.545.

Fiani, R. (2009). Teoria dos Jogos (3rd ed). Elsevier.

Hammond, J. S., Keeney, R. L., & Raiffa, A. (1998). The Hidden Traps in Decision Making. *Harvard Business Review: Decision making and problem solving*. https://hbr.org/1998/09/the-hidden-traps-in-decision-making-2.

IRAMUTEQ. Versão 0.7 alpha 2. Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires, 2014. Available at: http://www.iramuteq.org/documentation.

Leoneti, A. B. (2016). Utility Function for modeling Group Multicriteria Decision Making problems as games. *Operations Research Perspectives*, 3, 21-26. https://doi.org/10.1016/j.orp.2016.04.001

Leoneti, A. B., & de Sessa, F. (2017). A deviation index proposal to evaluate group decision making based on equilibrium solution. In *International Conference on Group Decision and Negotiation*, (pp. 101-112), Springer, Cham.

Leoneti, A. B., de Sessa, F., & Marques, M. T. (2015). Negociação com o uso de métodos de apoio à tomada de decisão: metodologia e prática. In *XVIII SEMEAD*, (pp. 1-16), FEA-USP, São Paulo.

Leoneti, A. B., & Prataviera, G. A. (2020). Entropy-norm space for geometric selection of strict Nash equilibria in n-person games. *Physica A*, 546, 1-6. https://doi.org/10.1016/j.physa.2020.124407

Porat, T., Delaney, B., & Kostopoulou, O. (2017). The impact of a diagnostic decision support system on the consultation: perceptions of GPs and patients. *BMC Medical Informatics and Decision Making,* 17(79), 1-9. https://doi.org/10.1186/s12911-017-0477-6

Ragsdale, C. T. (2019). *Modelagem de planilha e análise de decisão: uma introdução prática a business analytics* (8th ed). Cengage Learning.

Render, B., Stair Júnior, R. M., Hanna, M. E., & Hale, T. S. (2018). *Quantitative analysis for management* (13th ed). Pearson.

Shibl, R., Lawley, M. & Debuse, J. (2013). Factors influencing decision support system acceptance. *Decision Support Systems, 54*, 953-961. https://doi.org/10.1016/j.dss.2012.09.018

Sofo, F., Colapinto, C., Sofo, M., & Ammirato, S. (2013). *Adaptive Decision Making and Intellectual Styles*. Springer.

Sousa, Y. S. O. (2021). O Uso do Software Iramuteq: Fundamentos de Lexicometria para Pesquisas Qualitativas. *Estudos e Pesquisas em Psicologia, 21*(4), 1541-1560. https://doi.org/10.12957/epp.2021.64034

Simon, H. A. (1987). Making Management Decisions: the Role of Intuition and Emotion. *The Academy of Management Executive, 1*(1), 57-64. https://www.jstor.org/stable/4164720.

Ziotti, V. C. (2018). *Senso de justiça e comprometimento na decisão em grupo: análise do apoio de método multicritério baseado na teoria dos jogos* [Master's thesis, Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, Universidade de São Paulo]. Biblioteca Digital de Teses e Dissertações da USP. https://teses.usp.br/teses/disponiveis/96/96132/tde-07032019-105005/publico/VanessaCZiotti_Corrigida.pdf.

Ziotti, V. C., & Leoneti, A. B. (2020). Improving commitment to agreements: the role of group decision making methods in the perception of sense of justice and satisfaction as commitment predictors. *Pesquisa Operacional, 40*, 1-15.

Breaking Through Deadlock: Group Consensus Model Based on Regret Theory

Sainan Guo [1*], Rudolf Vetschera [1] and Jian Wu [2]

[1] Department of Business Decisions and Analytics, Vienna University, Vienna, Austria
[2] School of Economics and Management, Shanghai Maritime University, Shanghai, China
[*] sainan.guo@univie.ac.at

Abstract

Group decision-making is frequently confronted with the challenge of deadlock, where experts are unable to reach a consensus due to conflicting preferences and because no one wants to give in. This paper proposes a minimum regret consensus model in a group that considers the presence of regret emotions between decision makers' individual opinions and the group consensus opinion. It minimizes the total regret value to achieve consensus by all experts and thus suggests ways to promote collaboration and enhance the quality of the group decision-making process while minimizing the associated complaints and regret emotions. We further study how different levels of regret influence the outcome of the group decision and the possibilities for one member to prevail. We find that with increasing regret aversion, it becomes easier for experts holding an extreme opinion to dominate the group decision.

Keywords: group decision; inconsistent expert opinions; minimum regret consensus model; regret theory

1. Introduction

In the realm of decision-making, individuals often face diverse selections, leading to differing perspectives and decisions *(Bose et al., 1997)*. Group Decision-Making (GDM) is concerned with finding ways to reconcile these disparate views and opinions to arrive at a unified group view and decision. GDM, dating back centuries, involves collaboration among individuals to influence group objectives. Meanwhile, challenges arise as group decisions often diverge from individual choices, influenced by factors like an individual's role, group size, and controllable variables. These challenges can lead to deadlock, particularly when there is incomplete information among experts within an untrusted social network, and experts exhibit non-cooperative behavior *(Gai et al., 2023; Zhao et al., 2024)*. Consequently, researchers actively explore methodologies to enhance group decision-making processes, mitigating deadlock and striving for superior outcomes. Notably, models like the Minimum Cost Consensus Model (MCCM) have gained significant attention and recognition in addressing this challenge *(Ben-Arieh & Easton, 2007)*.

This research introduces a bridge between regret theory and consensus models and considers the 'cost' from an emotional perspective, recognizing that individuals within a group might experience regret if the group choice diverges from their individual choice. Utilizing the regret theory and the structure of minimum cost consensus models, we model the varying perceived complaint (regret) values among decision-makers (DMs) in the context of group consensus decision-making and introduce a novel group consensus model, the Minimum Regret Consensus Model.

The main goal of this study is to estimate how decision-makers with bounded rationality (regret emotion) influence the group consensus opinion while minimizing group complaints (regret values). We thus address a major challenge faced by MCCM, offering a more interpretable perspective on the 'cost' associated with consensus, thus advancing the understanding of GDM processes.

Interpreting deviations between group decisions and individual decisions of group members as causes of regret opens up another research question that we address in this paper: Group members who feel regret when the group decision deviates from their own opinion might wish for a situation in which they could determine the group decision, i.e. be a kind of 'dictator' in the group. We therefore study whether the presence of regret among group members influences the possibilities of exerting a dominant influence on the group. We determine under which circumstances one member can dominate the group decision (become in fact a dictator) and thus will not have any regret. We also present simulation results that examine the influence of different parameters (such as regret aversion coefficients, weights, and group size) on group consensus and dissatisfaction.

2. Regret Emotion and Satisfaction in Group Consensus

No matter which method is used to reach a group consensus, DMs commonly express regretful emotions upon comparing group consensus opinions to their individual opinions because of the differences between the two decisions *(Guo & Vetschera, 2023)*. Therefore, this negative regretful emotion will lead to DMs' dissatisfaction with the final group decision. Previous studies by *Taylor (1997)* confirm past findings about the impact of regret on predicted consumer satisfaction. Furthermore, as the expected evaluation of the unchosen alternatives increased, satisfaction with the selected alternative decreased. These observations have been substantiated in subsequent studies *(Herrmann et al., 1999; Tsiros & Mittal, 2000)*.

Therefore, it is important to explore the impact of regret emotion during the decision-making process, especially when comparing cases in decisions under risk and group decisions *(Bell, 1982; Loomes & Sugden, 1982; Chorus, 2012)*. This paper explores the impact of regret effect on DMs' satisfaction within the context of group consensus, defining group regret as an index for group dissatisfaction. For example, if the group consensus closely aligns with their individual opinions, satisfaction ensues. Conversely, greater dissatisfaction prevails if the group consensus significantly deviates from their individual choices. Building upon the model proposed by *Ben-Arieh & Easton (2007)*, our research extends and concentrates on achieving minimum complaint (maximum satisfaction) consensus by minimizing aggregated regret values in group decisions.

3. Minimum Regret Consensus Model

3.1. Construction of the model

Let $E = \{e_1, e_2, \dots, e_n\}$ be a set of n group members or experts, $n \geq 2$. Assume that they have originally normalized assessment opinions $O = (o_1, o_2, \dots, o_n)$, $o_i \in [0,1]$ for all $i = 1,2, \dots, n$. We also consider that experts have a different influence on the group, e.g. because of their different fields or level of knowledge. These are represented by a weight set $W = (w_1, w_2, \dots, w_n)$, where $0 \leq w_i \leq 1, \sum_{i=1}^{n} w_i = 1$. A consensus, representing the final group decision, is confirmed when all experts' individual opinions converge to a common value o', where $o' = o'_1 = o'_2 = \cdots = o'_n$ and $o' \in [0,1]$. The Minimum Regret Consensus Model (MRCM) model thus considers one unique consensus value. Extensions to different forms of consensus such as ε soft consensus *(Ben-Arieh & Easton, 2007)* would be possible.

Define a complaint (regret) for each expert, which occurs if and only if the final group decision differs from the expert's initial opinion. Higher levels of dissatisfaction arise when the group consensus significantly deviates from the expert's original choices. This refined definition of satisfaction and dissatisfaction establishes a more detailed understanding of expert emotions in the context of GDM.

The dissatisfaction of the i-th expert in the group can be quantified using the regret value $R_i = R_i(o', o_i, \delta)$. Here, $R(\cdot)$ represents a strictly increasing concave function, given by $R_i(o', o_i, \delta) = 1 - exp\ (-\delta|o' - o_i|)$, which depends on the group consensus opinion or final group decision o', the expert's initial opinion o_i, and the regret aversion level δ.

Similar exponential functions are widely used in regret theory (*Chorus, 2012; Bleichrodt et al., 2010*). The application of exponential functions to represent expert regret values can be explained simply by checking their mathematical attributes. Exponential functions are characterized by rapid growth as the input value increases. This attribute proves advantageous in regret modeling, as it mirrors the pronounced emotional response elicited by significant deviations from anticipated outcomes. During decision-making processes, individuals commonly experience more regretful feelings due to larger deviations from the optimal choice, which exponential functions capture well.

The regret function in this research offers a new measurement, capturing the psychological impact of adjustment between the group preference and an expert's initial preference, modulated by the regretful sensitivity parameter δ. Formally, the weighted consensus at minimum regret values in MRCM can be expressed as follows,

$$\min R(o') = \sum_{i=1}^{n} w_i * R_i(o', o_i, \delta).$$

The optimal consensus opinion at minimum regret values o^* in MRCM can be obtained as,

$$o^* = \underset{o' \in [0,1]}{argmin} \left\{ \sum_{i=1}^{n} w_i * R_i(o', o_i, \delta) \right\}$$

$$= \underset{o' \in [0,1]}{argmin} \{ w_1 * (1 - \exp(-\delta|o' - o_1|)) + w_2 * (1 - \exp(-\delta|o' - o_2|)) +$$

$$\ldots + w_n * (1 - \exp(-\delta|o' - o_n|)) \}.$$

The following proposition shows that the optimal consensus will always correspond to the opinion of one expert:

Proposition 1. If all $R_i(\cdot)$ are concave and monotonically increasing in $|o' - o_i|$, the minimum of R is located at one of the points o_i.

Detailed proof is given in the full working paper *(Guo et al., Available at SSRN 4710027)*. Utilizing this proposition simplifies the process of finding the optimal consensus value, which means, one just has to check all values of original individual opinions $(o_1, o_2, \ldots, o_n)$ by group members. The MRCM can be optimized as an approach to derive group consensus from experts' opinions. This model finds application in evaluating consensus dissatisfaction, especially in cases where selecting entirely new group opinions is not feasible, and the choices are limited to those provided by the experts. This scenario is particularly relevant when the pools of available opinions are fixed.

Based on our previous assumption that R_i is a strictly increasing concave function, the objective function can be reformulated as:

$$o^* = \underset{o' \in O}{argmin} \left\{ \sum_{i=1}^{n} w_i * R_i(o', o_i, \delta) \right\}.$$

3.2. Numerical Example

We now illustrate the MRCM for three distinct groups of different sizes n = 3, n = 5, n = 7. The initial opinions and weights of the group members are shown in Table 1.

Table 1. Original opinions and weights for three groups

		Weight w_i in group		
Member	o_i	**n = 3**	**n = 5**	**n = 7**
1	0.7902	0.3172	0.1798	0.1398
2	0.9499	0.3885	0.2202	0.1713
3	0.2031	0.2944	0.1669	0.1298
4	0.6125		0.0964	0.0750
5	0.0592		0.3367	0.2619
6	0.3840			0.0393
7	0.7176			0.1830

We use members with the same initial opinions in all groups, however their weights need to be adjusted proportionally to the different group sizes. Results of the model are shown in Table 2 for different levels of regret aversion, where we assume that all members have the same level of regret aversion.

Table 2. MRCM results for the groups in Table 1

	Group n = 3		Group n = 5		Group n = 7	
δ	o'	**Regret**	o'	**Regret**	o'	**Regret**
1	0.7902	0.1881	0.2031	0.2732	0.7176	0.2424
5	0.9499	0.4618	0.0592	0.5687	0.7176	0.5948
10	0.9499	0.5471	0.0592	0.6231	0.7176	0.7039

In the MCCM, as the number of group members increases, the consensus cost always rises. However, in the MRCM, it may happen that even for larger numbers of group experts, the group regret may not necessarily increase because of changing weights. Notably, with changing levels of the regret aversion coefficient δ ($\delta = 1 \rightarrow \delta = 5 \rightarrow \delta = 10$), the group minimum regret increases (0.1881 → 0.4618 → 0.5471; 0.2732 → 0.5687 → 0.6231; 0.2424 → 0.5948 → 0.7039). For the smaller groups, the consensus opinion tends to move towards the extreme ends of the range of opinions. This can be observed for the group with n = 3 members, where the consensus moved from the median value to the maximum values when δ reaches the value of 5, and also for the group with n = 5 members where the group opinion jumps to the lowest value also at δ = 5.

3.3. Dictatorship in the MRCM

As MRCM is applied to GDM problems where group consensus emerges from all experts' preferences, it might be interesting for group members to find out when their opinion prevails. Since the importance of experts in the group is represented by their weights, we therefore analyze the question which weight (i.e., how much power) a group member needs to make his or her opinion the group consensus. Specifically, we consider two particular group members:

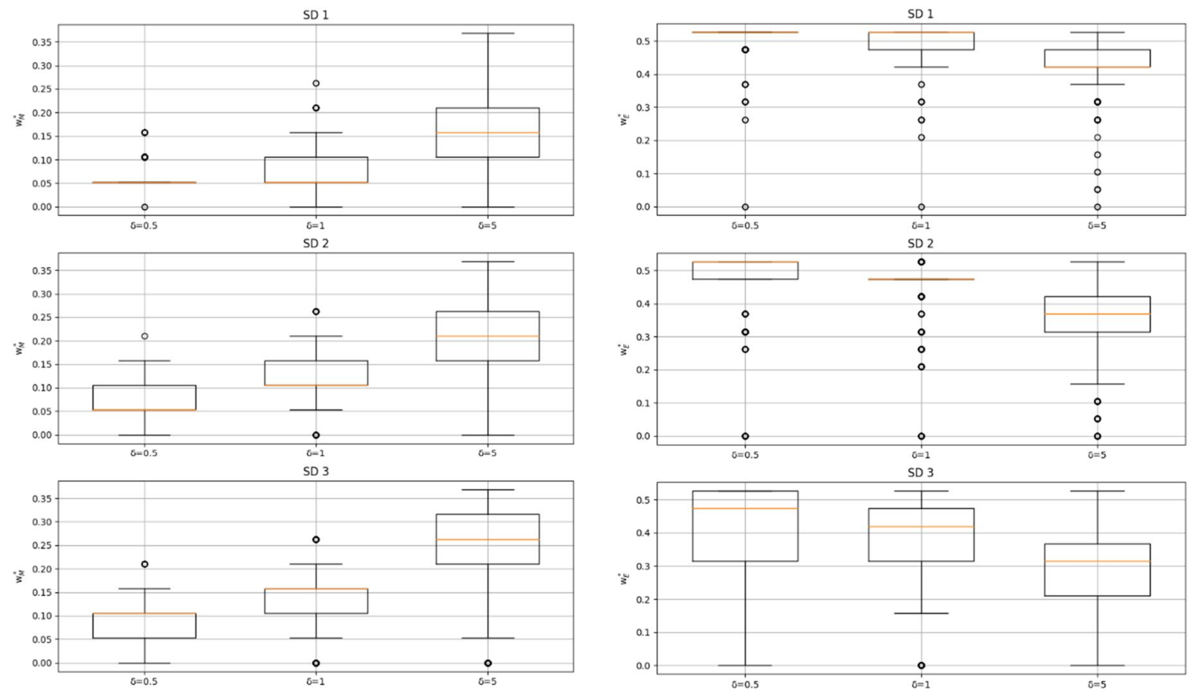
1. *Median Opinion Holder*: the expert holding the median opinion within the group (thus aligned with the majority).

2. *Extreme Opinion Holder*: the expert holding the most extreme opinion within the group (deviating from the majority).

This analysis illuminates the dynamics of power distribution and influence strategies within the GDM context. We estimate the weight threshold above which the expert’s opinion dominates the group opinion for both median and extreme opinion holders in GDM problems. Following the conclusion of Proposition 1, in this MRCM, the final group opinion will be optimized as an approach for deriving group consensus from experts' personal opinions. This implies that a specific expert's individual opinion will become the group consensus. In such a scenario, all experts aim to minimize the weight they contribute to the group to achieve the dictatorship. On the other words, the expert can be able to use smaller weight in the group (<50%) to influence the group decision reflects their personal decision. To control the heterogeneity of opinions, we use test cases with a predefined Standard Deviation (SD) of opinions.

3.4. Dictatorial Weight Threshold for Two Expert Types

We simulated groups of 5 experts (n = 5), generating their original opinion set $O = (o_1, o_2, \dots, o_n) \in [0,1]$ randomly for each simulation. The opinion sets were generated under distributions with three standard deviations (SD=0.1, SD=0.25, SD=0.4). We used three specified values (δ = 0.5, δ =1, δ = 5) of the regret aversion coefficient δ. Subsequently, we recorded all dictatorial weight thresholds (i.e., the minimum weight that would make that member’s opinion the group opinion) for the two types of experts. We denote these thresholds by, w_M^* for the Median Opinion Holder and w_E^* for the Extreme Opinion Holder. Simulation results for 2000 group consensus problems are presented in the box plots in Figure 1.



The trends depicted in the two box plots illustrate opposite patterns of two expert types during the group consensus process in MRCM model. Specifically, the minimum dictatorial weight value w_M^* for an expert holding a median opinion gradually increases with the increasing regret coefficient of the group. This implies that it becomes progressively more challenging to impose his individual choice as the group decisions. Conversely, for an expert holding an extreme opinion, the required dictatorial weight value w_E^* decreases with the rising regret coefficient of the group. This indicates an increasing advantage for the extreme opinion holder when seeking to assert dominance for his or her personal opinion within the group consensus.

4. Conclusions

This study delves into the challenge of achieving group consensus under conditions of bounded rationality. By combining the framework of the Minimum Cost Consensus Model with regret theory, we formulate a novel model named MRCM to achieve group consensus when regret emotions are present. This model aims to address the optimal group consensus while considering psychological regret feelings in group decisions. It also provides an alternative interpretation to the concept of costs in MCCM. Notably, through the proof of Proposition 1, this model makes the process of finding optimal values easier (from the original opinion set directly rather than by solving a possibly nonlinear optimization model).

It further analyzes under which conditions the expert holding a specific opinion will prevail in the group decision. We compare two types of experts, namely the Median Opinion Holder and the Extreme Opinion Holder. In our simulation, we analyze the dictatorial weight thresholds required for these two experts to assert dominance in GDM. The outcomes reveal diverging trends in the minimum weight necessary for dictatorship as the regret aversion increases in the group for both types. The results suggest that, as a median opinion holder, it becomes more and more difficult to dominate individual choice into group decision with the increase of regret aversion coefficient in the group. On the contrary, there is an increasing advantage for an extreme opinion holder to seek the dictatorship in the group with a stronger regret aversion level.

Moreover, our model focuses primarily on investigating how psychological emotions influence decision-making. A critical improvement involves translating these psychological impacts into quantifiable economic indicators. Such a transformation could foster a more robust integration of our model with MCCM, contributing to optimal group consensus and enhancing GDM processes, particularly in domains where emotions are pivotal, such as healthcare decision-making (patient treatment selection, healthcare resource allocation), policy formulation, and financial investment strategies.

References

Bell, D. E. (1982). Regret in decision making under uncertainty. *Operations research, 30*, 961-981.

Ben-Arieh, D., & Easton, T. (2007). Multi-criteria group consensus under linear cost opinion elasticity. *Decision Support Systems, 43*, 713–721.

Bleichrodt, H., Cillo, A., & Diecidue, E. (2010). A quantitative measurement of regret theory. *Management Science, 56*, 161–175.

Bose, U., Davey, A. M., & Olson, D. L. (1997). Multi-attribute utility methods in group decision making: past applications and potential for inclusion in GDSS. *Omega, 25*, 691–706.

Chorus, C. G. (2012). Regret theory-based route choices and traffic equilibria. *Transportmetrica, 8*, 291–305.

Gai, T., Cao, M., Chiclana, F., Zhang, Z., Dong, Y., Herrera-Viedma, E., & Wu, J. (2023). Consensus-trust driven bidirectional feedback mechanism for improving consensus in social network large-group decision making. *Group Decision and Negotiation, 32*, 45–74.

Guo, S., & Vetschera, R. (2023). Preference reversals in dynamic decision-making under uncertainty based on regret theory. *Managerial and Decision Economics, 44*, 1720–1731.

Guo, S., Vetschera, R., & Wu, J. Breaking Through Deadlock: Group Consensus Model Based on Regret Theory. *Available at SSRN 4710027*.

Herrmann, A., Huber, F., & Braunstein, C. (1999). A regret theory approach to assessing customer satisfaction when alternatives are considered. *ACR European Advances*.

Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal, 92*, 805–824.

Taylor, K. (1997). A regret theory approach to assessing consumer satisfaction. *Marketing Letters, 8*, 229–238.

Tsiros, M., & Mittal, V. (2000). Regret: A model of its antecedents and consequences in consumer decision making. *Journal of Consumer Research, 26*, 401–417.

Zhao, S., Wu, S., & Dong, Y. (2024). Managing non-cooperative behaviors and ordinal consensus through a self-organized mechanism in multi-attribute group decision making. *Expert Systems with Applications, 240*, 122571.

joint conference group and negotiation DECISION support systems technology Porto 2024

24th International Conference on Group Decision and Negotiation &
10th International Conference on Decision Support System Technology
Human-Centric Decision and Negotiation Support for Societal Transitions

Linguistic multi-criteria support for evaluation of negotiation offers

Marzena Filipowicz-Chomko[1*], Ewa Roszkowska[1]

1 Faculty of Computer Science, Bialystok University of Technology, Wiejska 45A, 15-351 Bialystok, Poland
[*] m.filipowicz@pb.edu.pl

Abstract

This paper aims to introduce a tool designed for the linguistic evaluation and rank-ordering of negotiation offers. The proposed approach utilizes a linguistic ideal-based multi-criteria method, linking linguistic terms with a numerical scale. The GDM2 measure is used to facilitate the measurement of the distance between options values represented on an ordinal scale. The key advantage of this method lies in its utilization of natural language, accommodating subjectivity, imprecision, and uncertainty in the negotiation offers evaluation. Furthermore, the incorporation of an ideal, representing the maximum linguistic score, helps prevent rank reversal when assessing new offers. Our proposal serves as an alternative for linguistically evaluated negotiation offers, where linguistic terms are expressed using fuzzy sets. The suggested approach is straightforward to implement in a negotiation support system and can be successfully utilized in various negotiation scenarios.

Keywords: multiple criteria decision making; preference analysis; linguistic evaluation negotiation offers; negotiation scoring system

1. Introduction

Building a negotiation scoring system using multi-criteria decision-making (MCDM) methods involves structuring the negotiation process, defining relevant criteria, and developing a systematic approach to evaluate and score negotiation offers (Wachowicz & Roszkowska, 2021). The MCDM approach enhances objectivity and transparency in the negotiation process, leading to more informed and strategic negotiation outcomes. Therefore, in the literature, we can find several propositions for determining the negotiation scoring systems using the MCDM methods: DR (Direct Rating) (Kersten & Noronha, 1999), AHP (Analytic Hierarchy Process) (Brzostowski et al., 2012; Mustajoki & Hämäläinen, 2000), TOPSIS (Technique for Order Preferences by Similarity to Ideal Solution) (Roszkowska & Wachowicz, 2012, 2015a; Wachowicz & Błaszczyk, 2013), MARS (Measuring Attractiveness near Reference Situations) (Górecka et al., 2016; Roszkowska & Wachowicz, 2015b), UTA (UTilités Additives) (Wachowicz & Roszkowska, 2022), ELECTRE (ÉLimination Et Choix Traduisant la Realité) (Wachowicz, 2010), among others.

The negotiation encompasses both quantitative aspects (e.g., cost, delivery time) and qualitative factors (e.g., warranty, relationship building). To address subjective or qualitative criteria that are challenging to precisely quantify, linguistic evaluation of negotiation offers can be particularly useful. Negotiators may employ linguistic terms such as 'very good,' 'good,' 'fair,' 'poor,' and 'very poor to articulate their preferences for different issues or options. These linguistic terms are often associated with numerical scales or fuzzy sets, allowing for the representation of uncertainty and vagueness in decision-making. The linguistic values can be represented in ordinal scale (Xu, 2012); through fuzzy sets (Herrera et al., 2008; Herrera & Herrera-Viedma, 2000); intuitionistic fuzzy sets (Zhang, 2014), or ordered fuzzy sets (Piasecki & Roszkowska, 2018). The assessment of negotiation offers linguistically has been explored in some papers (Filipowicz-Chomko et al., 2021; Piasecki & Roszkowska, 2018; Wachowicz, 2011), among others. Wachowicz (2011) for linguistic evaluation negotiation offers the proposed TOPSIS method with the use of GDM2 (Generalized Distance Measure). Piasecki & Roszkowska (2018) for building a negotiation scoring system applied Linguistic Fuzzy SAW or Fuzzy TOPSIS where linguistic values were represented by trapezoidal ordered fuzzy numbers. Filipowicz-Chomko et al (2021) to support negotiation proposed a fuzzy clustering model, where linguistic terms are represented by triangular fuzzy numbers.

The paper aims to present a novel linguistic ideal-based multi-criteria (LIBM) method that can be applied to evaluate negotiation offers and rank ordering them. This method relies on the concept of distance from the ideal solution and employs the GDM2 measure, which facilitates measuring the distance between values represented on an ordinal scale. Our proposition is an alternative for evaluation negotiation offers linguistically using fuzzy sets. The LIBM method belongs to the class of multi-criteria methods based on reference points, and it is particularly closely related to the Hellwig (Hellwig, 1968) or DARP (Roszkowska et al., 2020) method. The LIBM, similar to Hellwig and DARP, relies on the concept of measuring distances between alternative and ideal solutions. However, unlike the classical Hellwig or DARP approach, which is based on interval and ratio data, LIBM utilizes ordinal data. Therefore, the Euclidean distance implemented in Hellwig and DARP is replaced by the GDM2 measure, enabling the calculation of distances for ordinal data.

2. Linguistic multi-criteria negotiation support

In this section, we will present a multi-criteria negotiation support model that utilizes the Linguistic Ideal-Based Multi-Criteria (LIBM) method. The LIBM method falls under the category of multi-criteria methods based on reference points, and it is particularly closely related to the Hellwig (Hellwig, 1968) and DARP (Roszkowska et al., 2020) methods.

The LIBM method uses the concept of the distance between an object (alternative) and an ideal object and operates under the assumption that objects are assessed using an ordinal scale. The ordinal nature of the data necessitated the utilization of a specific measure for assessing the distance between objects. The assessment of an object relied on numerical relationships of 'equal to,' 'greater than,' and 'smaller than". In the LIBM method, we employed the distance measure for ordinal data introduced by (Walesiak, 1999).

Definition (Walesiak, 1999) Let $OS = \{1,2,\ldots,k\}$ be a numerical representation of the ordinal scale, where the higher number means "more preferable" and $O = \{O_1, \ldots O_n\}$ set of objects evaluated on ordinal scale OS with respect m criteria. Let $O_i = [x_{i1}, x_{2j}, \ldots, x_{im}]$, $O_k = [x_{k1}, x_{k2}, \ldots, x_{km}]$ be representation i-th (k-th) object respectively; x_{ij} (x_{kj}) $-$ evaluation of i-th (k-th)object with respect to j-th criterion; $x_{ij}, x_{kj} \in OS$ ($j = 1, \ldots m$; $i, k = 1,2,\ldots,n$). The GDM2 distance between O_i and O_k objects characterized by m criteria has the following form:

$$GDM2_{ik} = \frac{1}{2} - \frac{\sum_{j=1}^{m} \mathrm{w}_j a_{ikj} b_{kij} + \sum_{j=1}^{m} \sum_{l=1;\, l \neq i,k}^{n} \mathrm{w}_j a_{ilj} b_{klj}}{2\left[\left(\sum_{j=1}^{m} \mathrm{w}_j a_{ikj}^2 + \sum_{j=1}^{m} \sum_{l=1;\, l \neq i,k}^{n} \mathrm{w}_j a_{ilj}^2\right)\left(\sum_{j=1}^{m} \mathrm{w}_j b_{kij}^2 + \sum_{j=1}^{m} \sum_{l=1;\, l \neq i,k}^{n} \mathrm{w}_j b_{klj}^2\right)\right]^{\frac{1}{2}}}, \tag{1}$$

where $i, k, l = 1, \ldots, n$ - the number of objects, $j = 1, \ldots, m$ – the number of criteria, $\mathrm{w}_j -$ weights of j-th criterion, $x_{ij}(x_{kj}, x_{lj})$ – evaluation of i-th (k-th, l-th) object with respect j-th criterion.

For the ordinal scale, a_{ipj}, b_{krj} are given as

$$a_{ipj}(b_{krj}) = \begin{cases} 1 \text{ if} & x_{ij} > x_{pj}\ (x_{kj} > x_{rj}) \\ 0 \text{ if} & x_{ij} = x_{pj}\ (x_{kj} = x_{rj}), \text{for } p = k, l; r = i, l \\ -1 \text{ if} & x_{ij} < x_{pj}\ (x_{kj} < x_{rj}) \end{cases} \tag{2}$$

The linguistic multi-criteria support for evaluating negotiation offers comprises the following phases and steps:

Phase I: Building a scoring system for all potential negotiation offers.
Step 1. Determining the set of negotiation issues and the sets of negotiation options.

Let $C = \{C_1, C_2, \ldots, C_m\}$ be the set of issues and $X_j = \{x_j^i\}_{i=1,\ldots,n_j}$ is a set of n_j of options for j-th issue ($j = 1,2, \ldots m$).

Step 2. Determining the linguistic scale OS with its numerical representation and linguistically evaluating options from the sets X_j ($j = 1,2, \ldots m$) using this scale.

Let $OS = \{1,2, \ldots, k\}$ be a numerical representation of the linguistic scale, where the higher number means "more preferable". The linguistic scale should consist of 5, 7, 9, 11, or 13 linguistic values. As noted by Herrera and Herrera-Viedma (Herrera & Herrera-Viedma, 2000) or Xu (Xu, 2012) scale must have sufficient granularity to distinguish the performances exhibited by the option under evaluation. An example of a 7-point scale is presented in Table 1.

Table 1. Linguistic terms for the options rating- 7-point linguistic scale

Numerical representation linguistic scale (rating)	**Linguistic term**	**Label**
1	Very poor	VP
2	Poor	P
3	Medium-poor	MP
4	Fair	F
5	Medium-good	MG
6	Good	G
7	Very-good	VG

Step 3. Determining the criteria weight w_j which describes the importance of each criterion C_j ($j = 1, \ldots, m$), in the evaluation of the alternatives, assuming that

$$w_1 + w_2 + \ldots + w_m = 1. \tag{3}$$

Step 4. Building the negotiation space on the OS scale.

The numerical representation of the linguistic scale OS can be considered as the evaluation options comprising the negotiation offers. Therefore, without loss of generality, the set of feasible offers from the negotiation space N may be represented by the Cartesian product of m sets OS, i.e.:

$$N = \prod_{j=1,\ldots,m} OS_j = \{[r_1, \ldots, r_m]: r_j \in OS_j, j = 1, \ldots, m\}, \text{where } OS_j = OS \text{ for every } j. \tag{4}$$

Let us note the negotiation space consists of the k^m offers, where m is the number of issues, and k is the number of linguistic terms from the OS scale.

In this way, the negotiation offer can be represented as a vector $O_i = [x_{i1}.x_{i2}, \ldots, x_{im}]$, where $x_{ij} \in OS$ ($i = 1, \ldots, k^m$; $j = 1,2, \ldots, m$).

Step 5. Determining the best and the worst offer.

The best offer, denoted as

$$O^+ = [k, k, \ldots, k], \tag{5}$$

is determined as a vector with maximum linguistic evaluation from the scale.

The worst offer, denoted as

$$O^{-} = [1,1,\dots,1], \tag{6}$$

is determined as a vector with minimum linguistic evaluation from the scale.

Step 6. Calculating distances $GDM2_i^{+}$ between the offer O_i and the best offer O^{+}.

The distances $GDM2_i^{+}$ between offer O_i (i= $1,2,\dots,k^m$) and the best offer O^{+} are calculated using GDM2 measure (see formula 1).

Step 7. Calculating the value of the LIBM measure for the O_j offer according to the formula:

$$\text{LIBM}_{\,i} = 1 - \frac{GDM2_i^{+}}{GDM2^{+-}} \tag{7}$$

where: $GDM2_i^{+}$ is the generalized distance measure between the i-th offer and the best offer, $GDM2_i^{+-}$ is the generalized distance measure between the best and the worst offer ($i = 1,2,\dots,k^m$).

The values of the LIBM measure are standardized in the range <0; 1>. The offers are ordered according to the LIBM values, where the higher the LIBM value, the higher the ranking of the offers.

Let us observe that the $GDM2_i^{+}$, $GDM2^{+-}$ can be calculated using the R package dist.GDM {clusterSim} (available at https://search.r-project.org/CRAN/refmans/clusterSim/html/dist.GDM.html).

Finally, the set $\{\text{LIBM}_{\,i}\}_{i=1,\dots,k^m}$ consists of the evaluation of all potential negotiation offers from the negotiation space.

Phase II: Assessing and prioritizing negotiation offers within an actual negotiation process.

Step 8. Selecting assessments of particular offers that form the basis for the actual negotiation process from the set of all offers.

Let $A = \{A_1, A_2, \dots, A_m\}$ be the set of negotiation offers involved in the actual negotiation process. It is easy to see that $A \subseteq N = \prod_{j=1,\dots,m} OS_j$. Thus, all the offers from set A were previously assessed in Step 7.

3. Conclusions and the Future Research

In summary, this paper presents a linguistic ideal-based multi-criteria (LIBM) method designed for the evaluation and ranking of negotiation offers. The main advantage of this method stems from its reliance on natural language, allowing for the incorporation of subjectivity, imprecision, and uncertainty in the assessment of negotiation offers. Our proposal serves as an alternative for linguistic evaluation of negotiation offers through the use of fuzzy sets. The approach is straightforward to implement in a negotiation support system and proves effective in various negotiation scenarios.

Furthermore, the advantages of the proposed approach include:

- Easy linguistic evaluation of offers, with the potential for additional graphical support, through the use of stars, or sliders.
- Eliminating the requirement to assign points when assessing issues and options within those issues, as required by methods like DR. While linguistic evaluation may lead to a less precise assessment, it can, on the other hand, prove to be a good solution for decision-makers with low cognitive demand. This approach assists negotiators in preference analysis by mitigating biases and heuristics during the assessment of negotiation offers (Kersten et al., 2017; Roszkowska & Wachowicz, 2015c).
- The negotiator may encounter challenges when assigning an exact numerical score to options within the criterion (Roszkowska & Wachowicz, 2014a, 2014b). This could be attributed to their low numerical

skills as well as difficulties in interpreting numerical ratings. Linguistic evaluation might prove to be easier for them. Although it may be less precise due to the adopted linguistic scale, the results of the evaluation of negotiation offers could be satisfactory. This method can be particularly useful when issues are described by a few numbers of options. Otherwise, it can be employed as a classification method first. Next, a more detailed evaluation can be conducted by comparing only offers from the same category.

- The possibility of building a comprehensive system using the R package R package dist.GDM {clusterSim}, which calculates distances on an ordinal scale. For example, if we have a 7-point scale and two criteria, it results in 7^2 offers, with three: 7^3 offers, etc.
- A simple and intuitive interpretation of the LIBM value as the distance from the offer to the best offer. This kind of comparison is natural in negotiation scenarios.
- Introducing a new offer does not impact previous assessments and will not lead to a rank reversal.

However, a drawback of this approach is its potential limitation in addressing problems with a limited number of options within the criteria, stemming from a constrained set of linguistic terms. One solution to this scenario is to classify continuous issues into distinct classes.

Further studies should, among other objectives, concentrate on verifying the efficacy of the proposed linguistic approach in supporting negotiations. We aim to apply this linguistic multi-criteria model to evaluate negotiation offers in Inspire negotiation problem, where preference information about issues and options is presented both verbally and graphically (Kersten & Noronha, 1999). It might be interesting to juxtapose our results with a negotiation scoring system that relies on the DR approach (Kersten et al., 2017, 2018; Wachowicz et al., 2019).

Acknowledgment: The contribution was supported by the grant WZ/WI-IIT/2/22 from Bialystok University of Technology and was founded by the Ministry of Education and Science.

References

1. Brzostowski, J., Roszkowska, E., & Wachowicz, T. (2012). Using an analytic hierarchy process to develop a scoring system for a set of continuous feasible alternatives in negotiation. *Operations Research and Decisions*, *22*(4), Article 4.
2. Filipowicz-Chomko, M., Mierzwiak, R., Nowak, M., Roszkowska, E., & Wachowicz, T. (2021). Reducing Cognitive Effort in Scoring Negotiation Space Using the Fuzzy Clustering Model. *Entropy*, *23*(6), Article 6. https://doi.org/10.3390/e23060752
3. Górecka, D., Roszkowska, E., & Wachowicz, T. (2016). The MARS Approach in the Verbal and Holistic Evaluation of the Negotiation Template. *Group Decision and Negotiation*, *25*(6), Article 6. https://doi.org/10.1007/s10726-016-9475-9
4. Hellwig, Z. (1968). Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasoby i strukturę wykwalifikowanych kadr [Application of the taxonomic method to the typological division of countries according to the level of their development and the resources and structure of qualified personnel]. *Przegląd Statystyczny*, *4*(1968), Article 1968.
5. Herrera, F., & Herrera-Viedma, E. (2000). Linguistic decision analysis: Steps for solving decision problems under linguistic information. *Fuzzy Sets and Systems*, *115*(1), Article 1. https://doi.org/10.1016/S0165-0114(99)00024-X
6. Herrera, F., Herrera-Viedma, E., & Martínez, L. (2008). A fuzzy linguistic methodology to deal with unbalanced linguistic term sets. *IEEE Transactions on Fuzzy Systems*, *16*(2), 354–370.
7. Kersten, G., & Noronha, S. J. (1999). WWW-based negotiation support: Design, implementation, and use. *Decision Support Systems*, *25*(2), Article 2.
8. Kersten, G., Roszkowska, E., & Wachowicz, T. (2017). The Heuristics and Biases in Using the Negotiation Support Systems. In M. Schoop & D. M. Kilgour (Eds.), *Group Decision and Negotiation. A Socio-Technical Perspective* (pp. 215–228). Springer International Publishing. https://doi.org/10.1007/978-3-319-63546-0_16
9. Kersten, G., Roszkowska, E., & Wachowicz, T. (2018). Representative Decision-Making and the Propensity to Use Round and Sharp Numbers in Preference Specification. In Y. Chen, G. E. Kersten, R. Vetschera, & H. Xu (Eds.), *Group Decision and Negotiation in an Uncertain World. GDN 2018* (Vol. 315, pp. 43–55). Springer.

10. Mustajoki, J., & Hämäläinen, R. P. (2000). Web-Hipre: Global Decision Support By Value Tree And AHP Analysis. *INFOR: Information Systems and Operational Research*, *38*(3), Article 3. https://doi.org/10.1080/03155986.2000.11732409
11. Piasecki, K., & Roszkowska, E. (2018). On application of ordered fuzzy numbers in ranking linguistically evaluated negotiation offers. *Advances in Fuzzy Systems*, *2018*.
12. Roszkowska, E., Filipowicz-Chomko, M., & Wachowicz, T. (2020). Using individual and common reference points to measure the performance of alternatives in multiple criteria evaluation. *Operations Research and Decisions*, *30*(3). https://doi.org/10.37190/ord200305
13. Roszkowska, E., & Wachowicz, T. (2012). Negotiation support with fuzzy topsis. *Group Decision and Negotiations 2012. Proceedings. Editoria Universitaria, Federal University of Pernambuco: Recife*, 161–174.
14. Roszkowska, E., & Wachowicz, T. (2014a). Defining Preferences and Reference Points – A Multiple Criteria Decision Making Experiment. In P. Zaraté, G. E. Kersten, & J. E. Hernández (Eds.), *Group Decision and Negotiation. A Process-Oriented View* (pp. 136–143). Springer International Publishing. https://doi.org/10.1007/978-3-319-07179-4_15
15. Roszkowska, E., & Wachowicz, T. (2014b). SAW-Based Rankings vs. Intrinsic Evaluations of the Negotiation Offers – An Experimental Study. In P. Zaraté, G. E. Kersten, & J. E. Hernández (Eds.), *Group Decision and Negotiation. A Process-Oriented View* (pp. 176–183). Springer International Publishing. https://doi.org/10.1007/978-3-319-07179-4_20
16. Roszkowska, E., & Wachowicz, T. (2015a). Application of fuzzy TOPSIS to scoring the negotiation offers in ill-structured negotiation problems. *European Journal of Operational Research*, *242*(3), Article 3. https://doi.org/10.1016/j.ejor.2014.10.050
17. Roszkowska, E., & Wachowicz, T. (2015b). *Holistic evaluation of the negotiation template – comparing MARS and GRIP approaches* (G. Kersten, B. Kamiński, M. Jakubczyk, & P. Szufel, Eds.). Warsaw School of Economics Press,.
18. Roszkowska, E., & Wachowicz, T. (2015c). Inaccuracy in Defining Preferences by the Electronic Negotiation System Users. In *Outlooks and Insights on Group Decision and Negotiation* (pp. 131–143). Springer International Publishing. https://link-1springer-1com-1q12c56ak0c5b.eczyt.bg.pw.edu.pl/chapter/10.1007/978-3-319-19515-5_11
19. Wachowicz, T. (2011). Application of TOPSIS methodology to the scoring of negotiation issues measured on the ordinal scale. *Multiple Criteria Decision Making*, *6*, 238–260.
20. Wachowicz, T., & Błaszczyk, P. (2013). TOPSIS based approach to scoring negotiating offers in negotiation support systems. *Group Decision and Negotiation*, *22*(6), Article 6.
21. Wachowicz, T., Kersten, G. E., & Roszkowska, E. (2019). How do I tell you what I want? Agent's interpretation of principal's preferences and its impact on understanding the negotiation process and outcomes. *Operational Research*, *19*(4), Article 4.
22. Wachowicz, T., & Roszkowska, E. (2021). Holistic Preferences and Prenegotiation Preparation. In D. M. Kilgour & C. Eden (Eds.), *Handbook of Group Decision and Negotiation* (pp. 255–289). Springer. https://doi.org/10.1007/978-3-030-12051-1_64-1
23. Wachowicz, T., & Roszkowska, E. (2022). Can holistic declaration of preferences improve a negotiation offer scoring system? *European Journal of Operational Research*, *299*(3), 1018–1032. https://doi.org/10.1016/j.ejor.2021.10.008
24. Walesiak, M. (1999). *Distance measure for ordinal data*. *2*(8), 167–173.
25. Xu, Z. (2012). *Linguistic decision making.Theory and Methods.* Springer.
26. Zhang, H. (2014). Linguistic Intuitionistic Fuzzy Sets and Application in MAGDM. *Journal of Applied Mathematics*, *2014*, e432092. https://doi.org/10.1155/2014/432092

The level of altruism for running social contracts efficiently

Dmitry Gimon[1] [0000-0003-4515-7282] and Alexandre Bevilacqua Leoneti[2] [0000-0002-0744-8866]

[1] Park University, Parkville MO, USA dmitry.gimon@park.edu
[2] University of São Paulo, Ribeirão Preto, Brazil ableoneti@usp.br

Abstract

Leoneti and Gimon (2023) investigated whether the inclusion of altruism and reciprocity's parameters within material utility functions would allow for the selection of enhanced social solutions. The authors have found that for some of the most known noncooperative games, some level of altruism can change the outcomes, then, providing a way of solving the equilibrium selection problem in 2×2 games. Additionally, it has been found that for some of those games, some level of altruism had changed the topology of the game, which results in changing from a scenario with Nash equilibria that are not Paretian into another where the Pareto solutions are in equilibrium. The results found lead to the interesting question: if a society concern with altruism in their social contract, which would be the efficient level of altruism the players should assume to each of possible social interaction? Considering the periodic table of Robinson & Goforth (2005) that contains the 144 most important noncooperative games, the paper aims to investigate the amount of altruism that is required to play each of those games for measuring the altruism necessary to play efficient social contracts, which would also be a strategy to run social contracts efficiently.

Keywords: utility functions; social choice theory; social contract; Nash equilibrium

1. Introduction

Societies usually struggle for achieving a virtuous set of norms that could efficiently sustain their associated set of individuals consistently. These norms, also called social contract, place particular emphasis on the establishment of parameters for guiding the conduct of individuals within the occasions of social interaction that can occur within that society. According to Binmore (1994) like traffic signals, those norms are coordinating devices.

Among the social contracts, there is the one that is known as TIT-FOR-TAT, which is usually seem as nice, retaliatory, forgiving, and clear from the view of economists, mathematicians, political scientists, psychologists, and a sociologist scientist. Axelrod (1980) demonstrated that, cooperate on the first move and then do whatever the other player did on the previous move, potentially leads to a good degree of cooperation and that the presence of "niceness" is one feature within the class of social contracts that better perform. This feature can be driven by a "voluntary desire of giving", also known as altruism, which may or may not change the outcomes of social interactions (Fehr et al. ,1999).

For illustrating the TIT-FOR-TAT strategy, let us take the Prisoner's dilemma game of Figure 1. Suppose that the players are operating under the TIT-FOR-TAT and that had been altruistically choosing the cooperative state up to a given point. Consequently, their payoff would be (3,3) for every interaction.

	A	B
A	3,3	0,4
B	4,0	2,2

Figure 1: Prisoner's dilemma game

Now, suppose that at some point one player does not act altruistically leading the game to either the state (4,0) or (0,4), depending which player has defected. Consequently, the next interaction would result in any payoff but (3,3), since the player that was defeated would not play altruistically at the new interaction. Assuming a probability to the player that had previously defected of 50% to play altruistically in the new interaction, the expected payoff of the player that defected would be 2 (the average of the two interactions, 4 when defected and 0 when cooperate while the counterpart does not cooperate) while the expected payoff of the previously defeated player would be 2 (0 when was defeated and 4 when does not cooperate while the counterpart cooperates), which would balance the social outcome.

On the other hand, it should be noticed that assuming that there is a chance of 50% to the player that had previously defected play altruistically again following a voluntary desire of giving is a strong assumption. In fact, given the level of intrinsic coherence that a social contract might reach, such norms could be misinterpreted as nature given, instead as a human artifact (Binmore, 1994). In other words, it is expected in the TIT-FOR-TAT rationale that the player that defected would naturally return to play altruistically as soon the counterpart retaliate its first defection movement, which is a strong assumption.

However, it could also be considered that the player that had previously defected could consider to choose to not play altruistically again. In this scenario, the expected payoff of the player that defected would be 3 (the average of the two interactions, 4 when defected and 2 when does not cooperate while the counterpart also does not cooperate), which would be an incentive to the player to not play altruistically given the payoff's enhancement. It should be noticed that, in such scenario, that strategy would also decrease the payoff's counterpart, since the expected payoff of the previously defeated player would be 1 (0 when was defeated and 2 when does not cooperate while the counterpart also does not cooperate).

The incentive to keep the strategy of defecting would lead, therefore, to the disbalance of payoffs among the players, which would put in risk their society. Therefore, a first implication is that a society should care about altruism because, if they do not, the social norm would be only to operate in Nash equilibrium in the game of Prisoner's Dilemma. Other implications are that, if the society do not care about altruism, the symmetry found in games such as the Battle of the Sexes would also not be broken, and, finally, non-cooperative games would not be possible to turn into cooperative ones.

In this sense, Leoneti and Gimon (2023) investigated whether the inclusion of altruism and reciprocity's parameters within material utility functions would allow for the selection of enhanced social solutions. Particularly, the authors have found that for some of the most known noncooperative games, some level of altruism had changed the outcomes, then, providing a way of solving the equilibrium selection problem in 2×2 games. Additionally, it has been found that for some of those games, some level of altruism had changed the topology of the game, which results in changing from a scenario with Nash equilibria that are not Paretian into another where the Pareto solutions are in equilibrium.

The results found lead to the interesting strategic question: if a society concern with altruism in their social contract, which would be the efficient level of altruism the players should assume to each of possible social interaction? Considering the periodic table of Robinson & Goforth (2005) that contains the 144 most important noncooperative games, the paper aims to investigate the amount of altruism that is required to play each of those games for measuring the altruism necessary to play efficient social contracts, which would also be a strategy to run social contracts efficiently.

2. Method

Considering a common structure of the 2×2 games in Figure 2, Gimon and Leoneti (2020) have proposed a reefing utility function that allows to incorporate altruism of each of the players.

	A	B
A	(a,b)	(c,d)
B	(e,f)	(g,h)

Figure 2: General structure of 2x2 two-player games

The idea behind this refinement is to provide an elastic trade-off between "me" (user's own payoffs) and "us" (users' join payoffs) using Cobb-Douglas structure, which has been proposed as

$$\Psi_i\left(x_i^k, H_i^{t-1}\right) = \pi_i\left(x_i^{jk}\right)^{(1-\omega_i(H_i^{t-1}))} . \bar{\pi}(x_i^k)^{\omega_i(H_i^{t-1})} \quad (1)$$

where $\omega_i^{\mathrm{t}}\left(H_i^{t-1}\right) = \alpha_i . H_i^{t-1} + \beta_i$, and α_i is the perceived level of player i on the reciprocal altruism of the remaining players, and β_i is the level of the player i unconditional altruism. It can be seen that the refined payoffs allow to incorporate socially-geared solutions into individual payoffs, which then can be used to describe the games' topologies, i.e. Nash equilibria and their symmetry. In that sense, the proposed refinement may provide solutions to games with no Nash Equilibria. It may also provide a mechanism to distinguish undistinguishable solution in original games, and to do so in justifiable by socially-geared behaviours way.

As it was mentioned above, incorporating altruism has the potential to creates new games structures, where each of the payoffs (a,b,c,d,e,f,g,h) is replaced with the corresponding refined payoffs (a',b',c',d',e',f',g',h') using the Equation 1. Due to refinement those new game may qualify for different topologies, or may retain the same topologies as the original games. The change in the topology then indicate the qualitative change of the game outcomes due to player(s) altruism and reciprocity.

In order to test the structures of 144 games in the periodic table of Robinson & Goforth (2005) it has been selected a set of settings for four level of altruism, which are depicted in Table 1.

Table 1: Altruism level settings

Altruism level	ω_i	Description
1	0	No altruism
2	0.3	Moderate altruism
3	0.7	High altruism (p2)
4	1	Max altruism (p2)

All possible combinations of four levels of altruism (ω_1 and ω_2) produced 16 settings including $\omega_1 = \omega_2 = 0$ (original game). Consequently, the results of this study included 2,304 reviewed scenarios. For each game we looked at the original outcomes (number of Nash equilibria, and the number of symmetry); then we analyzed whether increasing the level of altruism for one or both players result in changing the game structure leading to the change of the number of equilibria and/or breaking the symmetry of the original equilibria. Additionally, we have investigated weather the level of altruism would allow the selection of a social solution based on the the search of equilibria that are also Paretian.

3. Results and discussion

The results of the game simulations are presented in the Table 2. Out of 144 game topologies, 18 had no Nash equilibria, 108 had one, and 18 had two Nash equilibria, and out of these latter, 5 games had symmetrical Nash equilibria.

Table 2: Distribution of Nash equilibria counts in original games

No. of Nash equilibria	Number of games/topologies
0	5 games (18 topologies)
1	26 games (52 topologies) + 56 topologies without name
2	9 games (5 topologies with symmetry, 13 topologies without symmetry)

Games with no Nash equilibrium and the ones with symmetrical equilibria are the games of particular interest as those games do not have analytical solution. The topology of games could have change when altruism-incorporating refining was conducted providing such solution (i.e. either resulting in a game with a single Nash equilibrium, or in a game with multiple but non-symmetrical equilibria, from which a social outcome could be selected by tacking the Paretian equilibrium). Table 3 presents the results of minimum altruism values that resolve games with zero Nash equilibria.

Table 3: Games with zero Nash equilibria and the conditions for their resolutions

Game name	No. of topologies	Min level of altruism to achieve a solution (ω_1, ω_2)	
		Single Nash eq.	Two Nash eq. (no symmetry)
Cuban Missile Crisis	2	(0, 0.7) or (0.7, 0)	(0, 1) or (1, 0)
Cycle	10	(0, 0.7) or (0.7, 0)	(0, 1) or (1, 0)
Inspector-Evader	2	(0, 0.7) or (0.7, 0)	(0, 1) or (1, 0)
Pursuit of the Israelites	2	(0, 0.7) or (0.7, 0)	-
Total Conflict	2	-	(0, 1) or (1, 0)

For all five games the altruism levels were found. Three games (Cuban Missile Crisis, Cycle, and Inspector-Evader) require high level of altruism (0.7) of at least one player for the game to be transformed into a game with a single Nash equilibrium, and maximum level of altruism (1.0), of at least one player, for the game to be transformed into one with two (non-symmetrical) Nash equilibria.

In no simulated settings Pursuit of the Israelites game (with no Nash equilibria) had been transformed into a game with two equilibria, while high level of altruism of one player was sufficient for the game to be transformed into one with one equilibrium. No settings resulted in the Total Conflict game to be transformed into a game with single equilibrium, while two non-symmetric equilibria were a result of at least one player having maximum level of altruism. Note, that the transformed games may no longer meet the requirements of the original game's naming (for example, a Total Conflict with one fully altruistic side have a conflict nature). Additionally, actual threshold levels of altruism may differ from once we tested, and can be found analytically.

Following, Table 4 shows the results of the analysis for three games (5 topologies) with originally symmetrical solutions.

Table 4: Games with two symmetric Nash equilibria and the conditions for their resolutions

Game name	No. of topologies	Min level of altruism to break a symmetry (ω_1, ω_2)
Battle of Sexes	3	(0.3, 0) or (0, 0.3)
Chicken	1	(0.3, 0) or (0, 0.3)
Hero	1	(0.3, 0) or (0, 0.3

All three games require a minimum level of altruism (0.3 in the case of experiment) to differentiate two Nash equilibria (i.e. break the symmetry). Note, that the tested level of altruism does not represent an actual threshold for topology transformation. For instance, any level of altruism (as long as it is not equal among the players) will result in transforming those games to the games with asymmetrical Nash equilibria.

Let's us consider the example of Battle of Sexes game closer. In a case of symmetric original payoffs of cooperating strategies the choice between Nash Equilibria cannot be explained. However, refined payoffs allow for such choice (i.e. a strategy preferable by a less altruistic player becomes more desirable by both player due to the change of the payoffs for the more altruistic player. This situation has an interesting consequence observed in Gimon and Leoneti (2020) when a reciprocal altruism results in alternation of Nash equilibria in repetitive Battle of Sexes game, which means provides a sustainable approach to the choice of strategies that leads to long-term fairness of outcomes.Another example worth discussion is Chicken game that has been also used to describe Cold War scenarios (Russel, 1959). In that game one side (Soviet Union or USA) may avoid mutual destruction by collision on the expanse of reputational costs. The explanation of our model can be done by representing lower consideration of those reputational costs by a higher the level of altruism demonstrated by one side.

As it can be noticed in the results, the adjustment of the altruism's level is a fundamental strategy for playing social contracts efficiently. There are games, such as the Battle of the Sexes' game, which require a little level of altruism for breaking the symmetry among possible equilibria for providing a social outcome. In such situations, a perceived level of altruism is necessary for improving significantly the social welfare. On the other hand, there are games, such as Total Conflict, which require a great level of altruism for solving the game. It should be observed that particular game there is the necessity of one player to be fully altruistic for the game get two equilibria, from which a the Paretian one can be selected, due to the asymmetry among them.

4. Conclusions

By means of a numerical evaluation of 144 noncooperative games this study aimed to estimate the necessary amount of altruism for a society run its social contract. It has been observed that for a certain class of games a little level of altruism is suffice for breaking the symmetry found in games such as the Battle of the Sexes, which would be an important strategy for the society stands on its social contract. For other games, it a high level of altruism is necessary for the game to be solved using Paretian equilibrium, given the break of symmetry among the equilibria found. Future research should focus on the levels of altruism necessary for transforming the topology of a specific game into another, e.g., from a scenario of low possibilities of coordinating strategies into another with great opportunities for coordinating strategies.

References

Axelrod, R. (1980). Effective choice in the prisoner's dilemma. Journal of conflict resolution, 24(1), 3-25.

Binmore, K. G. (1994). Game theory and the social contract: just playing (Vol. 2). MIT press.

Gimon, D., & Leoneti, A. (2020). A refined utility function for modeling "me" vs."us" allo-cation in a sequential group decision process. EURO Journal on Decision Processes, 8(1-2), 1-12.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. The quarterly journal of economics, 114(3), 817-868

Leoneti, A. B.; Gimon, D. . The Effects of Altruism and Reciprocity on the Stability of Nash Equilibrium in Two-Player 2x2 Games. In: 23rd International Conference on Group Decision and Negotiation, 2023, Tokyo. Proceedings of 23rd International Conference on Group Decision and Negotiation, 2023.

Robinson D & Goforth D. 2005. The Topology of the 2×2 Games: A New Periodic Table. Routledge, New York, NY

A new approach for group decision making: a software combining the Strategic Choice Approach and the Analytic Hierarchy Process

Isabella M. Lami[1]

[1] Interuniversity Department of Regional and Urban Studies and Planning (DIST), Politecnico di Torino, Viale Mattioli 39, 20125 Turin, Italy.
isabella.lami@polito.it

Abstract

This research investigates how technology affords group decision making.
The presentation illustrates a new multi-methodological framework combining a Problem Structuring Methods (PSMs) and a Multi-Criteria Decision Analysis (MCDA) to address the problem of composing facts and values in the decision-making process of policymaking, recently proposed to the academic discussion with an article published on EJOR (Lami and Todella, 2023).

Keywords: Strategic Choice Approach; Analytic Hierarchy Process; new software, decision making, interaction

1. Introduction

The importance of problem structuring—and Problem Structuring Methods (PSMs) specifically—for Multi-Criteria Decision Analysis (MCDA) has been acknowledged in the literature and practices of the last decades.

This depends on the recognition that problem structuring is central to providing a richer view of the problematic situation for the subsequent phases of MCDA. In terms of the contributions that PSMs make to MCDA, problem structuring is indeed widely recognized as a helpful improvement, as it makes it possible to understand and articulate a problem's different values and perspectives and then identify alternative courses of action for dealing with it, exploring the uncertainties, and developing alternatives. However, while the literature has scrutinized the crucial advantages that PSMs could bring to MCDA, little attention has been directed to an opposite point of view, reflecting on any aspects or weaknesses shown by PSMs in the process that can be balanced through integration with MCDA. Many recent MCDA applications focus on broadly supporting different types of problems, with respect to: (i) choosing issues, in terms of selection among alternatives; (ii) sorting issues, in terms of classification of alternatives to categories; (iii) ranking issues, in terms of ordering of alternatives.

In this sense, MCDA for PSMs could allow for a better exploration of the alternatives themselves, facilitate communication, and support shared and transparent solution finding, using the involved stakeholders' preferences to deal with prioritization and evaluation.

2. Method

The new approach is composed by the combination of Strategic Choice Approach (SCA, Friend and Hickling, 2006) and Analytic Hierarchy Process (AHP, Saaty, 2008).

SCA enables the detection of relevant issues in the decision problems and their articulation in alternatives. At the same time, the integration with AHP allows the hierarchization of alternatives in an aggregated evaluation and, in so doing, discussion of the problem to be faced in a more transparent—and more structured—manner.

This approach relies on a new software called MuVAM (Multi-Values Appraisal Methodology), which is the focus of this presentation. MuVAM, designed by the author and DEM future, a software house, is a web application designed to support decision-making processes related to complex problems, equipped with a very effective online graphical interface. MuVAM works with a series of algorithms in the background to guide participants through five steps aimed at: representing the problem, identifying possible solutions, identifying decision criteria and their weights, comparing solutions and illustrating results. Lastly, it can be used by groups working in real time, in presence or remotely, or in asynchronous mode, allowing them to participate in the same workshop from anywhere, even deferred.

The software has been used in several structured workshops held in various European universities (in Turin, Paris, Bratislava, Tirana), in master's degree courses, doctoral courses and master's courses for lifelong learning. The interaction within the groups during the workshops, aimed at a reflection on the creation of real collaboration also supported by the software, was analysed essentially through the reports written by the participants (who answered a series of questions provided by the lecturers) and through reports written by an observer within each group. The questions asked to the participants mainly concerned three areas: i) the influence of the methodology on group dynamics (consensus building, dissent management, interaction between individual and group); ii) the performativity of the software; iii) timing and process. The observer's report, on the other hand, was not guided by precise questions but was more related to what could be an ethnographic observation: the observer had to shadow an individual group and transcribe the group's interactions, from literal quotations of certain group’s exchanges to the annotation of specific behaviors. The presentation illustrates some of these applications to show the potential of MuVAM to contribute to a real collaborative effort.

References

J. Friend, A. Hickling, Planning under pressure, *Routledge*, 2012.

I.M. Lami, E. Todella, «A multi-methodological combination of the strategic choice approach and the analytic network process: From facts to values and vice versa», *European Journal of Operational Research*, Volume 307, Issue 2, pp. 802-812, 2023

T. Saaty, Decision making with the analytic hierarchy process, *International journal of services sciences,* vol. 1, pp. 83-98, 2008.

Multicriteria Group Decision Model for Land Regularization in Brazil

João Batista Sarmento dos Santos-Neto[1*], Carolina Lino Martins[1], Kassia Tonheiro Rodrigues[1], Nadya Kalache[1] and Saulo Gomes Moreira[1]

[1] Federal University of Mato Grosso do Sul (UFMS), Campo Grande, Brazil
[*] joao.sarmento@ufms.br

Abstract

The agrarian reform began to be discussed by the government in Brazil in the late 1950s. Currently, the National Agrarian Reform Program (PNRA) seeks to enable access to land by small farmers by distributing public lands divided into areas. In this context, it is necessary within the scope of public policies for territorial and land governance to use models to define priorities in the occupation of the land in the country. Therefore, this study proposes a multicriteria group decision model to rank the candidate families interested in joining the PNRA to select beneficiaries for public policies aligned with the federal government's strategic social, environmental, and economic goals. For this, we applied the Multicriteria method PROMETHEE II for ranking the beneficiaries using the Group Decision Support System (GDSS) function of the Visual PROMETHEE software. After this, we analyzed the results and performed a sensitivity analysis.

Keywords: Multicriteria; Agrarian Reform; PROMETHEE; Group Decision; Land Tenure.

1. Introduction

Land regularization is the term used to describe government initiatives aimed at granting land tenure and acknowledging property rights to people residing in unofficial settlements. To enhance living conditions and provide access to basic services for informal lands, it is advised that these measures be incorporated into rural upgrading plans (Araújo et al., 2022). Acquiring informal land renders the development of cities and their residents unfeasible. According to Childress et.al (2021), this practice imposes high burdens on citizens and costs on the government, such as illegal access to land tenure, and lack of government services related to infrastructure, in addition to indirect costs on social programs aimed at health, education, housing, and criminal violence. The government uses land regularization as an alternative to minimize the negative consequences arising from irregular settlements due to these factors.

In Brazil, the National Institute for Colonization and Agrarian Reform (INCRA) is responsible for all policies related to land regularization. One aspect of the Brazilian land issue is the prioritization of beneficiaries for land occupation and use. In 2022, the Federal Court of Auditors (TCU) published a report that identified the main risks associated with public administration that undermine the quality of government services and the effectiveness of policies promoted by public authorities. The report highlights the lack of integration among government institutions and the complexity and fragmentation of the current laws as the reasons for the limited knowledge about land occupation and land use capacity in Brazil. Also on this list, the TCU reports that evidence of irregularities has been identified in the criteria for selecting families within the National Agrarian Reform Program (PNRA), with numerous financial losses for the country.

Multicriteria Group Decision Models enable this type of analysis by incorporating multiple decision-makers (DM's) and their diverse perspectives on a given problem. Self-awareness is crucial for effective group decision-making, as each individual brings unique strengths and weaknesses to the table. Seghezzo et al. (2016)

conducted a study on land tenure and use conflicts in the Chaco region of Salta, Argentina. The study utilized a participatory and multicriteria evaluation model that identified several variables that directly or indirectly explained a significant proportion of the conflicts. The findings of the study are supported by the literature. Nyeko (2012) reports that multicriteria decision-making based on Geographic Information Systems (GIS) has a high potential for application in land use planning. However, there is a gap in the literature regarding studies aimed at planning the suitability of land use on rural land. Gebre et al. (2021) note that there are few studies in the area of intelligent land use planning. Therefore, research is needed to focus on the role of rural land allocation.

Thus, this study proposes a multicriteria group decision model for ranking beneficiaries of the Brazilian Agrarian Reform Program. The model is aligned with the federal government's strategic social, environmental, and economic goals. To achieve this, the study uses the PROMETHEE II method for group decision, since this strategy provides a complete analysis, through variations in the model's parameters, resulting from different scenarios and preferences. According to Maria-Paraskevi et al. (2022), the method is highly adaptable due to its simplicity, which satisfies DM’s. This research contributes by demonstrating how DM’s influence the ranking compared to the existing methodology adopted by INCRA. The ranking is based on established scenarios, as DM’s are directed to choose between 'preference' or 'indifference' among alternatives. This method remains clear, simple, and stable, as noted by Lorena et al. (2022).

2. A group decision-making model for ranking the beneficiaries of the land regularization program

Group decision-making is the process of converging diverse individual preferences into a single collective preference, aiming to achieve a consensus of preferences coming from a set of DM’s in an organization (Causil and Morais, 2021). For the context of this study, the PROMETHEE II method (Preference Ranking Organization Method for Enrichment Evaluation) and the PROMETHEE GDSS are used, as they are easy-to-understand methods for DM’s, and to allow obtain individual and collective classifications considering multiple objectives, respectively. In addition, the ROC (Rank Order Centroid) method is implemented to estimate the value of the weights.

In Brazil, to benefit as a settler, the rural worker needs to participate in a selection process, through public notice, which evaluates the set of candidates with scores measured by parameters defined by INCRA. To incorporate the preferences of DM’s in the selection process, a decision model was proposed and later tested, in six steps, as shown in Figure 1.

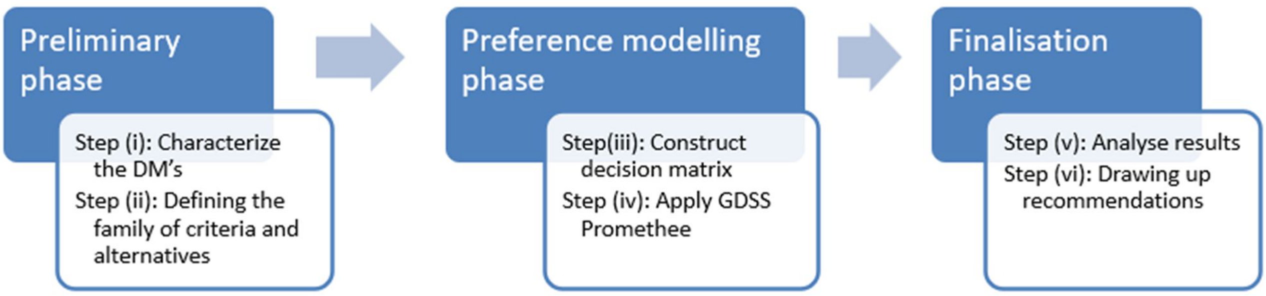


Figure 1: proposed model for ranking beneficiaries.

Step 1. Characterize the DM’s: The DM is the person responsible for the decision and for establishing relationships, constraints, preferences, and judgment of values that influence the decision process (YOSHIURA et al., 2023). At this stage, an interdisciplinary team is selected as responsible for analyzing the problem considering the multiple objectives of the context.

Step 2. Defining the family of criteria and alternatives: The definition of the family of criteria can occur from the structuring of the problem together with the DM’s. In this case, it was decided to use in the model the same criteria proposed by INCRA, which were validated with the DM's and will be presented in section 3.

Step 3. Construct Decision Matrix: To construct the decision matrix we simulated data for an intra-criterion evaluation. For this, we generate random data for a uniform distribution.

Step 4. Apply GDSS PROMETHEE: To apply the GDSS PROMETHEE, we used the GDSS function of the Visual Promethee software. Visual PROMETHEE is a computational tool that allows exploring numerical and graphical resources.

Step 5. Analyse Results: In this step, to analyze the results, we did a sensitivity analysis. For this, the model parameters must be changed to observe possible variations in the model outcomes. We performed a sensitivity analysis modifying the criteria weight values in a range of 20%.

Step 6. Drawing up recommendations: In step 6 we recommended the final ranking of beneficiaries.

3. Real application

A real application was developed to evaluate the applicability of the proposed model. In the preliminary phase, we characterized the decision makers (DM) and defined the evaluation criteria and the set of alternatives for the decision problem. In this phase, an interdisciplinary team was responsible for analyzing the problem: a politician who usually works with family farming issues (DM1), a director of a public agrarian support agency (DM2), a researcher in the field of family farming with several publications in specialized journals (DM3), program beneficiaries (DM4) and, as a supra decision maker (SDM), the regional superintendent of INCRA was selected.

To define the set of criteria, we validated with the DM's the criteria adopted by INCRA in the process of beneficiary evaluation. Table 1 describes the seven criteria used to evaluate the alternatives. To test this model, the set of alternatives was defined as a hypothetical set of 10 beneficiaries.

Table 1. Criteria definition

Criteria	**Definition**	**Aim**	**Range**
I - Family Size and Workforce	Size of the family and its workforce.	Maximization	0-15
II – Length of Residence in the Municipality	Criterion applied to a family unit that has resided for a longer time in the municipality in which the settlement is located.	Maximization	0-15
III – Female-Headed Family	Family in which the woman is responsible for most of the income.	Maximization	0 or 1
IV – Family or Individual Camp Member	Family unit or individual member of a camp located in the Municipality in which the settlement is located.	Maximization	0 or 1
V - Family of Aggregated Rural Workers	Family unit of a rural worker who resides in the property intended for the settlement project.	Maximization	0 or 1
VI - Time in Agricultural Activity	The criterion applied to the proven time of exercise of agrarian activity by the family unit.	Maximization	0-15
VII – Monthly Family Income	The criterion applied to monthly family income, graded under the terms declared in the Unified Registry for Social Programs of the Brazilian Government.	Minimization	0-5

Next, in the preference modeling phase, we simulated data for an intra-criterion evaluation to construct the decision matrix. To do this, we generated random data for a uniform distribution. In step 4, we used the PROMETHEE II multi-criteria method to rank the beneficiaries. For this, we used the GDSS function of the Visual PROMETHEE software. Table 2 shows the performance of each attribute for each outcome in the decision matrix.

Table 2. Decision matrix

Criteria	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
I	13,70	12,60	11,20	9,70	7,00	10,00	9,90	12,80	8,50	8,50
II	7,00	7,00	6,00	2,00	7,00	4,00	12,00	1,00	0,00	8,00
III	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00
IV	0,00	1,00	1,00	1,00	1,00	0,00	0,00	0,00	1,00	1,00
V	1,00	1,00	0,00	1,00	1,00	1,00	1,00	0,00	0,00	1,00
VI	12,00	1,00	5,00	10,00	5,00	15,00	4,00	4,00	10,00	14,00
VII	4,00	3,00	3,00	5,00	2,00	5,00	4,00	5,00	1,00	2,00

For the PROMETHEE application, the criteria weight values and preference functions were different for each DM. The Rank-Order Centroid (ROC) method was used to estimate the criteria weight values, which only provides a ranking of the attributes by the DM's. ROC is an important method for the elicitation procedure that directly consists of ordering from the most important to the least important objectives or criteria of the decision for assigning criteria weights (CAUSIL and MORAIS, 2021). To define the preference or indifference thresholds (p and q, respectively), some functions are required that have been established with the DM's with the support of Visual PROMETHEE. As an example, Table 3 shows the parameters used to apply the model for decision-maker 1. The same parameters were set for the other DM's in the problem.

Table 3. Weight, Objectives, and Preference Functions by DM1

Criteria	Weight	Aim	Preference function	Parameters
I	0,07	Maximization	Linear	P = 2,00; Q = 1,00
II	0,11	Maximization	Usual	Not applicable
III	0,04	Maximization	Usual	Not applicable
IV	0,16	Maximization	Usual	Not applicable
V	0,23	Maximization	Linear	P = 2,00; Q = 1,00
VI	0,37	Maximization	Usual	Not applicable
VII	0,02	Minimization	Linear	P = 2,00; Q = 1,00

Figures 2, 3, 4, and 5 illustrate the individual ranking of beneficiaries based on PROMETHEE II for decision-maker preferences. There is a clear similarity in behavior between DM's 1 and 2 and between DM's 3 and 4. This similarity is explained by the way the DM's ranked the criteria. For example, candidate 10 performed well in the criteria "time spent in agricultural activity" and "family or individual living in a settlement", which have higher weight values for DM's 1 and 2. Conversely, candidate 1 performed very well on the criteria "monthly family income" and "family size and labor force", which have the highest weight for decision makers 3 and 4.

Rank	action	Phi	Phi+	Phi-
1	Candidato 10	0,4402	0,5048	0,0646
2	Candidato 1	0,2945	0,4530	0,1585
3	Candidato 6	0,2831	0,4333	0,1503
4	Candidato 4	0,0915	0,3276	0,2362
5	Candidato 7	-0,0731	0,2114	0,2845
6	Candidato 5	-0,1294	0,1765	0,3059
7	Candidato 9	-0,1306	0,2470	0,3776
8	Candidato 2	-0,2127	0,1894	0,4021
9	Candidato 3	-0,2164	0,1611	0,3775
10	Candidato 8	-0,3469	0,1179	0,4649

Figure 2: Individual ranking by decision-maker 1

Rank	action	Phi	Phi+	Phi-
1	Candidato 4	0,1478	0,2579	0,1101
2	Candidato 10	0,1449	0,2659	0,1210
3	Candidato 1	0,0959	0,2696	0,1737
4	Candidato 6	0,0712	0,2469	0,1756
5	Candidato 3	0,0348	0,2031	0,1683
6	Candidato 2	0,0128	0,2059	0,1931
7	Candidato 9	-0,0270	0,1877	0,2147
8	Candidato 5	-0,0838	0,1513	0,2351
9	Candidato 7	-0,1676	0,1118	0,2794
10	Candidato 8	-0,2289	0,1466	0,3755

Figure 3: Individual ranking by decision-maker 2

Rank	action	Phi	Phi+	Phi-
1	Candidato 1	0,3754	0,3981	0,0227
2	Candidato 6	0,2595	0,3688	0,1093
3	Candidato 4	0,2074	0,3240	0,1166
4	Candidato 8	0,1979	0,3769	0,1789
5	Candidato 7	0,0742	0,2225	0,1483
6	Candidato 2	0,0481	0,2531	0,2050
7	Candidato 3	-0,1184	0,1698	0,2882
8	Candidato 10	-0,1988	0,1382	0,3370
9	Candidato 5	-0,4023	0,0547	0,4570
10	Candidato 9	-0,4432	0,0661	0,5093

Figure 4: Individual ranking by decision-maker 3

Rank	action	Phi	Phi+	Phi-
1	Candidato 1	0,3627	0,4175	0,0548
2	Candidato 7	0,2266	0,3479	0,1213
3	Candidato 8	0,1514	0,3836	0,2322
4	Candidato 2	0,1454	0,3178	0,1724
5	Candidato 6	0,1389	0,3482	0,2093
6	Candidato 4	0,0985	0,3152	0,2167
7	Candidato 3	-0,0534	0,2349	0,2883
8	Candidato 10	-0,1730	0,1802	0,3532
9	Candidato 5	-0,3261	0,1193	0,4454
10	Candidato 9	-0,5709	0,0426	0,6135

Figure 5: Individual ranking by decision-maker 4

Finally, in the finalization phase, the sensitivity analysis was performed to verify the robustness of the method. For this purpose, adjustments were made to the parameters to examine possible variations in the model results. Specifically, the sensitivity analysis was performed by varying the weight values of the criteria within a 20% range. Figure 6 shows the sensitivity of the ranking to changes in the weights given to the DM's, rather than to the criteria. No significant difference in results was observed for any of the DM's.

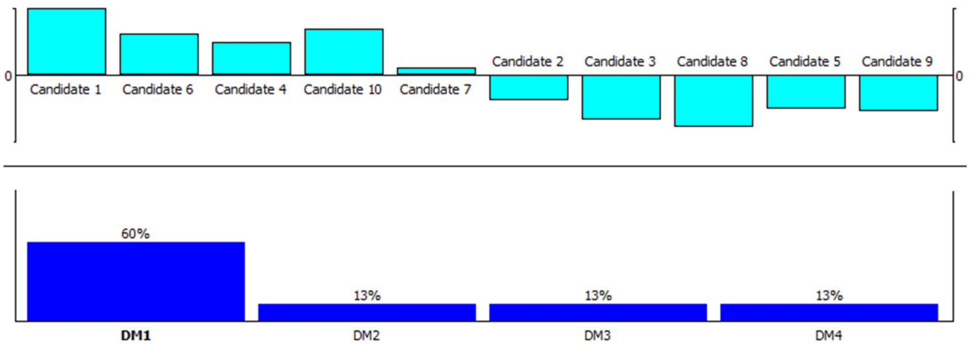


Figure 6: Sensitivity analysis for decision-maker 1

Figure 7 shows the final collective ranking of the candidates. It can be seen that Candidate 1 achieved the best performance, due to its outstanding individual results across all DM's. The applicability of the group decision model is evident, as it allows the determination of the alternative with the best outcome by considering the individual candidate performances for each decision maker and taking into account the final decision of the SDM. In the group decision model, the SDM does not make the final decision, but is responsible for group consensus achievement. In this context, the primary role of the SDM was to provide an aggregation perspective, ensuring results that comprehensively incorporate the viewpoints of all decision makers. The SDM was responsible for aggregate the obtained result, drawing upon their expertise and holistic understanding of the subject matter.

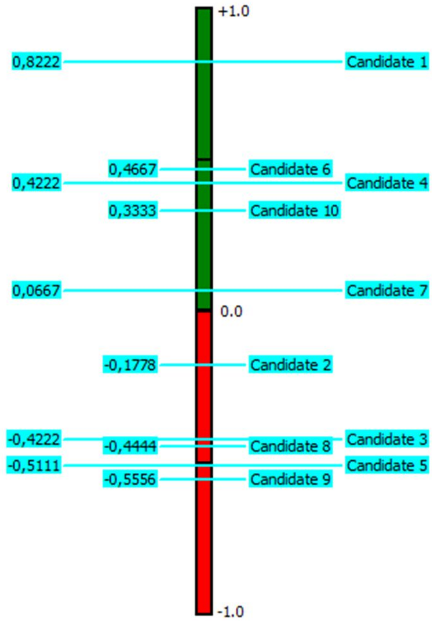


Figure 7: Collective ranking by the SDM

The proposed framework integrates theoretical and practical considerations and provides a streamlined approach to complex decision-making in agricultural land distribution. By addressing the challenges of limited resource allocation and social equity, the model aims to improve the efficiency of agrarian reform project management.

This contribution to the literature provides a practical tool for agrarian reform project managers that promotes transparent and data-driven decision-making. The multi-criteria approach streamlines family identification and selection, promoting efficiency in different contexts. Management implications include resource optimization, conflict minimization, and equitable opportunity distribution, in line with the objectives of the National Agrarian Reform Program (PNRA). The model facilitates the selection of beneficiaries based on defined criteria, ensuring efficient allocation of resources to the PNRA, and targeting those in need with potential for productive land utilization. Moreover, through the implementation of objective and transparent criteria, it can mitigate disputes and conflicts associated with land distribution and agricultural resources. Consequently, it advances a more equitable distribution of opportunities for land and agricultural resource access, taking into account variables such as social and occupational background, family structure, and engagement in agricultural pursuits, among others.

INCRA has a systematized scoring regulation for selecting beneficiaries of the PNRA, based on criteria that have also been incorporated into the model presented in this study. Despite having well-defined criteria and scores, this method does not incorporate the perspectives of DM's in its modeling. As a result, decisions that are not suitable may be made, and considered suboptimal solutions. Additionally, DM's may not support the results obtained, leading to resistance to implementing actions and a lack of commitment to results. The incorporation of DM preferences enhances the robustness of the proposed model, which outperforms its predecessor by incorporating multiple perspectives. This inclusive approach allows for a more comprehensive representation of the complexities of selecting families for agrarian settlements, strengthening the model's adaptability to nuances and multifaceted considerations. The resulting model aligns with PNRA's goals, remains faithful to the DM's preferences, and addresses specific issues.

4. Conclusions

Through this study, we proposed and tested a useful multicriteria group decision model that could be used for the Brazilian Agrarian Reform Program. A multicriteria model was proposed and applied in the selection of beneficiary families that will occupy landholdings in settlement projects destined for the PNRA. This methodology reinforced the importance of considering the different perspectives of the DM's since their preferences regarding the criteria served as a basis for determining individual rankings. As main contributions, we innovated in defining a model with 6 steps for ranking beneficiaries to land location, taking into account the preferences of a group of DM's using the PROMETHEE II and GDSS PROMETHEE implemented with ROC.

Acknowledgments

This study was financed by the Ministry of Agriculture, Livestock and Supply (MAPA), the National Institute of Agrarian Reform (INCRA) and the Federal University of Mato Grosso do Sul (UFMS).

References

Araújo, R. P. Z. de, Moura, A. C. M., & Nogueira, T. D. A. (2022). *Creating Collaborative Environments for the Development of Slum Upgrading and Illegal Settlement Regularization Plans in Belo Horizonte, Brazil* (pp. 86–112). https://doi.org/10.4018/978-1-7998-9090-4.ch005

Causil, O. M., & Morais, D. C. (2021, December). A Multicriteria Group Decision-making Model for Selecting a Perishable Food Packaging System Using an Outranking Method. In *2021 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)* (pp. 467-471). IEEE.

Childress, M., Carter, S., & Barki, E. (2021). Fit-for-Purpose, Private-Sector Led Land Regularization and Financing of Informal Settlements in Brazil. *Land,* 10(8), 797. https://doi.org/10.3390/land10080797

Gebre, S. L., Cattrysse, D., Alemayehu, E., & Van Orshoven, J. (2021). Multi-criteria decision making methods to address rural land allocation problems: A systematic review. *International Soil and Water Conservation Research*, 9(4), 490-501.

Lorena, A. L. F., Costa, P. C. S., Lima, E. S., Lima, T.B. ERM-POP Model: improving government initiatives towards Enterprise Risk Management implementation. ICDSST 2022 Proceedings – online version the EWG-DSS *2022 International Conference on Decision Support System Technology*. p. 127. Maio, 2022.

Maria-Paraskevi, B., Jason, P., Aikaterini, G. and George, A. Multi-criteria Decision making for a green and more sustainable nanotechnology and nanoproducts. ICDSST 2022 Proceedings – online version the EWG-DSS *2022 International Conference on Decision Support System Technology*, pp. 80, 2022.

Nyeko, M. (2012). GIS and multi-criteria decision analysis for land use resource planning. *Journal of Geographic Information System*, 4(4). https://doi.org/10.4236/jgis.2012.44039

Seghezzo, L., Venencia, C., Buliubasich, E. C., Iribarnegaray, M. A., & Volante, J. N. (2017). Participatory, multi-criteria evaluation methods as a means to increase the legitimacy and sustainability of land use planning processes. The case of the Chaco region in Salta, Argentina. *Environmental management*, 59, 307-324.

Yoshiura, L. J. M., Martins, C. L., Costa, A. P. C. S., & Santos-Neto, J. B. S. D. (2023). A multicriteria decision model for risk management maturity evaluation. *Pesquisa Operacional*, 43, e270186.

joint conference
group and negotiation DECISION support systems technology
Porto 2024

24th International Conference on Group Decision and Negotiation &
10th International Conference on Decision Support System Technology
Human-Centric Decision and Negotiation Support for Societal Transitions

Building a cardinal scoring system for alternatives evaluated by various experts within disjoint subsets

Lihi Dery[1*], and Tomasz Wachowicz[2]

[1] Ariel University, Ariel, Israel
[2] University of Economics in Katowice, Poland
[*] nlihin@gmail.com

Abstract

In certain situations, distinct groups of decision-makers are assigned the task of evaluating different sets of alternatives. Consequently, these disparate decision-making groups assess disjoint sets of alternatives, and the challenge arises to aggregate these assessments effectively. A common example of such a scenario is in peer assessment settings within educational environments, where students are divided into groups, and each group is responsible for assessing the work of its members. To address this challenge, we build upon the ranking-to-rating (R2R) approach and introduce a method that incorporates an integrated scale. This scale combines original ratings with fractional Borda counts, thereby enhancing the detail and range of the evaluations. This methodology not only tackles the inherent subjectivity and biases present in evaluations but also provides a mechanism to equalize integrated ratings. This is particularly useful in instances where significant differences arise due to the application of fractional Borda counts, ensuring a more balanced and fair aggregation of assessments from different decision-making groups.

Keywords: peer-assessment; disjoint group decisions; integrated evaluation scales

1. Introduction

There are situations in which some alternatives are assessed by a group of decision-makers (DMs) while a disjoint group of DMs assesses other alternatives. Then, the goal is to construct a unified rating of all the alternatives. An example of this could be the evaluation process of candidates for a job or a Ph.D. Some candidates could be evaluated by a committee on one day, and other candidates could be evaluated by another committee on another day. Another example is a peer assessment activity, in which students evaluate each other's works. Distinct groups of students are formed, and the peer assessment takes place for each member of the group, by the members of that group. In such problems, the issue of comparing assessments assigned by different DMs arises, even if they use a unified rating scale. This is because they may differ in behavioral traits, leading to different interpretations of the scale and its (explicitly or implicitly) associated linguistic labels (Silva & Morais, 2014; Wachowicz & Roszkowska, 2023). For instance, some DMs may consider alternatives to be very good and assign them ratings from 80 to 90 points on [0; 100]-range scale, reserving a score of 100 for potentially superior offers that may potentially emerge in the future or similarly avoiding the use of extreme linguistic labels in line with the central tendency bias phenomenon (Pimentel, 2019). However, some others might be less restrictive and rate qualitatively similar offers with a total 100-point rating, considering them to meet all established evaluation criteria fully. This leads to an evident problem where constructing a joint list based on opinions of decision-makers interpreting the same scales differently could disadvantage alternatives assessed by those inclined towards central tendency bias compared to those who have no issue using extreme ratings. Hence, there is a need to develop a mechanism for comparing assessments by different decision-makers, which can be designed differently depending on the type of problem and the decision-making context under consideration.

In this paper, we address a specific decision-making problem, the context of which involves peer assessment of projects presented by students in their academic courses. These projects are evaluated by students themselves within academic groups, as involving students in the assessment process of their own and their peers' achievements is recognized as a way to enhance engagement, improve outcomes, and build additional skills (Adesina et al., 2023; Double et al., 2020). The results of peer reviews are expected to contribute to the final grades the students obtain from the course; therefore, they should be comparable between groups to ensure that all good projects are equally well rewarded by the teacher, no matter from which group they come from and how much biased the students within each group could be while granting their scores to the projects they evaluated. To compare the projects between groups, we employ an approach based on the notion of meta-DM, which evaluates selected projects from different groups. For example, when evaluating students using peer-assessment, a meta-DM would be the course instructor. Such an approach allows for adequately rescaling within-group ratings resulting from student evaluations and merging them into a joint evaluation scale. The student's evaluation of projects is performed using the ranking-to-rating approach (R2R) proposed earlier by Dery (2023) for within-group evaluation of projects. Its advantage is that it reduces the number of potential ties, often observed in student peer review problems (Panadero et al., 2013), compared to alternative techniques of tie-breaking (Fabre, 2021). However, using the original R2R approach for between-group evaluations requires designing a procedure for determining the cardinal ratings from within-group evaluations. Additionally, an approach for between-group cardinal ratings comparison for meta-DM needs to be constructed that will allow merge within-group results into one ranking list. These two latter elements are the original contributions proposed within this manuscript.

2. Determining cardinal ratings from R2R within-group evaluations

Dery (2023) has proposed an evaluation schema that allows building ranked ratings for alternatives under evaluation, named Ranking-to-Rating (R2R) (see Dery, 2023). In her problem, she considers a group of n DMs (voters), $V = \{v_1, v_2, \dots, v_n\}$, evaluating a set of m alternatives (projects), $C = \{c_1, c_2, \dots, c_m\}$. The R2R evaluation mechanism amounts to determining the Ranked Rating Set ψ_i for i-th DM (originally named voter, and here – a student), which consists of a list of following pairs

$$\psi_i = \{(s_{i,1}, p_{i,1}), \dots, (s_{i,m}, p_{i,m})\}, \tag{1}$$

each describing the scores ($s_{i,j}$) and rank position ($p_{i,j}$) assign to alternative c_j by DM v_i. It is suggested to use coarse scales (such as 5- or 7-point, e.g., $D = \{1,2, \dots, 7\}$) to declare $s_{i,j}$ values. In such a situation, one may easily assign accompanying linguistic etiquettes to subsequent levels of such a scale, which makes DM evaluation cognitively easier.

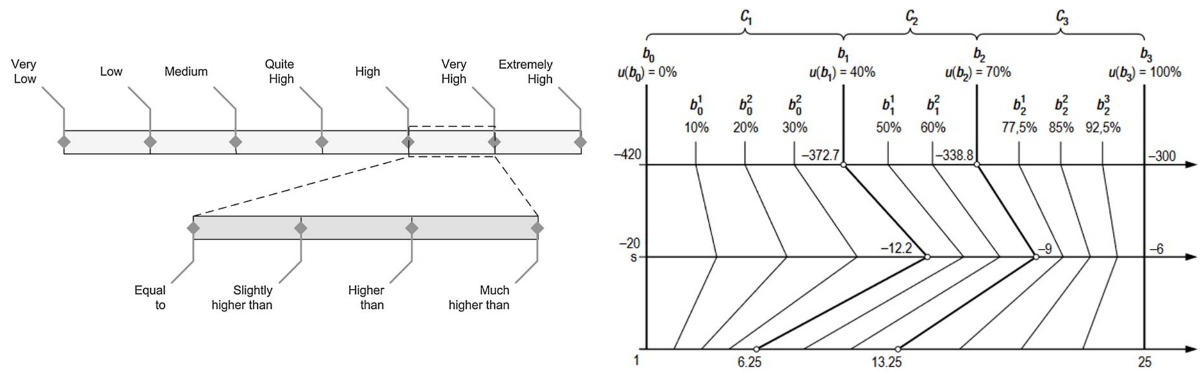
Ranked rating sets of all DMs comprise the group preference information regarding the alternatives $\Psi = \{\psi_i\}_{i=1,\dots,n}$. The group rank rating sets Ψ is then used in an R2R aggregation phase to determine a compromise rank order of alternatives. First, the median rank $\tilde{p}_j$ is determined for each alternative c_j out of all individual $p_{i,j}$. Fine-tuning procedures are introduced to break the ties if some alternatives share the same rank. One of them uses the Borda count, i.e.,

$$B_j = \sum_{i=1}^{n} b_{i,j} \tag{2}$$

where $b_{i,j}$ is the number of alternatives that c_j outperforms according to its rank position $p_{i,j}$ in ψ_i. Thus, if two alternatives have the same median evaluation score, the one with a higher Borda count value will be considered better. Other voting rules beside Borda can be used.

Even though R2R allows for breaking ties in many situations and determining the total order of alternatives, it cannot be directly applied when comparing performances of alternatives from different groups. It is because the Borda-based fine-tuning mechanism does not modify the cardinal ratings $s_{i,j}$ but only provides additional information to be used as a second element in construction lexical order, unfortunately, dependent in value on the size of the set of alternatives C. Therefore, modifications in Dery's original approach may be required to produce more granular cardinal ratings out of the R2R evaluation process that will be easier for cross-group comparisons.

From the technical viewpoint, the problem we are dealing with is to refine the cardinal scale, as in the case of the integrated rating scale proposed earlier by Brzostowski and Wachowicz (2014) for evaluating negotiation alternatives in the NegoManage system (see Figure 1a) or the cardinal rating of alternatives classified into selected quality categories, proposed by Wachowicz (2010) as part of calibrating the results of the ELECTRE-TRI sorting method (see Figure 1b). As for coarse scales, individual resolution levels can be perceived as certain quality categories; therefore, we will use these terms interchangeably in the subsequent part of the work.



(a) (b)

Figure 1: Integrated scales for extending scoring precision by using higher- and lower-level quality categories with accompanying ratings for the classic direct rating (a) and ELECTRE-TRI (b)

We need to calibrate the resolution levels of our coarse scale ($s_{i,j}$) – an equivalent of a higher level of evaluation (Fig. 1a) or a category of quality C_k (Fig. 1b) – by additional information derived from B_j – an equivalent of a lover level of evaluation (Fig. 1a) or subprofiles b_k^l (Fig. 1b). As the calibration should be independent of the size of the set C, we propose to use a modified Borda count normalized by the size of this set. Doing so, we will express the information about the fraction of better and worse alternatives in percentage terms. Therefore, we replace formula (2) by

$$B_j^{\%} = \frac{B_j}{(m-1)n}, \tag{3}$$

where in the denominator, we use $(m-1)$ because when comparing the number of situations where alternatives are surpassed, we never compare an alternative to itself.

Fractional Borda count may be used to determine the alternatives evaluation on the new integrated scale built out of a series of $\left(s_{i,j}, p_{i,j}\right)$ values using a routine similar to the original R2R approach. In this routine, we will use the median values $\tilde{p}_j$ calculated based on the original R2R idea as indicators of the higher level scale, and fractional Borda counts $B_j^{\%}$ as an indicator of a lower-level scale. As a result, each alternative c_j is assigned a score $\tilde{p}_j$, adjusted by the proportion of c_j's quality resulting from comparing its $B_j^{\%}$ to other alternatives with the same median value as $\tilde{p}_j$ and $B_{j'}^{\%}$ of alternative $c_{j'}$ considered best within those having a median equal to $\tilde{p}_j - 1$ (i.e., assigned to a lower category on the higher-level scale). Following this idea, such a score can be calculated from the following formula

$$\ddot{s}_j = \tilde{p}_j - (1 - \tilde{B}_j^{\%}), \tag{4}$$

where $\tilde{B}_j^{\%}$ is a category-specified, rescaled fractional Borda count for jth alternative. It can be computed for each category/median $\tilde{p}_j \in D$ as

$$\tilde{B}_j^{\%} = \frac{B_j^{\%} - \min\left(\min_{c \in C: \tilde{p}_c = \tilde{p}_j} (B_c^{\%}); c_{\tilde{p}_j}^{ref}\right)}{\max_{c \in C: \tilde{p}_c = \tilde{p}_j} (B_c^{\%}) - \min\left(\min_{c \in C: \tilde{p}_c = \tilde{p}_j} (B_c^{\%}); c_{\tilde{p}_j}^{ref}\right)}. \tag{5}$$

In formula (5), $c_{\tilde{p}_j}^{ref}$ is a referential lower bound of the fractional Borda count used for scaling values $B_j^{\%}$ within resolution level (category) $\tilde{p}_j$ of the higher-level scale, and it is determined in the following way:

$$c_{\tilde{p}_j}^{ref} = \begin{cases} 0 & \text{if } \tilde{p}_j = 1 \text{ or } A = \{c: \tilde{p}_c \leq \tilde{p}_j - 1\} = \emptyset \\ \max_{c: \tilde{p}_c \leq \tilde{p}_j - 1} \left(B_c^{\%}\right) & \text{otherwise} \end{cases}. \tag{6}$$

3. Building a joint rating system for between-group comparisons

The notion of integrated scale presented in section 2 may be easily used to score alternatives within a single group of DMs. However, here we are considering a broader decision-making context in which there are t groups of students, each group V^k consists of n^k DMs (students), who assess m^k alternatives (projects):

$$V = \{V^1, V^2, \dots, V^t\}, V^k = \{v_1^k, v_2^k, \dots, v_{n^k}^k\}, C^k = \{c_1^k, c_2^k, \dots, c_{m^k}^k\}.$$

The assessment is conducted using the R2R approach, which results in determining the ranked rating sets $\Psi^k = \{\psi_1^k, \psi_2^k, \dots, \psi_{n^k}^k\}$, each consisting of a set of ratings and ranks $\psi_i^k = \{S_i^k, P_i^k\}$ expressed by the ith DM from the kth group, where $S_i^k = \{s_{i,j}^k\}_{j=1,\dots,m^k}$ is a set of cardinal ratings for alternatives $c_j^k \in C^k$ directly assigned by the DM using a predetermined scale D, and $P_i^k = \{p_{i,j}^k\}_{j=1,\dots,m^k}$ is a set of ranking positions for these alternatives. The ranked rating sets Ψ^k are used to produce sets of scores on integrated and scales $\ddot{S}^k = \{\ddot{s}_1^k, \ddot{s}_2^k, \dots, \ddot{s}_{m^k}^k\}$. Since the integrated scale is universal, one may argue that the alternatives may be compared directly using their $\ddot{s}_j^k$ values. Such between-group evaluation requires building a joint set of projects $C = \prod_{k=1}^t C^k$ and corresponding joint set of scores and sorting them according to descending $\ddot{s}_j^k$ values.

However, it must be emphasized that when the alternatives are evaluated within groups, the same scores obtained by two alternatives from different groups (e.g., c_7^1 and c_1^2 with $\ddot{s}_7^1 = \ddot{s}_1^2 = 4.5$) may not adequately represent their relative quality. This is because the general quality of the project within each group may differ. In one group, all projects may be poor, while in the other, they are good. However, all DMs in both groups may wish to differentiate scores among projects and use all grades from the scale (which is not too ample, e.g., 5-point only), as suggested by (Wachowicz & Roszkowska, 2023). On the other hand, in one group, peers could have different perceptions of what is good, very good, or average (i.e., interpret linguistic labels associated with the scale differently than DMs in another group). They might also evaluate their projects strategically, wishing to obtain higher scores than other projects from other groups. Therefore, before establishing the final integrated scores across all projects from all groups, some scores-review procedure should be proposed.

In this paper, we consider a scenario for such a scores-review procedure in which a meta-DM or a supervisor (here, the teacher) exists, who may additionally perform a between-group evaluation of some selected projects from all groups. In such a scenario, the procedure may be the following:

1. Selecting two extreme projects (i.e., the best and worst one) from each group to form a between-group reference set of projects $R = \left\{\tilde{c}_{\text{worst}}^k = \arg\max_{c_j^k \subset C^k} \ddot{s}_j^k, \tilde{c}_{\text{best}}^k = \arg\max_{c_j^k \subset C^k} \ddot{s}_j^k\right\}_{k=1,\dots,t}$.
2. Scoring each project from R with a selected scoring method (preferred by the supervisor) and mapping the result on the scale originally used to determine the within-group integrated scores. Formally, the scoring

mechanism f ensures the following:

$$f: R \rightarrow \tilde{S}, \text{ where } f(\tilde{c}^k_{\text{worst}}) = \tilde{s}^k_{\text{worst}} \text{ and } f(\tilde{c}^k_{\text{best}}) = \tilde{s}^k_{\text{best}} \text{ for each } k.$$

3. Rescaling the integrated scores within each group $\ddot{S}^k$ using the rescaled integrated scores $\tilde{s}^k_j$ provided by meta-DM.

Can you please provide me with accessLet me know if there is anything else I can help you with.to this document?For each $c^k_{j'} \subset C^k \backslash R$, the rescaled integrated rating $\tilde{s}^k_{j'}$ is obtained through linear interpolation between best and worst rescaled integrated ratings, $\tilde{s}^k_{best}$ and $\tilde{s}^k_{worst}$. This interpolation preserves the position of the point$c^k_{j'}$ in the rating space $\ddot{S}^k$ between the ratings of the extreme alternatives $\tilde{c}^k_{\text{worst}}$ and $\tilde{c}^k_{\text{best}}$, namely $\ddot{s}^k_{worst}$ and $\ddot{s}^k_{best}$

$$\tilde{s}^k_{j'} = \tilde{s}^k_{worst} + \left(\tilde{s}^k_{best} - \tilde{s}^k_{worst}\right)\frac{\ddot{s}^k_{j'} - \ddot{s}^k_{worst}}{\ddot{s}^k_{best} - \ddot{s}^k_{worst}}. \tag{7}$$

4. Example

We will present below the process of constructing a common evaluation system for peer-reviewed assessments of alternatives within two groups of student projects. This problem is part of Dery's experiment (2023), where students presented assignment projects for their academic course and assessed them using the R2R mechanism. We consider two groups of students, A and B. In group A, 27 students evaluated 10 projects, while in group B, 29 different students assessed 10 other projects. In scoring projects, students operated with a 5-point rating scale with associated linguistic labels: {"needs improvement," "fair," "good," "very good," "outstanding"}. In the case of ties, they had to compare offers pairwise, thereby constructing a total order. In Table 1, we present the results of the R2R analysis for both groups, i.e., median ratings along with Borda counts.

Table 1. Results of R2R within-group analysis for groups A and B

Group A				Group B			
Project	$Med(s_j)$	B_j	$B^{\%}_j$	Project	$Med(s_j)$	B_j	$B^{\%}_j$
a1	4	56	0,23	b1	5	131	0,50
a2	5	158	0,65	b2	4	100	0,38
a3	4	50	0,21	b3	5	202	0,77
a4	5	148	0,61	b4	5	170	0,65
a5	5	126	0,52	b5	4	85	0,33
a6	5	182	0,75	b6	4	139	0,53
a7	4	95	0,39	b7	4	112	0,43
a8	5	162	0,67	b8	4	73	0,28
a9	4	88	0,36	b9	4	126	0,48
a10	5	150	0,62	b10	5	167	0,64

Following the procedure presented in section 2, we determine cardinal scores $\ddot{s}_j$ using formula $\psi_i = \{(s_{i,1}, p_{i,1}), \ldots, (s_{i,m}, p_{i,m})\}$, (1)(4). It requires identifying first, according to formula (6), the values of referential lower bound for each scale levels resulting from median analysis, here 4 and 5, for both groups respectively. We obtain: $c^{ref,A}_5$=0.39 (as the maximun $B^{\%}_j$ for alternatiuves with $Med(s_j)$<5 equals to 0.39, and $c^{ref,A}_4$=0 (as there are no alternatives with median ratings lower than 4). For group B, we obtain: $c^{ref,B}_5$=0.53 , and $c^{ref,B}_4$=0. Now, determining $\tilde{B}^{\%}_j$ according to formula (5) requires identifying the scaling bases (max and min values). The minimum values may be either $c^{ref}_{\tilde{p}_j}$ or the minimum $B^{\%}_j$ of alternatives that category. Consequently, for project a8 we have

$$\tilde{B}^{\%}_{a8} = \frac{0.67-\min(0.52;0.39)}{0.75-\min(0.52;0.39)} = 0.77 \rightarrow \ddot{s}_{a8} = 5-0.23 = 4.77,$$

as $c_5^{ref,A} < \min_{c\in C^A:\tilde{p}_c=5}\left(B_c^{\%}\right)$, the latter equal to 0.52. However, for b4 we have

$$\tilde{B}^{\%}_{b4} = \frac{0.65-\min(0.5;0.53)}{0.77-\min(0.5;0.53)} = 0.55 \rightarrow \ddot{s}_{b4} = 5-0.45 = 4.55.$$

because $c_5^{ref,B} > \min_{c\in C^B:\tilde{p}_c=5}\left(B_c^{\%}\right)$, the latter equal to 0.5.

Similar calculations performed for all remaining alternatives lead to determining the complete set of scores $\ddot{S}^k$, shown in Table 2.

Table 2. Integrateg cardinals scores $\ddot{s}_j^k$ for groups A and B

Group A		Group B	
Project	$\ddot{s}_j^A$	**Project**	$\ddot{s}_j^B$
a1	3,59	b1	4
a2	4,75	b2	3,72
a3	3,53	b3	5
a4	4,61	b4	4,55
a5	4,36	b5	3,61
a6	5	b6	4
a7	4	b7	3,81
a8	4,77	b8	3,53
a9	3,93	b9	3,91
a10	4,63	b10	4,51

Now, we assume that the meta-DM compares extreme offers from both sets to scale them, according to our three-step procedure.

Step. 1. The reference set of extreme alternatives is defined $R = \{a3, a6, \mathrm{b}3, \mathrm{b}8\}$.

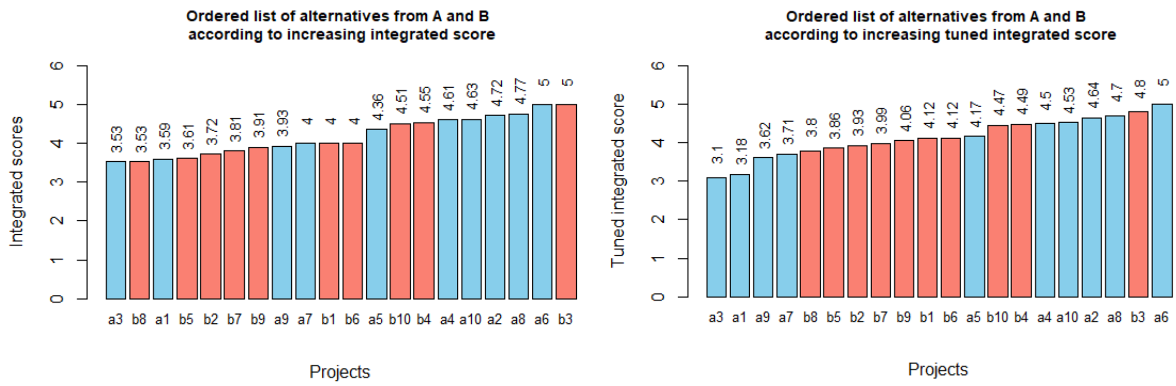
Step. 2. Meta-DM compares the best and worst options within each group and decides that a6 is somewhat better than b3, hence $\tilde{s}^A_{\mathrm{best}}$=5 but $\tilde{s}^B_{\mathrm{best}}$=4.8. Furthermore, she considers b8 better than a3, and assigns the following scores: $\tilde{s}^A_{\mathrm{worst}}$=3.1 and $\tilde{s}^B_{\mathrm{worst}}$=3.8.

Step 3. We perform a within-group rescaling, according to formula (7) using $\tilde{S}$ values, which lead to changes in scores, as shown in Table 3.

One may easily see that introducing meta-DM-based rescaling resulted in changes in the rank orders of alternatives in between-group comparisons. If the original integrated scores $\ddot{s}_j^k$ are used to rank order alternatives from A and B groups, a9 is considered 13 on the list. However, when the ranking is built using tuned $\tilde{s}_j^k$, a9 falls down the list and is ranked as 18. Similar changes may be observed for alternatives a1, a3, a7, b2, b3, b5, b7, b8 and b9. The differences in rank orders are visualized in the graphs shown in Figure 2.

Table 3. Rescaled integrated cardinals scores $\tilde{s}_j^k$ for groups A and B

Group A		Group B	
Project	$\tilde{s}_j^A$	**Project**	$\tilde{s}_j^B$
a1	3,18	b1	4,12
a2	4,64	b2	3,93
a3	3,1	b3	4,8
a4	4,5	b4	4,49
a5	4,17	b5	3,86
a6	5	b6	4,12
a7	3,71	b7	3,99
a8	4,7	b8	3,8
a9	3,62	b9	4,06
a10	4,53	b10	4,47



(a) (b)

Figure 2: Ranked alternatives from groups A and B according to original integrated ratings $\ddot{s}_j^k$ (a) and tuned integrated ratings $\tilde{s}_j^k$ (b). Group A is depicted in red and group B is in blue.

5. Conclusions

We have proposed a comprehensive mechanism for determining comparable cardinal ratings for problems where different DMs evaluate different alternatives, and meta-DM can provide additional tuning information for integrating different rating systems obtained this way. Practical examples of this scenario include creating one ranked list of student grades from disjoint sets of peer- assessment evaluations, or determining the top-k candidates for a program or a job, when candidates were given a score by different comittees (for example, a few comittees evaluate candidates in parallel).

However, one needs to be aware of certain issues related to the mechanisms adopted in our approach. One of them is the equalization of integrated ratings of those alternatives, which, according to the median determined out of the individual ratings, differed significantly on the higher-level (rating) scale. This may occur when an alternative with a lower median rating surpasses another alternative with a higher median on the second-level scale; here, fractional Borda counts. In such a case, formula (5) results in the latter receiving the same integrated rating as the former. In our example, such a situation may be observed for alternatives b1 and b6. However, it can be easily resolved by a minor modification of formula (4), which would take into account the minimal shift of alternatives from the higher category by adding to their performances the smallest interpretable value on the adopted scale (e.g., 0.01).

Another issue is the differences in evaluating alternatives by peers and a potential meta-DM. Our algorithm suggested that the meta-DM should only evaluate extreme alternatives in groups. The empirical observations of Dery's experiments proved that the instructor (meta-DM) consistently evaluated these two alternatives the same way as the students. That is, the teacher always indicated the best alternative for students as better than the one indicated by students as the worst. However, it should be noted that if the meta-DM decides to include some intermediate alternatives from different groups in the set R, this evaluation of alternatives within one group might differ from the integrated rating obtained from student evaluations. In other words, rank reversals in the meta-DM's assessment compared to peer assessments might occur. This is entirely possible because we are dealing with subjective assessments, and the preferences of the meta-DM may be arranged differently. In such a case, the procedure from section 3 will assign all alternatives between these two new reference points linearly interpolated scores that will reverse the ranks between all of them. Therefore, having detected rank reversal, it would be necessary to verify carefully, in interaction with the meta-DM, which alternatives, according to his assessment $\tilde{s}_j^k$ do not maintain the order resulting from the assessments $\ddot{s}_j^k$. To minimize preference collisions, these alternatives should be equalized in evaluation, and the rest should be scaled in relation to other neighbors who maintain the preference order defined by peers.

In further developing our method, we will focus on alternative ways of combining rating systems, such as when there is no reliable meta-DM in the problem, and the evaluation process should involve, for instance, a new group of decision-makers composed of selected representatives of peers.

References

Adesina, O. O., Adesina, O. A., Adelopo, I., & Afrifa, G. A. (2023). Managing group work: The impact of peer assessment on student engagement. *Accounting Education*, *32*(1), 90–113. https://doi.org/10.1080/09639284.2022.2034023

Brzostowski, J., & Wachowicz, T. (2014). NegoManage: A System for Supporting Bilateral Negotiations. *Group Decision and Negotiation*, *23*(3), 463–496.

Dery, L. (2023). *Forming a ranking from tied evaluations: A case of an online, interactive student peer assessment system* (arXiv:2212.13774). arXiv. https://doi.org/10.48550/arXiv.2212.13774

Double, K., McGrane, J., & Hopfenbeck, T. (2020). The Impact of Peer Assessment on Academic Performance: A Meta-analysis of Control Group Studies | Educational Psychology Review. *Educational Psychology Review*, *32*, 481–509.

Fabre, A. (2021). Tie-breaking the highest median: Alternatives to the majority judgment. *Social Choice and Welfare*, *56*(1), 101–124. https://doi.org/10.1007/s00355-020-01269-9

Panadero, E., Romero, M., & Strijbos, J.-W. (2013). The impact of a rubric and friendship on peer assessment: Effects on construct validity, performance, and perceptions of fairness and comfort. *Studies in Educational Evaluation*, *39*(4), 195–203. https://doi.org/10.1016/j.stueduc.2013.10.005

Pimentel, J. L. (2019). Some Biases in Likert Scaling Usage and its Correction. *International Journal of Sciences*, *45*(1).

Silva, V. B. S., & Morais, D. C. (2014). A group decision-making approach using a method for constructing a linguistic scale. *Information Sciences*, *288*, 423–436. https://doi.org/10.1016/j.ins.2014.08.012

Wachowicz, T. (2010). Decision support in software supported negotiations. *Journal of Business Economics and Management*, *11*(4), 576–597.

Wachowicz, T., & Roszkowska, E. (2023). How well may the direct linguistic declarations substitute AHP in defining accurate criteria weights? *International Transactions in Operational Research*. https://doi.org/10.1111/itor.13339

Group Decision and Negotiation
- Collaborative Decision Making -

Conference papers and extended abstracts

24th International Conference on Group Decision and Negotiation &
10th International Conference on Decision Support System Technology
Human-Centric Decision and Negotiation Support for Societal Transitions

Understanding expectation disconfirmation and trust in human-AI teams: Polynomial modeling and response surface analysis

Xusen Cheng[1], Xiaoping Zhang[1], Triparna de Vreede[2], and Gert-Jan de Vreede[2]

[1] Renmin University of China, 100872 Beijing, China
[2] University of South Florida, 33620 Tampa, FL, USA

Abstract

Trust is an important phenomenon in the context of team collaboration. Trust is especially important for human-AI teams (HATs) because the inclusion of an artificial intelligence (AI) teammate disrupts the interpersonal relationships of traditional human-human teams. However, little research focuses on trust development and its effects in HATs. Drawing upon expectation disconfirmation theory, this study will investigate how trust expectation disconfirmation influences trusting intention toward the human teammate versus the AI teammate and how these effects depend on expectation maturity. We propose a theoretical model to guide our investigation.

Keywords: five AI Teammate; trust; expectation disconfirmation; expectation maturity

1. Introduction

Galvanized by machine learning and deep learning, the role of artificial intelligence (AI) is transitioning from a simple tool to a teammate, which promotes the emergence of a new form of team—human-AI teams (HATs) [1, 2]. Unlike traditional passive IT tools that wait to be used, an AI teammate is a proactive entity that seeks information, proposes solutions, adapts to changing situations, and learns from past interactions [1–3]. However, the inclusion of an AI teammate disrupts the interpersonal relationships of traditional human-human teams [1], which may yield several negative consequences. For instance, the presence of an AI teammate may be detrimental to team cognition, communication, and implicit coordination [4, 5].

A high level of trust may increase an individual's willingness to take risks and be vulnerable to the actions of an AI teammate in HATs [6]. Trust is a salient issue and plays an important role in increasing team cohesiveness, communication, and coordination [7–9]. Trust promotes knowledge exchange and knowledge combination among team members [10, 11], which contributes to team performance [12]. Trust is particularly important for HATs as it can foster knowledge transfer between human and AI team members.

Trust is not static; it dynamically changes over time [13]. According to trust theory [14], trust is either strengthened, weakened, or even slashed depending on whether expectations for the behaviors of team members are met or violated. This process of adjusting trust in response to the discrepancy between initial expectation and perceived performance of team members is referred to as trust calibration [15]. Trust calibration facilitates the balance between trust belief and trust intention. Whereas trust beliefs are expectations that are considered cognitive components and come before trust intention, trust intention is a willingness to depend on trustees [9, 14]. However, little research on HATs investigates the impact of unmet trust-in-teammate expectations on trusting intention. Clarifying these effects is crucial because unmet trust may hurt team interpersonal relationships.

Expectation disconfirmation theory (EDT) offers a holistic perspective to study expectations and (dis)confirmation [16–19]. As trust research claims that trust is built upon expectation confirmation or disconfirmation, our study draws upon expectation disconfirmation theory (EDT) to examine trust in the AI teammate and trust in the human teammate in the context of human-AI teaming [20]. It is worth noting that trust disconfirmation effects may have different magnitudes on trusting intention in HATs, depending on the type of teammate. The phenomenon of algorithm aversion suggests that people lose confidence more quickly in AI that has made the same mistakes as humans [21]. Therefore, the same level of negative trust disconfirmation may lead to a greater decrease in trusting intentions toward the AI teammate compared to the human teammate. Similarly, positive disconfirmation may also result in such an asymmetric effect. Therefore, our first research question is:

RQ1: How does the impact of trust expectation disconfirmation on trusting intention toward the teammate vary with the type of teammate (human vs. AI)?

The trust literature also indicates that trust disconfirmation effects may not be linear [14]. Oliver [22] believes that nonlinear effects are more pronounced when individuals have “stronger” initial expectations. Lankton et al. [14] refer to “stronger” initial expectations as “expectation maturity”. Direct experience with technology may lead to more mature initial expectations than first impressions or reputational information. It was confirmed that nonlinear effects of technology trust disconfirmation exist when users have more mature initial trust expectations, including negative effects of positive trust disconfirmation, asymmetry between negative and positive trust disconfirmation, and curvilinear effects of trust disconfirmation. Therefore, our second research question is:

RQ2: How does expectation maturity affect the trust disconfirmation effects on trusting intention toward different types of teammates (human vs. AI)?

2. Human-AI hybrids

While certain aspects of a task (e.g., accuracy, reliability, speed, comprehensiveness, scalability) may align well with the competencies of AI agents, others (e.g., empathy, creativity, fairness) may be more compatible with the capabilities of humans [23]. Therefore, the combination of humans and AI’s capabilities can augment each other. As the competency of AI continues to evolve, humans are increasingly collaborating with AI agents to perform tasks. Rai et al [23] identified three modes of human-AI hybrids at the task level: substitution, augment, and assemblage. For task substitution, AI agents can deliver knowledge and replace humans [24]. For instance, many e-commerce companies have deployed AI chatbots that replace customer service employees to recommend products, take orders, and address customer requests [25, 26]. More studies paid attention to AI augmentation capabilities for humans, with AI generating knowledge or decisions to expand humans’ cognitive ability, help humans make decisions, and enhance humans’ decision confidence [23, 24]. AI augmentation has been applied in scenarios such as medical diagnosis, financial investment, recruitment, etc. With the dramatic development of deep learning, AI has also been used for task assemblage, where human and AI agents collaborate dynamically as an integrated unit to perform a task [23]. For instance, surgeons and AI-enabled robots have worked together to perform minimally invasive surgery [27]. Most existing literature focuses on individuals’ attitudes, perceptions, intentions, and behaviors in the contexts of AI substitution and AI augmentation. As an emerging human-AI hybrid mode, AI assemblage receives limited attention. Therefore, our research focus, HAT, is a typical human-AI hybrid for assemblage, in which human teammates share the same decision-making power with AI teammates.

3. Trust expectations disconfirmation model in human-AI teaming

Trust refers to “the willingness to be vulnerable and having confident expectations of the other party” [28]. Trust usually involves three concepts: trusting beliefs (i.e., positive expectations of the other party), trusting intentions (i.e., willingness to be vulnerable), and trusting actions (i.e., risky behaviors) [29]. Trust develops by confirming or disconfirming trust expectations (i.e., trusting beliefs). Negative disconfirmation (i.e., performance is worse than expectations) reduces trusting intention, regardless of the maturity of initial trust expectations [14]. In the initial calculus-based trust stage, trust violations (i.e., negative disconfirmation) jeopardize the relationship and undermine future dependence on the trustee because the trustor is very sensitive and cautious about the degree of risk [20, 30, 31]. When trust is knowledge-based, negative disconfirmation raises doubts about the perceptual capacity of the trustee, causing the trustor to feel unsettled and experience reduced trust. In the identification-based stage of trust, negative disconfirmation could undermine the original relationship, and even if repaired, it may be difficult to recover and reach the original level of trust [32].

Similarly, positive disconfirmation usually increases trusting intentions toward the trustee because only by doing so can an equilibrium between performance and expectations be reached [30]. However, IS and marketing literature has found that positive disconfirmation can also produce negative consequences when initial expectations are more mature [14, 22, 33]. Based on direct interaction or experience, trustors develop mature trust expectations toward the trustee and have high confidence in their own judgments. Trustors who have mature expectations perceive more cognitive dissonance, leading them to be stressed and wary [14].

Finally, even with the same level of confirmation, regardless of whether it is positive or negative, trustors may develop varying degrees of trust intentions toward different types of trustees (human teammate vs. AI teammate). Studies on algorithm aversion suggest that people quickly reduce their trust in AI when AI makes the same mistakes as humans [21]. Even when algorithms perform better than humans, people are less willing to rely on AI.

In short, drawing on EDT, this study investigates how trust expectation disconfirmation influences trusting intention toward the human teammate versus the AI teammate and how these effects depend on expectation maturity. A longitudinal study will be conducted, in which team members will collaborate three times in a row. According to Lankton et al., more mature expectations are developed based on more experience with the trustee so that the trustor feels more confident [14]. Therefore, we believe that team members develop less mature trust expectations in the first collaboration and develop more mature expectations in the last collaboration. Our research model is shown in Fig. 1.

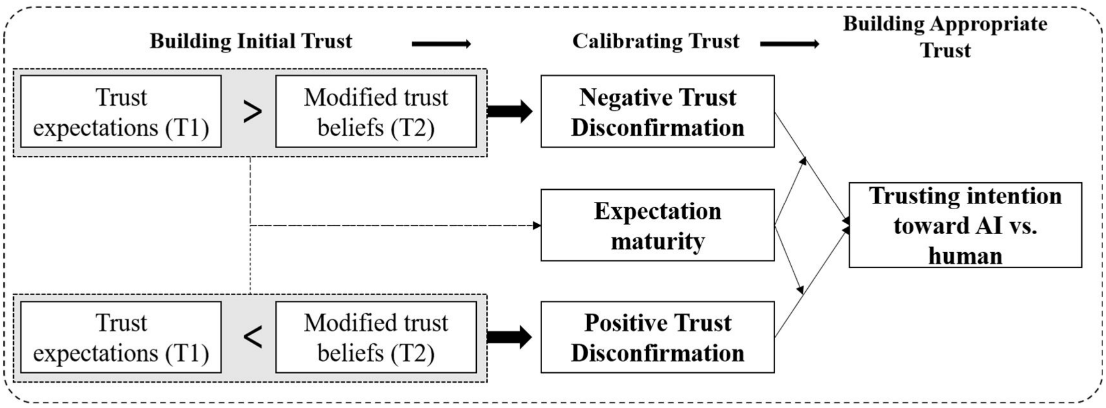


Figure 1: Research model.

4. Future research

Our research is in progress. We have developed an AI-human collaboration system that enables HATs to perform their teamwork. Subsequently, we plan to conduct a laboratory experiment to collect data on users' perceptions of human-AI teaming. Moreover, we plan to conduct polynomial modeling and response surface analysis to examine the nonlinear effect of expectation disconfirmation. Polynomial modeling is a strong method to study nonlinear effects. Given that polynomial equations are often difficult to interpret, researchers use response surface methodology to better interpret the results. Finally, we plan to conduct semi-structured interviews to corroborate and supplement the findings of the quantitative study.

Acknowledgments. The authors acknowledge the National Natural Science Foundation of China (GrantNo. 72271236) for providing funding for part of this research.

References

1. Dennis, A.R., Lakhiwal, A., Sachdeva, A.: AI Agents as Team Members: Effects on Satisfaction, Conflict, Trustworthiness, and Willingness to Work With. J. Manage. Inform. Sys. 40, 307–337 (2023).

2. Seeber, I., Bittner, E., Briggs, R.O., de Vreede, T., de Vreede, G.-J., Elkins, A., Maier, R., Merz, A.B., Oeste-Reiß, S., Randrup, N., Schwabe, G., Söllner, M.: Machines as Teammates: A Research Agenda on AI in Team Collaboration. Inf. Manage. 57, 103174 (2020).

Hauptman, A.I., Schelble, B.G., McNeese, N.J., Madathil, K.C.: Adapt and overcome: Perceptions of adaptive autonomous agents for human-AI teaming. Comput. Hum. Behav. 138, 107451 (2023).

McNeese, N.J., Schelble, B.G., Canonico, L.B., Demir, M.: Who/What Is My Teammate? Team Composition Considerations in Human–AI Teaming. IEEE. Hum-Mach. Sys. 51, 288–299 (2021)

Musick, G., O'Neill, T.A., Schelble, B.G., McNeese, N.J., Henke, J.B.: What Happens When Humans Believe Their Teammate is an AI? An Investigation into Humans Teaming with Autonomy. Comput. Hum. Behav. 122, 106852 (2021).

Mayer, R.C., Davis, J.H., Schoorman, F.D.: An Integrative Model of Organizational Trust. Acad. Manage. Rev. 20(3), 709–734 (1995)

Jarvenpaa, S.L., Shaw, T.R., Staples, D.S.: Toward Contextualized Theories of Trust: The Role of Trust in Global Virtual Teams. Infor. Syst. Res. 15, 250–267 (2004).

Kuo, F., Yu, C.: An Exploratory Study of Trust Dynamics in Work-Oriented Virtual Teams. J. Comput-Mediated. Comm. 14, 823–854 (2009).

Robert, L.P., Denis, A.R., Hung, Y.-T.C.: Individual Swift Trust and Knowledge-Based Trust in Face-to-Face and Virtual Team Members. J. Manage. Inform. Sys. 26, 241–279 (2009).

Collins, C.J., Smith, K.G.: Knowledge Exchange and Combination: The Role of Human Resource Practices in the Performance of High-Technology Firms. Acad. Manage. J. 49, 544–560 (2006).

Reinholt, M., Pedersen, T., Foss, N.J.: Why a Central Network Position Isn't Enough: The Role of Motivation and Ability for Knowledge Sharing in Employee Networks. Acad. Manage. J. 54, 1277–1297 (2011).

Choi, S.Y., Lee, H., Yoo, Y.: The Impact of Information Technology and Transactive Memory Systems on Knowledge Sharing, Application, and Team Performance: A Field Study. MIS. Q. 34, 855–870 (2010).

Zahedi, F. "Mariam", Song, J.: Dynamics of Trust Revision: Using Health Infomediaries. J. Manage. Inform. Sys. 24, 225–248 (2008).

Lankton, N.K., McKnight, D.H., Wright, R.T., Thatcher, J.B.: Using Expectation Disconfirmation Theory and Polynomial Modeling to Understand Trust in Technology. Infor. Syst. Res. 27, 197–213 (2016).

De Visser, E.J., Peeters, M.M.M., Jung, M.F., Kohn, S., Shaw, T.H., Pak, R., Neerincx, M.A.: Towards a Theory of Longitudinal Trust Calibration in Human–Robot Teams. Int. J. Soc. Robotics. 12, 459–478 (2020).

Oliver, R.L.: A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions. J. Market. Res. 17, 460–469 (1980).

Brown, S.A., Venkatesh, V., Kuruzovich, J., Massey, A.P.: Expectation confirmation: An examination of three competing models. Organ. Beh. Hum. Dec. Proc. 105, 52–66 (2008).

Brown, S.A., Venkatesh, V., Goyal, S.: Expectation Confirmation in Technology Use. Inform. Sys. Res.

23, 474–487 (2012).

Darke, P.R., Ashworth, L., Main, K.J.: Great expectations and broken promises: misleading claims, product failure, expectancy disconfirmation and consumer distrust. J. of the Acad. Mark. Sci. 38, 347–362 (2010).

Lewicki, R.J., Tomlinson, E.C., Gillespie, N.: Models of Interpersonal Trust Development: Theoretical Approaches, Empirical Evidence, and Future Directions. J. Manage. 32, 991–1022 (2006).

Dietvorst, B.J., Simmons, J.P., Massey, C.: Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err. J. Exp. Psychol. Gen. 114, 114–126 (2014)

Oliver, R.L.: Hedonic Reactions to the Disconfirmation of Product Performance Expectations: Some Moderating Conditions. J. Appl. Psych. 61, 246–250 (1976).

Rai, A., Constantinides, P., Sarker, S.: Editor's comments: Next-Generation Digital Platforms: Toward Human–AI Hybrids. MIS. Q. 43, iii–ix (2019).

Chen, L., Hsieh, J.J.P.-A., Rai, A.: How Does Intelligent System Knowledge Empowerment Yield Payoffs? Uncovering the Adaptation Mechanisms and Contingency Role of Work Experience. Infor. Syst. Res. 33, 1042–1071 (2022).

Chandra, S., Shirish, A., Srivastava, S.C.: To Be or Not to Be …Human? Theorizing the Role of Human-Like Competencies in Conversational Artificial Intelligence Agents. J. Manage. Inform. Sys. 39, 969–1005 (2022).

Luo, X., Tong, S., Fang, Z., Qu, Z.: Frontiers: Machines vs. Humans: The Impact of Artificial Intelligence Chatbot Disclosure on Customer Purchases. Market. Sci. 38, 937–947 (2019).

Sergeeva, A.V., Faraj, S., Huysman, M.: Losing Touch: An Embodiment Perspective on Coordination in Robotic Surgery. Org. Sci. 31, 1248–1271 (2020).

Cheng, X., Fu, S., Druckenmiller, D.: Trust Development in Globally Distributed Collaboration: A Case of U.S. and Chinese Mixed Teams. J. Manage. Inform. Sys. 33, 978–1007 (2016).

Jones, S.L., Shah, P.P.: Diagnosing the Locus of Trust: A Temporal Perspective for Trustor, Trustee, and Dyadic Influences on Perceived Trustworthiness. J. Appl. Psych. 101, 392–414 (2016).

McEvily, B., Perrone, V., Zaheer, A.: Trust as an Organizing Principle. Org. Sci. 14, 91–103 (2003).

Van Der Werff, L., Buckley, F.: Getting to Know You: A Longitudinal Examination of Trust Cues and Trust Development During Socialization. J. Manage. 43, 742–770 (2017).

Lewicki, R., Bunker, B.: Trust in relationships: A model of development and decline. In: Bunker BB, Rubin JZ, eds. Conflict, Cooperation, and Justice. pp. 133–173. , Jossey-Bass, San Francisco (1994)

Han, E., Yin, D., Zhang, H.: Bots with Feelings: Should AI Agents Express Positive Emotion in Customer Service? Infor. Syst. Res. 34, (2022).

An Analysis of Shipper-Shipping Company Relationships for Enhancing Sustainability: Application of Fuzzy-set Qualitative Comparative Analysis

Chang-Soo Kim[1], Saeyeon Roh[2*], Mohamed Haddoud[3], Adah-Kole Onjewu[4] and Young-Joon Seo[5]

[1] Pusan National University, Pusan, South Korea
[2] University of Plymouth, Plymouth, United Kingdom
[3] The British University in Dubai, Dubai, United Arab Emirates
[4] Northumbria University, United Kingdom
[5] Kyongpook National University, Daegu, South Korea
[*] saeyeon.roh@plymouth.ac.uk

Abstract

This study aims to evaluate the extent of collaboration between shippers and shipping companies by investigating the factors to enhance shipper-shipping company relationships. It employs a fuzzy-set configuration approach to capture the complex interactions underlying these factors in the context of South Korean shipping companies. Based on data from a sample of 167 respondents, the findings revealed three complex paths with factors including Trust, Fairness, Transparency, and Mutuality lead to high and low Sustainability in shaping the relationship. Theoretically, the findings are an exception to extant the comprehensive inter-firm cooperation and collaboration models. Practically, the attention of practitioners in South Korea and other similar context was drawn.

Keywords: Shipper-Shipping Relationship; Sustainable; Maritime logistics; South Korea; fsQCA

1. Introduction

Throughout the globe, the extent and structure of international seaborne transport have been impacted by the increasingly globalised nature of commerce. Sea transport carries over 80 percent of the world's merchandise (UNCTAD, 2023), and approximately three-fifth of commodities by value are annually transported internationally through container shipping (Mason and Nair, 2013). These instances indicated that shipping logistics plays a central role not only in the supply chains (SCs) of individual countries but also on a global scale.

Nevertheless, with the increasing openness of the global economy facilitated by the World Trade Organisation (WTO) and Free Trade Agreements (FTA), the shipping industry consistently experience recurring crises akin to the cyclic nature of the global financial crisis (Kuo et al., 2017). Global container carriers have faced challenges in navigating the aftermath of the crises, grappling with a sluggish economy and excess capacity.

In the context of the relationship between shippers and logistics companies, with a few exceptions like the recent COVID-19 pandemic, shippers have typically maintained a dominant advantage over shipping lines throughout the majority of shipping cycles. These cycles have shortened from 14.9 years to 8 years over the last 22 cycles (Stopford, 2009). In highly competitive industries such as logistics, where shippers (buyers) can easily switch shipping companies (suppliers), shippers (customers) are generally seen as wielding more power in supplier-customer relationships (Golicic, 2007). Additionally, Midoro et al. (2005) observe that shippers' contractual power has grown due to the imbalance between supply can demand, as well as intense competition in container shipping, manifesting in the following ways. The expansion of manufacturers on a global scale has required carriers to adapt to new demands from shippers for global goods delivery, leading to a persistent surplus of fleet capacity. Shipping companies are challenged by the increased costs associated with providing globalised services. The entry of new carriers into shipping markets has heightened competition globally (Slack et al., 2002). In this context, major carriers were compelled to implement strategies that enhance economies of scale, such as increasing vessel sizes and engaging in mergers and acquisitions. This was driven by the significant challenge of maintaining stable freight rates in an intensely competitive business environment (Midoro et al., 2005). Put differently, the pursuit of economies of scale has exerted pressure to maximise freight capacity on ships, leading shipping lines to accept whatever price the shipping market offers and prioritise cost reduction (Notteboom et al., 2010).

In this fiercely competitive and shippers-dominated market, survival strategies for shipping companies may prioritise efforts to establish and sustain freight transportation contracts with shippers to maximise ship space utilisation. Given that the revenue of shipping companies relies solely on shippers, who can easily switch to other shipping providers, the success of shipping companies hinges on shippers adopting a collaborative attuite toward them (Talley and Ng, 2013). Shippers' collaborative behaviours, such as ensuring reasonable profits, sharing additional costs, and entering into long-term contracts, can greatly assist shipping lines struggling to navigate their challenges. Consequently, there is a growing interest in recognising the necessity for sustainable relationships between shipping companies and shippers.

In summary, this study addresses existing voids and advances the literature by exploring the interactions underlying the development of sustainable shipper-shipping company relationships vis-à-vis transparency, fairness, mutuality, and trust. Survey data from shipping companies registered in South Korea returned were examined using a fuzzy-set Qualitative Comparative Analysis (fsQCA). The fsQCA is a case-based analytical tool that can identify complex combinations of factors that lead to a given outcome (i.e. sustainable relationship). Accordingly, this study addresses two pressing research questions: (1) What transparency, fairness, mutuality, and trust factors support the development of sustainable relationship? (2) Are the factors influencing autonomous or heteronomous sustainable relationship development?

The rest of the study is arranged as follows: Section 2 develops the conceptual framework through a literature review, Section 3 explain the method and instruments, and Section 4 and 5 present the calibration, necessity, and sufficiency analyses, which are features of the fsQCA. Finally, Section 7 discusses the results and concludes the study.

2. Conceptual Framework

2.1. Theoretical perspectives

Inter-company relationships can be examined using various theoretical frameworks, such as transaction cost theory (TCT), resource-based theory (RBT), resource dependency theory (RDT), contingency theory (CT), social exchange theory (SET), and stoical capital theory (SCT) (Kim et al. 2010; Cao et al., 2010).

Mainly, TCT centres on determining whether a transaction is executed more efficiently within a single vertically integrated firm or by independent contractors operating in the external market under market governance (Geyskens et al. 2006). Transactions costs may emerge due to asset specificity, uncertainty, or transaction frequency, potentially causing market failure, thereby favouring vertical integration (Williamson et al., 1975). Collaborative relationships between firms can mitigate transaction costs and enhance transaction stability by minimising opportunistic behaviour among partners (Kim et al., 2010).

The resource-based view (RBV) seeks to explain a firm's sustainable competitive advantage by emphasising resources that are, valuable, difficult to imitate, and hard to substitute (Bromiley & Rau, 2016). RBV, particularly focused on a firm's internal resources and capabilities, posits that these factors contribute to a competitive edge (Dyer and Singh, 1998). Resources encompass what a firm possesses or controls, while capabilities refer to the firm's ability to effectively use these tangible or intangible resources. In the context of collaborative networks with other firms, RBV provides a theoretical foundation for cooperation. Relational theory (RT) extends RBV to inter-organisational relationships (Acedo et al., 2006) and serves a complementary theory, exploring the origins of competitive advantage within inter-firm networks (Dyer and Singh, 1998).

As per RDT, firms dependent on mutual exchanges for essential resources, tend to engage in cooperation, particularly in the face of uncertainty, fostering interdependence with their partners (O'Toole and Donaldson, 2000). Successful responses to market demands hinge on collaboration and support from these partners. RDT and RBT frequently suggest that collaboration between firms is instrumental in enhancing their overall performance (Ramathan and Gunasekaran, 2014). CT suggests that the optimal management approach for an organisation depends on its specific organisational environment. A successful 'fit' occurs when an organisation aligns effectively with its environment, leading to improved performance (Betts, 2003). However, as business environments differ and influence the correlation between organisation and performance, collaboration between companies can be beneficial, especially in rapidly changing environments, to enhance this alignment.

SET delves into how the interactions among participants in an exchange process are shaped by rewards and costs (Molm, 1991). As actors assess the rewards and costs associated with an exchange, their decision to engage in the exchange process leads to the formation of relationships aimed at maximising benefits and minimising costs (Nunkoo and Ramkissoon, 2012). In the context of inter-firm relationships, SET literature highlights justice as a primary factor, followed by trust and commitment (Morgan and Hunt, 1994).

Social capital refers to 'networks along with shared norms, values, and understandings that promote cooperation within or among groups' (OECD, 2011, p. 41). The presence of high levels of trust within social capital can decrease the probability of opportunism and the need for costly monitoring processes, ultimately reducing transaction costs. Social capital has the potential to foster cooperative behaviour and serves as a resource that a firm can leverage (Adler and Kwon, 2002).

2.2. The determinants of sustainable shipper-shipping company relationship

Spekman et al. (1998) introduced a step function illustrating the shift from open market negotiations to cooperation, coordination, and collaboration. Cooperation, as an initial relationship, differs from collaboration, which entails higher levels of trust and commitment. Golicic et al. (2003) posit that collaboration represents the highest level of inter-firm relationships. Cooperation involves lower levels of trust, commitment, and mutual dependence and is less active compared to collaboration. The distinguishing factors among the three concepts lie in the extent of trust and commitment. Sustainability, as proposed, encompasses commitment and a long-term orientation (1998). To confirm the relevance of the determined factors, two rounds of Q-sorting and a pilot test were conducted with 21 experts including director-level civil servants, members of Korea Shipowners' Association, senior researchers of Korea Maritime Institute and university professors.

Transparency

Transparency in relationship establishment is achieved through clear communication, shared information, and advance agreements, fostering open and straightforward connections between parties (Kim et al., 2018). This transparency is composed of elements such as information sharing, communication, and formalisation. In supply chain management (SCM), a crucial aspect of collaboration involves information sharing, serving as a key component of cooperation (Kumar et al., 2014). Establishing an efficient supply chain (SC) requires private data exchange among partners, and a reluctance to share pertinent information can pose challenges in forming SC relationships (Richey et al., 2010). On the contrary, detailed information exchange not only facilitates between decision-making but also contributes to the efficiency of the SC by enhancing visibility (Min et al., 2005). Communication in the supply chain is essential for identifying opportunities and addressing improvement needs, with collaborative communication involving frequency, means, direction, and influence among members. Formalisation, characterised by clear rules and procedures, plays a crucial role in influencing decisions and working relationships, contributing to transparency and standardisation within the supply chain. (Min et al., 2005; Cao et al., 2010; Mohr & Nevin, 1990; Dwyer & Oh, 1987; Daughterty et al., 2005).

Fairness

Fairness denotes the degree to which one partner treats another with equity and justice, drawing upon principles of justice and reciprocity (Konovsky, 2000). While fairness can be framed within collaborative buyer-supplier relationships and has an impact on the level of cooperation between partners (Kim et al. 2010), it remains an underexplored aspect in supply chain research. Traditionally, fairness has been characterised by procedural and distributive justice (Kumar et al., 1995). Procedural justice pertains to how fairly a firm and its personnel engage with a partner (Konovsky, 2000). Distributive justice, on the other hand, outlines the equitable distribution between partners, aligning with their contributions and ensuring mutually satisfactory levels of cooperation (Harland, 2004). Given that the establishment of trust among partners is contingent on just interactions (Dwyer et al., 1987).

Mutuality

In the context of interdependence and the RDT, mutuality signifies reciprocal relationships between two organisations, where partners treat and support each other equitably based on a shared understanding (Thomson et al., 2009). Mutuality encompasses elements such as resource sharing, collaborative knowledge creation, joint problem-solving/performance measurement, and goal congruence (Kim et al., 2018). Resource sharing involves investing in and utilising assets and capabilities with SC partners (Cao et al., 2010). Joint knowledge creation facilities a clearer understanding and more effective responses to dynamic markets and environments (Malhotra et al., 2005). Collaborative problem-solving entails amicably setting disputes and disagreements among partners (Kumar et al., 2014), leading to mutually beneficial process improvements. The establishment of cross-functional teams and personnel co-location lead to virtual integration of supply chain processes, while joint performance measurement is crucial for the success of collaborative efforts. (Lee et al., 2016; Min et al., 2005; Seo et al., 2014). Goal congruence, reflecting mutual alignment in company beliefs and values, is achieved when supply chain partners fulfil objectives, emphasising the importance of management commitment and financial/non-financial investment for sustain collaborative relationships (Cao et al., 2010; Min et al., 2005).

Trust

Trust is fundamentally influence by fairness, and cultivating fairness can foster mor efficient and high-quality relationships among firms (Kim et al., 2018). Trust pertains to the level of confidence a partner company inspires in terms of trustworthiness, good faith, and the fulfilment of obligations. It signifies the extent to which partners perceive each other as reliable (Ganesan, 1994). Trust can be viewed as a belief or expectation that one particular will not exploit the vulnerability inherent in their relationship or transaction, a vulnerability arising from the acceptance of inevitable risks (Lane, 2000). Credibility and honesty constitute two essential components of trust (Eyuboglu et al., 2003), while credibility and benevolence serve as two dimensions for assessing trust (Wang et al., 2008). Establishing trust can lead to a reduction in various costs associated with pre-negotiation, contract finalisation, and post-transaction activities (Ryu et al., 2007). Trust also contributes to diminishing anxiety and uncertainty between partners (Wang et al., 2008), thereby lowering transaction costs (Ganesan 1994, Kwon & Suh, 2004). Moreover, trust acts as a restraint on opportunistic behaviour by one partner and deters a dominant partner from exerting undue power over a weaker counterpart (Ganesan, 1994; Mei & Dinwoodie, 2005).

Sustainability

Sustainability serves as a continual reinforcement of collaborative partnership with another entity, incorporating principles of dedication and a focus on the long term. These principles extend the duration of relationships, establishing loyalty and fostering long-term expectations (Gardner et al., 1994). The successful implementation of SC performance necessitates a high level of trust and commitment. Commitment involves an implicit commitment to sustain the relationship between partners, allowing them to accommodate each other's shortcomings (within reasonable limits) and coordinate actions rather than taking advantage of circumstances (Dwyer & Oh, 1987; Min et al., 2005).

Within SCs, organisational commitment is comparable to a willingness on the part of the less dominant party to engage in a prolonged relationship, with the quality of the relationship improving as commitment levels rise (Nyaga & Whipple, 2011). Long-term orientation reflects a partner's desire to establish a lasting partnership (Ganesan, 1994). Effective SCM necessitates the establishment, continuation, and fortification of enduring partnership with SC collaborators (Kim et al., 2020). As partners develop trust in each other through the success of their collaborative endeavours, collaborative relationships and cooperation are likely to be strengthened (Min et al., 2005).

The research model is illustrated in Figure 1.

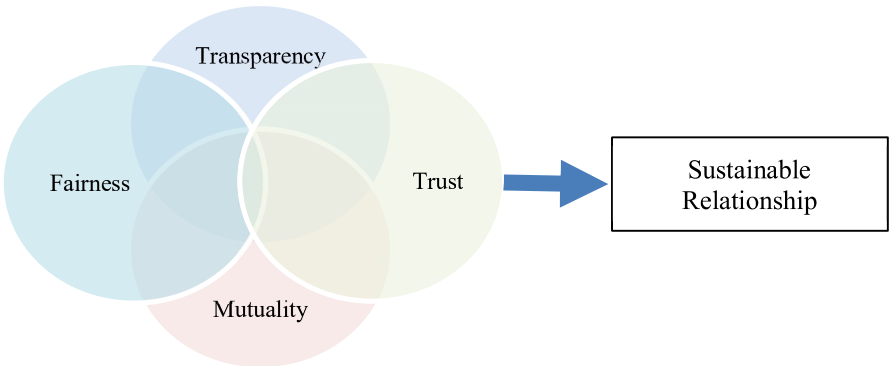


Figure 1. Research Model

3. Method

This study investigates how transparency, fairness, mutuality, and trust impact the development of sustainable relationship in South Korea's shipping industry. The shipping industry is marked by intense competition, with shippers generally wielding more influence in relationships with service providers (Golicic, 2007). In this competitive context, the common approach in relationship research, focusing on one side of the relationship, was adopted (Golicic & Mentzer, 2006).

The shipping industry in South Korea provides a valuable environment for examining inter-firm cooperation and collaboration, with substantial presence of 70 million deadweight tonnes of registered vessels in 2016. South Korea is home to six prominent liner shipping companies, with major shippers such as Samsung, Hyundai, and Lotte. On a global scale, South Korea holds the sixth position in steel production, fifth in steel consumption, fourth in iron ore imports, and fifth in coal consumption (UNCTAD, 2023).

South Korea accommodated 183 ocean-going and 723 coastal shipping companies as reported by the Ministry of Oceans and Fisheries (2022). The Korea Shipowners' Association supplied contact information for all ocean-going companies and 241 coastal shipping companies. A survey questionnaire accompanied by two subsequent reminder emails and phone calls. Out of the 174 responses received, 167 were deemed usable, excluding one with missing data and six that did not provide complete engagement. The overall response rate of 25.5% and an active response rate of 39% indicated a representative sample (Neuman, 2014) (Table 1). The measures utilised are detailed in Appendix A, including their sources. All items were evaluated using a 5-point Likert scale.

Table 1. Demographic Data for 167 Respondents

Variables	Frequency	Percentage (%)
Shipping registered		
Oceangoing	89	53.3
Coastal	78	46.7
Vessel type		
Container	16	9.6
Bulk carrier	66	39.5
Tanker	51	30.5
Others	34	20.4
Work experience		
< 5 years	18	10.8
5 – 9 years	28	16.8
10 – 19 years	90	53.8
≥ 20 years	31	18.6
Job title		
(Senior) Director/CEO	36	21.6
Department manager	58	34.7
Manager/Deputy department manager	54	32.3
Staff/Assistant manager	19	11.4
Employee		
< 10	18	10.8
10 – 49	60	35.9
50 – 99	43	25.7
100 – 199	21	12.6
≥ 200	25	15.0
Contract period		
< 1 year	55	32.9
1 – 2 years	56	33.5
3 – 9 years	28	16.8
≥ 10	28	16.8

4. Constructs' reliability and validity

Before performing the fsQCA, an assessment of the reliability and validity of the constructs was conducted. In this study, a structural equation modelling protocol was employed using WarpPLS version 8.0 (Kock, 2022). Reliability was evaluated using the Composite Reliability (CR) and Cronbach's Alpha (α), while validity was scrutinised through the Average Variance Extracted (AVE) and factor loadings for convergent validity, as well as the square roots of AVEs for discrimination validity. Collinearity was examined using the variance inflation factor (VIF) (Hair et al., 2016). The result of the measurement model (Table 2) demonstrates satisfactory scores in terms of reliability (with CR and α values below the 0.7 threshold), convergent validity (with AVE scores above the 0.5 cut-off and all loadings above 0.6, as shown in Appendix A), and collinearity (with VIF levels below or equal to 5). Discriminant validity is also confirmed, as the square roots of AVE are greater than the diagonal (Hair et al., 2011). These scores collectively indicate a high level of a measurement quality.

Table 2. CR, Cronbach's α, AVE and VIF

Factor Variable	Cronbach's α	CR	AVE	VIF
Transparency	0.925	0.947	0.817	3.655
Fairness	0.955	0.963	0.764	3.726
Mutuality	0.928	0.949	0.823	/
Trust	0.920	0.944	0.808	2.934
Sustainability	0.914	0.940	0.796	2.487

5. fsQCA analysis

The fsQCA method is employed to identify conditions, often in the form of combinations, that result in a specific outcome (Fiss et al., 2013; Ordanini et al., 2014). This approach incorporates contrarian cases, which deviate from the general trend of the data (Woodside, 2014), thereby minimising the issues related to unobserved heterogeneity (Schneider and Wagemann, 2010). Developed by Ragin (2000), this study utilises software version 3.0 (Ragin and Davey, 2016). fsQCA provides comprehensive insights by revealing (1) multiple solutions that can successfully lead to the same outcome (equifinality) and (2) addressing complex causality issues by capturing latent interactions among a set of antecedents (Sclittgen et al., 2016; Wong et al., 2018).

5.1. Calibration

This constitutes the initial stage of the fsQCA analysis. Fuzzy scores were derived from Likert scale values by delineating three qualitative thresholds, reflecting fuzzy set scores and their corresponding data values (Ragin, 2009). These thresholds include full membership, crossover point (0.5), and full non-membership (0) (Ragin, 2009). Identifying these thresholds could be based on the 95th, 50th, and 5th percentiles, although caution has been advised against sample-based calibration, whenever possible (Greckhamer et al., 2018). For Likert scales, thresholds such as 4 [agree], 3 [neutral], and 2 [disagree] (Pappas and Woodside, 2021), or 5 [strongly agree], 3 [neutral], and 1 [strongly agree] (Laouiti et al., 2022) are applicable. In this study, the latter threshold precedent was selected to denote full membership, the crossover point, and non-membership.

5.2. Necessity analysis

This step ensues after the calibration process. The conditions (or variables) essential for sustainable relationship were identified. A necessary condition implies that its presence is required but may not alone be sufficient for the outcome to transpire (Kent, 2015). Necessity is affirmed when a condition demonstrates a consistency score of at least 0.9 (Legewie, 2013). Consistency gauges the extent to which cases in the sample sharing a condition or configuration agree to exhibit the desired outcome (Ragin, 2008). Table 3 indicates that Trust has the highest consistency score of 0.945, which explains the necessitates confirmation through a sufficiency analysis.

Table 3. Necessity Analysis

	Consistency	Coverage
Transparency	0.851	0.847
~ Transparency	0.577	0.637
Fairness	0.841	0.870
~ Fairness	0.585	0.620
Mutuality	0.770	0.934
~ Mutuality	0.655	0.602
Trust	0.945	0.802
~ Trust	0.482	0.659

5.3. Sufficiency analysis

After determining the necessary conditions, a sufficiency analysis was undertaken. During this phase, the combinations leading to sustainable relationships were identified by generating truth tables that encompass logically plausible combinations of conditions (Ragin, 2008). To pinpoint the relevant configuration for the desired outcome, thresholds for the minimum number of cases associated with those combinations (referred to as the frequency threshold) and the appropriate consistency score (referred to as the consistency cut-off) were set (Woodside and Zhang, 2012). In this study, the former was established at four cases, aligning with Ragin's (2008) recommendation to use higher thresholds for larger samples. For the latter, a cut-off value of at least 0.80 was applied. The consistency score for the combination should be at least 0.75 for the combination to be deemed consistent.

Three solution types were considered: parsimonious, intermediate, and complex. An intermediate solution is interpreted as the midpoint between parsimonious and complex solution (Kent, 2015). The summarised results can be found in Table 4. Each combination is accompanied by scores for consistency and coverage. Similar to the significance value in multivariate techniques, consistency reflects “the degree to which the cases sharing a given combination of conditions...agree in displaying the outcome in question”. Coverage reflects “the degree to which a cause or causal combination ‘accounts for’ instances of an outcome” (Ragin, 2008: 44), indicating the empirical importance of sufficient configurations (Ordanini et al., 2014). Coverage can be raw or unique, where raw coverage may overlap with other combinations, while unique coverage is exclusive to certain combinations (Beynon et al. 2016). The overall solution is also presented to signify the extent to which outcomes can be determined by a set of configurations, akin to R-square value in multivariate methods (Woodside, 2014). Additionally, core and complementary (peripheral) conditions were identified. Core conditions demonstrate a robust causal association with outcomes, while peripheral elements exhibit a weaker association (Fiss, 2011).

As depicted in Table 4, three configurations associated with high sustainable relationship emerged. All three involves high levels of Trust (confirming its necessity), with the first combination adding Transparency, the second configuration involving Fairness, and last combination adding both Transparency and Fairness. From the raw coverage scores, the first two combinations bear relatively similar empirical relevance, and the last configuration exhibits higher relevance. The overall solution coverage was 0.86, reflecting the proportion of the sustainable relationship covered by the three solutions.

Table 4. Intermediate Solutions for High Sustainability

Solutions	Raw Coverage	Unique Coverage	Consistency
TRA*TRU*~MUT	0.540	0.041	0.924
TRU*FAI*~MUT	0.540	0.041	0.904
TRU*TRA*FAI	0.769	0.270	0.936
Solution Coverage	0.851		
Solution Consistency	0.901		

5.4. Negation analysis

The fsQCA is an asymmetric analysis, enabling researchers to identify conditions that result in the absence of the outcome under consideration, namely sustainable relationships (Kent, 2015; Woodside & Zhang, 2013). The principle of causal asymmetry posits that the conditions predicting the presence of an outcome can differ from those predicting its absence (Pappas & Woodside, 2021). The negation analysis, presented in Table 5, illustrates the combinations leading to the absence of sustainability. Consequently, two possible combinations have surfaced. The first involves low levels of Fairness and Mutuality, while the second includes low levels of mutuality and transparency with high levels of trust. The first solution exhibit higher empirical relevance to the second solution. The two solutions have an overall coverage of 0.87.

Table 5. Intermediate Solutions for Low Sustainability

Solutions	Raw Coverage	Unique Coverage	Consistency
~FAI*~MUT	0.836	0.254	0.861
~TRA*MUT*TRU	0.620	0.038	0.903
Solution Coverage	0.874		
Solution Consistency	0.840		

6. Discussion and Conclusion

The structural equation analysis shows that Trust is the only significant predictor for sustainability, whereas Transparency, Mutuality and Fairness do not hold a significant include. The model explains 66% of sustainability. On the other hand, the fsQCA analysis confirms that Trust is indeed a necessary condition to reach sustainability. However, it also shows that Trust is not sufficient and must be combined with other factors to reach sustainability. Here, trust needs to be combined with either (1) transparency and fairness, (2) transparency and low mutuality or (3) fairness and low mutuality. The three solutions explain 90% of sustainability. Once could perhaps argue here that when there is no mutuality, transparency or fairness (in addition to trust) would lead to high sustainability. Moreover, when both transparency and fairness are added to trust, sustainability will be reached regardless of the level of mutuality.

As for the conditions leading to low sustainability, it appears that when low mutuality is combined with either low fairness or low transparency, even with high trust, no sustainability will be achieved. Hence, confirming that trust is not sufficient, and that low mutuality could a double-edged sword, i.e when combined with low factors, it will lead to low sustainability, while when combined with high factors, it will not prevent high sustainability. Here, the necessity analysis for low sustainability confirms low mutuality to be a necessary condition.

This study employs a configurational approach, revealing that sustainable relationships are a complex outcome influenced by various combinations of conditions. This supports the notion of equifinality, suggesting that multiple pathways can lead to the same outcome. The findings suggest practical implications for fostering sustainable relationships, emphasising the importance of combining Trust with Transparency and Fairness. The findings can guide practitioners in developing strategies for achieving and maintaining sustainable relationships. Current study contributes to the theoretical advancements in understanding the complexity of conditions influencing sustainable relationships. Offering methodological insights and practical implications for both researchers and practitioners in the field.

Appendix A. Items and Loadings

	Loading	Reference
Transparency		
Shippers exchange relevant and timely information with our firm	0.899	Cao et al. (2010), Daugherty et al. (2006), Dwyer & Oh (1987), Kim et al. (2018), Kim et al. (2020), Kim et al. (2022), Kumar et al. (2014), Min et al. (2005), Mohr & Nevin (1990) and Richey et al. (2010)
Shippers and our firm communicate smoothly with each other	0.843	
Shippers make communication with our firm open and two-way	0.937	
The relationship between shippers and our firm is understood clearly and transparently through prior agreements	0.913	
Fairness		
Shippers do not discriminate our firm against other shipping companies	0.861	Bensaou (1997), Griffith et al. (2006), Harland et al. (2004), Hornibrook et al. (2009), Kim et al. (2018), Kim et al. (2020), Kim et al. (2022), Konovsky (2000), Korsgaard et al. (1995), Kumar et al. (1995), Nassrini & Robinson (2013) and Simatupang & Sridharan (2005)
Shippers try to comply with the regulations related to business transactions	0.898	
Shippers make an effort to guarantee reasonable and just profits for our firm	0.943	
Shippers make an effort to bear reasonably and justify any additional risks, burden, and costs related to delivery with our firm	0.911	
Mutuality		
Shippers understand our firm's services well and are willing to assist us	0.896	Cao et al. (2010), Kumar & Banerjee (2014), Kim et al. (2018), Kim et al. (2020), Kim et al. (2022), Malhotra et al. (2005), Min et al. (2005) and Thomson et al. (2000)
Shippers are willing to provide their facilities and equipment for our firm	0.899	
Shippers are willing to provide financial support for our firm	0.688	
Shippers are willing to assist our firm in overcoming any difficulties we face	0.844	
Shippers and our firm agree common implementation plans and objectives	0.931	
Shippers and our firm, as equal business partners, decide together the availability level of our facilities and equipment	0.932 0.883	
Shippers and our firm identify together customer needs related to delivery	0.907	
Shippers and our firm identify together customer needs related to delivery		
Trust		
Overall, shippers are trustworthy	0.889	Anderson and Narus (1990), Eyuboglu et al. (2003), Fleming et al. (2020), Ganesan (1994), Golicic (2007), Kumar et al. (1995), Kwon & Suh (2005), Min et al. (2005) and Wang et al. (2008)
We believe the good faith offered by shippers	0.931	
We believe that shippers fulfil their contractual obligations	0.781	
We believe that shippers benefit our firm	0.876	
Sustainability		
The relationship between shippers and our firm is stable	0.872	Dwyer & Oh (1987), Ganesan (1994), Gardner et al. (1994), Hornibrook et al. (2009), Kim et al. (2018), Kim et al. (2020), Kim et al. (2022), Min et al. (2005) and Nyaga & Whipple (2011)
The relationship between shippers and our firm will last and strengthen	0.922	
Shippers try to develop new business plans or ideas together with us	0.918	
Shippers enhance their relationship with our firm by expanding markets jointly	0.916	

References

Acedo, F.J., Barroso, C., & Galan, J.I. 2006. The Resource-Based Theory: Dissemination and Main Trends. *Strategic Management Journal*, 27, 621-636.

Adler, P.S., & Kwon, S.W. 2002. Social Capital: Prospects for a New Concept. *The Academy of Management Review*, 27, 17-40.

Anderson, J.C. & Narus, J.A. 1990. A Model of Distributor Firm and Manufacturer Firm Working Partnerships. *Journal of Marketing*, 54(1), 42-58.

Bensaou, M. 1997. Interorganisational Cooperation: The Role of Information Technology an Empirical Comparison of US and Japanese Supplier Relations. *Information Systems Research*, 8(2), 107-124.

Betts, S.C. 2003. Contingency Theory: Science or Technology? *Journal of Business & Economics Research*, 1, 123-130.

Beynon, M.J., Jones, P., & Pickernell, D. 2016. Country-based Comparison Analysis Using fsQCA Investigating Entrepreneurial Attitudes and Activity. *Journal of Business Research*, 69(4), 1271-1276.

Bromiley, P., & Rau, D. 2016. Operations Management and the Resource Based View: Another View. *Journal of Operations Management*, 41, 95-106.

Cao, M., Vonderembse, M.A., Zhang, Q., & Ragu-Nathan, T.S. 2010. Supply Chain Collaboration: Conceptualisation and Instrument Development. *International Journal of Production Research*, 48(22), 6613-6635.

Dwyer, F.R., Schurr, P.H., & Oh, S. 2004. Developing Buyer-Seller Relationships. *Journal of Marketing*, 51, 11-27.

Daughtert, P.J., Richey, R.G., Roath, A.S., Min, S., Chen, H., Arndt, A.D., & Genchev, S.E. 2006. Is Collaboration Paying Off for Firms? *Business Horizons*, 49(1), 61-70.

Dwyer, F.R. & Oh, S. (1987). Output Sector Munificence Effects on the Internal Political Economy of Marketing Channels. *Journal of Marketing Research*, 24(4), 347-358.

Dyer, J.H., & Singh, H. 1998. The Relational View: Cooperative Strategy and Sources of Interorganisational Competitive Advantage. *Academy of Management*, 23, 660-679.

Eyuboglu, N., Ryu, N., & Tellefsen, T. 2003. Current and Future Interdependence. *Journal of Marketing Channels*, 11(1), 3-26.

Fiss, P. 2011. Building Better Causal Theories: A Fuzzy Set Approach to Typologies in Organization Research. *Academy of Management Journal*, 54(2), 393-420. m

Fiss, P., Marx., A., & Cambre, B. 2013. Configuration Theory and Methods in Organisational Research: Introduction. In: *Configuration Theory and Methods in Organisational Research*, Emerald Group Publishing Limited.

Fleming, A., Ogier, E., Hobday, A.J., Thomas, L., Hartog, J.R., & Hass, B. 2020. Stakeholder Trust and Holistic Fishery Sustainability Assessments. *Marine Policy*, 111, Article 103719.

Ganesan, S. 1994. Determinants of Long-Term Orientation in Buyer-Seller Relationships. *The Journal of Marketing*, 58(2), 1-19.

Geyskens, I., steenkamp, J., & Kumar, N. 2006. Make, Buy, or Ally: A Transaction Cost Theory Meta-Analysis. *Academy of Management*, 48, 519-543.

Golicic, S.L. 2007. A Comparison of Shipper and Carrier Relationship Strength. *International Journal of Physical Distribution and Logistics Management*, 37(9), 719-739.

Golicic, S.L. & Mentzer, J.T. 2006. An Empirical Examination of Relationship Magnitude. *Journal of Business Logistics*, 27, 81-108.

Golicic, S.L., Foggin, J.H., & Mentzer, J.H. 2003. Relationship Magnitude and Its Role in Interorganisational Relationship Structure. *Journal of Business Logistics*, 24, 57-75.

Greckhamer, T., Furnari, S., Fiss, P., & Aguilera, R. 2018. Studying Configurations with Qualitative Comparative Analysis: Best Practices in Strategy and Organisation Research. *Strategic Organization*, 16(4), 482-495.

Griffith, D.A., Harvey, M.G., & Lusch, R.F. 2006. Social Exchange in Supply Chain Relationships: The Resulting Benefits of Procedural and Distributive Justice. *Journal of Operations Management*, 24(2), 85-98.

Hair Jr. J., Ringle, C., Sarstedt, M. 2011. PLS-SEM: Indeed a Silver Bullet. *Journal of Marketing Theory and Practice*, 19(2), 139-152.

Hair Jr. J., Sarstedt, M., Matthews, I., Ringle, C. 2016. Identifying and Treating Unobserved Heterogeneity with

FIMIX-PLS: Part I-Method. *European Business Review*, 28(1), 63-76.

Harland, C., Zheng, J., Johnsen, T., & Lamming, R.A. 2004. Conceptual Model for Researching the Creation and Operation of Supply Networks. *British Journal of Management*, 15, 1-21.

Hornibrook, S, Fearne, A., & Lazzarin, M. 2009. Exploring the Association Between Fairness and Organisational Outcomes in Supply Chain Relationships. *International Journal of Retail & Distribution Management*, 37, 790-803.

Kim, K.K., Park, S.H., Ryoo, S.Y., & Park, S.K. 2010. Inter-Organizational Cooperation in Buyer-Supplier Relationships: Both Perspectives. *Journal of Business Research*, 63(8), 863-869.

Kim, C.S., Dinwoodie, J., & Roh, S. 2020. Developing Measurement Scales of Collaboration in Shipping Logistics. *International Journal of Logistics Research and Applications*, 24(5), 495-510.

Kim, C.S., Dinwoodie, J., & Seo, Y.J. 2018. Inter-Firm Cooperation and Collaboration in Shipper-Shipping Company Relationships for Enhancing Sustainability. *Sustainability,* 10(10), 3714.

Kim, C.S. Roh, S., & Seo, Y.J. 2022. Development of Collaborative Spirit Indices: The Case of South Korea's Maritime Industry. *The Asian Journal of Shipping and Logistics*, 38, 84-98.

Kim, K.K., Park, S.H., Ryoo, S.Y., & Park, S.K. 2010. Inter-organisational Cooperation in Buyer-Supplier Relationships: Both Perspectives. *Journal of Business Research*, 63, 863-869.

Kent, R. 2015. *Analysing Quantitative Data: Variable-based and Case-based. Approaches to Non-experimental Datasets*. Sage.

Kock, N. 2022. WarpPLS User Manual: Version 8.0. ScriptWrap System, Laredo, TX.

Konovsky, M.A. Understanding Procedural Justice and Its Impact on Business Organizations. 1995. *Journal of Management*, 26, 489-511.

Korsgaard, M.A., Schweiger, D.M., & Sapienza, H.J. 1995. Building Commitment, Attachment, and Trust in Strategic Decision-Making Teams: The Role of Procedural Justice. *Academy of Management*, 26(3), 489-511.

Kumar, G. & Banerjee, R.N. 2014. Supply Chain Collaboration Index: An Instrument to Measure the Depth of Collaboration. *Benchmarking An International Journal*, 21

Kumar, N., Scheer, L.K., & Steenkam, J.B.E. 1995. The Effects of Supplier Fairness on Vulnerable Resellers. *Journal of Marketing Research*, 32, 54-65.

Kuo, S.Y., Lin, P.C. & Lu, C.S. 2017. The Effects of Dynamic Capabilities, Service Capabilities, Competitive Advantage, and Organisational Performance in Container Shipping. *Transportation Research Part A: Policy and Practice*, 95, 356-371.

Kwon, I.W.G. & Suh, T. 2004. Factors Affecting the Level of Trust and Commitment in Supply Chain Relationships. *Journal of Supply Chain Management*, 40(1), 4-14.

Lane, C. 2000. Introduction: Theories and Issues in the Study of Trust. In *Trust Within and Between Organisations*, edited by C. Lane, and R. Bachmann, 1030. Oxford: Oxford University Press.

Laouiti, R., Haddoud, M., Nakara, W., & Onjewu, A. 2022. A Genger-based Approach to the Influence of Personality Traits on Entrepreneurial Intention. *Journal of Business Research*, 142, 819-829.

Lee, H.Y., Seo, Y.J., & Dinwoodie, J. 2016. Supply Chain Integration and Logistics Performance: The Role of the Supply Chain Dynamism. *The International Journal of Logistics Management*, 27(3), 668-685.

Legewie, N. 2013. An Introduction to Applied Data Analysis with Qualitative Comparative Analysis. In: *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, 14(3).

Mei, Z. & Dinwoodie, J. 2005. Electronic Shipping Documentation in China's International Supply Chains. *Supply Chain Management: An International Journal*, 10(3), 198-205.

Malhotra, A., Gosain, S., & Sawy, O.A.E. 2005. Absorptive Capacity Configurations in Supply Chains: Gearing for Partner-Enabled Market Knowledge Creation. *MIS Quarterly,* 29, 145-187.

Mason, R., & Nair, R. 2013. Strategic Flexibility Capabilities in the Container Liner Shipping Sector. *Production Planning and Control,* 24(7), 640-651.

Midoro, R., Musso, E., & Parola, F. 2005. Maritime Liner Shipping and the Stevedoring Industry: Market Structure and Competition Strategies. *Maritime Policy and Management*, 32(2), 89-106.

Ministry of Oceans and Fisheries (MoF) 2022. *Government's Internal Data*, Ministry of Oceans and Fisheries, Sejong, Korea.

Min, S., Roath, A.S., Daugherty, P.J., Genchev, S.E., Chen, H., Armdt, A.D., & Richey, R.G. 2005. Supply Chain Collaboration: What's Happening? *The International Journal of Logistics Management, 16(2), 237-256.*

Mohr, J. & Nevin, J.R. 1990. Communication Strategies in Marketing Channels: A Theoretical Perspective. *Journal of Marketing*, 54(4), 20-38.

Molm, L.D. 1991. Affect and Social Exchange: Satisfaction in Power-Dependence Relations. *American Sociological Review*, 56, 475-493.

Morgan, R.M., & Hunt, S.D. 1994. The Commitment-Trust Theory of Relationship Marketing. *Journal of Marketing,* 58, 20-38.

Narssirnia, P. & Robinson, R. 2013. Supply Chain Integration and Chain Efficiency: The Importance of Changing Perspective toward Whole Chain Integration. *Journal of Applied Sciences and Engineering Management*, 1(1), 13-23.

Neuman, W.L. 2014. *Social Research Method*, Pearson: Harlow, UK.

Notteboom, T., Rodrigue, J.P., & De Monie, G. 2010. The Organisational and Geographical Ramifications of the 2008-09 Financial Crisis on the Maritime Shipping Port Industries. In *Integrating Seaports and Trade Corridors,* edited by P. Hall, R. McCalla, C. Comtois and B. Slack, 31-46. Aldershot: Ashgate.

Nunkoo, R., & Ramkissoon, H. 2012. Power, Trust, Social Exchange and Community Support. *Annals of Tourism Research*, 56, 475-1023.

Nyaga, G.N. & Whipple, J.M. 2011. Relationship Quality and Performance Outcomes: Achieving a Sustainable Competitive Advantage. *Journal of Business Logistics*, 32(4), 345-360.

Ordanini, A., Parasuraman, A. & Rubera, G. 2014. When the Recipe is More Important than the Ingredients: A Qualitative Comparative Analysis (QCA) of Service Innovation Configurations. *Journal of Service Research*, 17(2), 134-149.

Organisation for Economic Co-operation and Development (OECD). 2001. *The Well-Being of Nations: The Role of Human and Social Capital*. Organisation for Economic Co-operation and Development: Paris, France.

O'Toole, T., & Donaldson, B. 2000. Relationship Governance Structures and Performance. *Journal of Marketing Management*, 16, 327-341.

Pappas, I. & Woodside, A. 2021. Fuzzy-set Qualitative Comparative Analysis (fsQCA): Guidelines for Research Practice in Information Systems and Marketing. *International Journal of Information Management*, 58, 102310.

Ragin, C. 2008. *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. University of Chicago Press, Chicago.

Ragin, C. 2009. Qualitative Comparative Analysis Using Fuzzy Sets (fsQCA). In: Rihoux, B. & Ragin, C. (Eds). *In Configurational Comparative Methods*. Sage, Thousand Oaks, CA.

Ragin, C. & Davey, S. 2016. Fuzzy-Set/Qualitative Comparative Analysis 3.0. University of California, California.

Ramanathan, U., & Gunasekaran, A. 2014. Supply Chain Collaboration: Impact of Success in Long-Term Partnerships. *International Journal of Production Economics*, 147, 252-259.

Ryu, S., Park. J.E., & Min. S. 2007. Factors of Determining Long-Term Orientation in Interfirm Relationships. *Journal of Business Research*, 60(12), 1225-1233.

Schlittgen, R., Ringle, C., Sarstedt, M., & Becker, J. 2016. Segmentation of PLS Path Models by Iterative Reweighted Regressions. *Journal of Business Research*, 69(10), 4583-4592.

Schneider, C. & Wagemann, C. 2010. Qualitative Comparative Analysis (QCA) and Fuzzy-Sets: Agenda for a Research Approach and a Data Analysis Technique. *Comparative Sociology*, 9(3), 376-396.

Seo, Y.J., Dinwoodie, J., & Kwak, D.W. 2014. The Impact of Innovativeness on Supply Chain Performance: Is Supply Chain Integration a Missing Link? *Supply Chain Management: An International Journal*, 19(5/6), 733-746.

Simatupang

Slack, B., Comtois, C., & McCalla, R. 2002. Strategic Alliance in the Container Shipping Industry: A Global Perspective. *Maritime Policy and Management*, 29(1), 65-76.

Spekman, R.E., Kamauff, J.W., & Myhr, N. 1998. An Empirical Investigation into Supply Chain Management. *Supply Chain Management*, 3, 53-67.

Stopford, M. 2009. *Maritime Economics.* 3rd edition. New York: Routledge.

Talley, W.K., & Ng. M. 2013. Maritime Transport Chain Choice by Carriers, Ports and Shippers. *International Journal of Production Economics*, 142(2), 311-316.

Thomson, A.M., Perry, J.L., & Miller, T.K. 2009. Conceptualising and Measuring Collaboration. *Journal of Public Administration Research and Theory*, 19(1), 23-56.

United Nations Conference on Trade and Development (UNCTAD). 2023. *Review of Maritime Transport, 2023.* UNCTAD/RMT/2023. Geneva: United Nations Convention on Trade and Development.

Wang, C.I., Siu, N.Y.M., & Barnes, B.R. 2008. The Significance of Trust and Renqing in the Long-Term Orientation of Chinese Business-to-Business Relationships. *Industrial Marketing Management*, 37(7), 819-824.

Williamson, O.E. 1975. *Markets and Hierarchies, Analysis and Antitrust Implications: A Study in the Economics of Internal Organisation*, Free Press: New York, NY, USA.

Wong, T., Haddoud, M., Kwok, Y., He, H. 2018. Examining the Key Determinants Towards Online Pro-Brand and Anti-Brand Community Citizenship Behaviours: A Two-Stage Approach. *Industrial Management and Data Systems*, 118(4), 850-872.

Woodside, C., Brabenec, T., Gao, P., & Tang, Z. 2021. The Business Strategy, Competitive Advantage and Financial Strategy: A Perspective from Corporate Maturity Mismatched Investment, *Journal of Competitiveness*, 31(1), 164-181.

Woodside, A. & Zhang, M. 2012. Identifying X-Consumers Using Causal Recipes: “Whales” and “Jumbo Shrimps” Casino Gamblers. *Journal of Gambling Studies*, 28(1), 13-26.

24th International Conference on Group Decision and Negotiation &
10th International Conference on Decision Support System Technology
Human-Centric Decision and Negotiation Support for Societal Transitions

Multi-Method Approach Combining Predictive Analytics and Qualitative Inputs to Model Decisions of Non-Partisan Councils

Colin Williams [1,2*] and Liping Fang [2]

[1] Canadore College, North Bay, Canada
[2] Toronto Metropolitan University, Toronto, Canada
[*] c.williams@torontomu.ca

Abstract

Public policy issues are inherently complex, requiring various trade-offs between competing priorities. Quantitative approaches like predictive analytics offer new possibilities for comprehending individual and collective political decisions. This study investigated the efficacy of layered quantitative approaches for modelling voting behaviours of democratic institutions regarding public policies. An analytical model combining the random forest algorithm with statistical regression and clustering analyses is developed in this study. Using a non-partisan municipal council in the City of Toronto as a case study, over 7,000 voting records spanning 2006 to 2022 were analyzed. The proposed model demonstrated high predictive accuracy, correctly classifying 89.5% of votes. Four distinct voting blocs emerged, paralleling the political spectrum commonly found in Canada. The presented model enables strategically designing policy proposals that may achieve a majority support. This study provides public administrators with an evidence-based means to advance policies, e.g., climate action, that require decisive political responses. Overall, predictive analytical techniques offer insight into decision-making mechanics within democratic systems, improving accountable and effective policy formulation and implementation.

Keywords: Predictive analytics, public policy, voting behaviour, and political decision-making

1. Introduction

Climate change represents an existential threat to our livelihood, necessitating policy responses from public and government institutions (Herndon, 2018). However, enacting impactful environmental policies has proven challenging, with many political and public figures downplaying or ignoring scientific evidence (Fischer, 2019). Overcoming these barriers to decisive climate action requires the application of new policy tools that enable public administrators to comprehensively understand and act effectively in government.

Traditional policy modelling prioritizes financial projections, often appearing to neglect equally relevant social, cultural, and political factors (Luderer et al., 2016). Predictive analytics provides policy tools that look beyond superficial variables and comprehensively comprehend complex decisions (Wamba et al., 2017). This would, in turn, enable public administrators to successfully craft policy recommendations for adoption and implementation within democratic institutions (Homsy, 2018). Public administrators' knowledge of their respective subject matter could be more positioned to develop policy recommendations with an increased likelihood of adoption if appropriately applied. Nonetheless, the ethical aspects of such policy tools must also be considered.

The application of policy tools, such as predictive analytics, to potentially influence politicians' voting behaviours raises ethical concerns about the appropriate role of public administrators. As Busuioc (2021) asserted, transparency alone does not ensure accountability when algorithms shape government decisions.

Public administrators must provide political officials with sound justifications rooted in expertise and the public interest. Public administrators are fallible and may produce misguided policy recommendations. The model presented may enable administrators to manufacture political consensus through data mining and behavioural trend analysis. This highlights the need for thoughtful oversight and democratic checks on attempts to engineer or control political outcomes. Analytics should inform, not subvert, democratic deliberation on the public good.

As Desouza and Jacob (2017) identified, predictive analytics offers significant yet underutilized possibilities to elucidate the factors, including psychological, driving political decision-making on policy matters. While big data techniques have transformed private sector fields, applications in the public sphere remain nascent (Mergel et al., 2016). The Theory of Planned Behaviour (TPB) was employed as a theoretical framework to understand and predict the voting behaviour of political officials on environmental issues. According to the TPB, an individual's behaviour is shaped by three factors: attitude, subjective norms, and perceived behavioural control (Ajzen, 1991). It is asserted that political officials can be treated as individual decision-makers regarding their votes on policy, and therefore, TPB applies in understanding their voting choices. While subjective norms and attitudes are difficult to mine directly from data, political officials' past voting records on similar issues were used to indicate these variables. It is posited that political affiliation and past voting records may constitute predictors in modelling the voting behaviour of Councillors. Additionally, Keeney's (1996) value-focused thinking is directly incorporated, which suggests that individuals engage in decision-making by identifying alternatives aligned with their values. Value-focused thinking complemented the TPB by explaining how values may drive the decision to support or oppose environmental policies.

The efficacy of the proposed approach in this study could be evaluated by focusing on layering analytical approaches and incorporating machine learning and predictive modelling for explicating the mechanics of official political voting decisions. It was hypothesized that predictive analytics could achieve a high level of accuracy in modelling voting behaviours of non-partisan governing bodies, irrespective of the subject matter. While this study focused on a single non-partisan council in the City of Toronto, Canada, the approach demonstrated that it could be applied to other democratic institutions without political affiliations.

In a "post-truth" world, predictive analytics appears to offer a powerful tool for developing democratically accountable, empirically justifiable public policies. Predictive analytics is positioned to provide data-driven comprehension of political decision-making, enabling public administrators to craft policies that gain majority political support. This study layered psychological and predictive analytical techniques to derive an evidence-based approach to enable public institutions via their public administrators to craft policy recommendations that are efficacious and likely to be adopted – transforming the political-administrative dichotomy.

2. Methods

This study employed a mixed methods approach, leveraging quantitative predictive analytics and machine learning algorithms to model voting behaviours, then supplemented with qualitative insights from political officials. The specific techniques utilized include random forest modelling, regression analysis, hierarchical clustering, and algorithmic modelling to analyze over 7,000 voting records of the Toronto City Council spanning 2006 to 2022.

The population comprised all Toronto City Council voting members over this period. As Mergel et al. (2016) discussed, public policy research requires extensive data gathering across long timescales, hence the lengthy study period enabling analysis of trends and clusters. Records from the publicly available Toronto Open Data Portal contained detailed accounts of every policy proposal and member's votes (Hardy & Maurushat, 2017). Rather than surveys prone to bias, actual voting history provided an objective measure of politicians voting decisions. The votes analyzed pertained to environmental proposals, classified based on content analysis of proposal titles and texts. The dependent variable was a binary outcome of whether the proposal passed or failed. Independent variables included policy type, content, timing, financial impacts, and member voting history on related measures.

- *Policy type*: The type or category of policy being voted on, such as climate, financial, public health impacts, land use, transportation/roads, etc.

- *Content*: This variable captures the environmental policy proposal's specific content or subject matter.
- *Timing*: This refers to the electoral cycle or year the policy vote occurred.
- *Financial impacts*: This variable measured whether the proposed environmental policy had financial impacts or costs for the government.
- *Member voting history on related measures*: This variable represented each council member's past voting record on previous, related environmental policy proposals to account for their predisposition on environmental issues.

First, random forest modelling was used to identify significant predictors and clustering algorithms to discern voting blocs amongst members. The former addressed overfitting limitations of decision trees via ensemble modelling aggregating predictions across hundreds of trees (Probst et al., 2019). Second, clustering analysis enabled computational pattern recognition to categorize high-dimensional data (Vogl et al., 2020). Analysis was performed using RStudio and Python packages for machine learning and statistical analysis. The methodology uniquely combined the rigour of data science with qualitative validation and framing per scholars Desouza and Jacob (2017). Ten semi-structured interviews were conducted with long-time politicians to assist in interpreting the statistically derived behavioural groupings.

3. Analysis

The dataset contained 7,137 total votes across the 16-year study period. The random forest model comprised 500 constituent decision trees. Predictor importance was assessed using mean decreases in accuracy and Gini coefficient. The trained model was validated on a test set of 30% of voting records, which achieved 89.5% predictive accuracy on whether a given proposal passed or failed. The mean decrease accuracy and Gini importance plots identified policy content and timing as the most influential predictors of voting outcomes. This aligned with the theoretical framework emphasizing perceived attitudes and norms as behavioural drivers (Ajzen, 1991). Content analysis of proposal texts enabled partial quantification of attitudes based on policy focal points.

Hierarchical clustering of Council members' voting records identified four distinct blocs visibly separated in the dendrogram. The four clusters were characterized as progressive, moderate, centrist, and conservative per the interview insights from long-serving Councillors. They reflected different priorities and political ideologies regarding environmental policy's appropriate scope and prioritization. Cluster membership significantly predicted voting behaviours. For instance, progressive bloc members voted affirmatively on climate change or sustainability proposals at rates 17% higher than moderate counterparts and 51% higher than conservatives. Yet, the overlap between the clusters was non-negligible, suggesting room for persuasion on specific proposals. Thirteen percent (13%) of climate change proposals garnered unanimous support, while 41% passed with only one or two dissenting votes. This indicates framing and tailored justification of policies can overcome some partisan differences. Even among the conservative bloc, outright dismissal of environmental proposals was rare, with only 4% voting negatively as a bloc.

4. Findings and Remarks

This study demonstrated the potential of layering quantitative techniques like computational predictive analytics, machine learning, statistical modelling, and data mining to achieve explanatory and predictive insights into the complex, multifaceted drivers underlying political voting decisions and public administrators' policymaking on critical issues such as climate action. While quantitative methods are increasingly used in public policy research, this work uniquely integrates advanced computational analytics with theoretical framing to model voting behaviours rather than just focusing on attitudes or policy outcomes, as others have done (Vogl et al., 2020). The results align with Desouza and Jacob's (2017) contention that big data analytics need to be more utilized in the public sector despite their immense potential to elucidate the nuanced intricacies of policymaking processes. By computationally analyzing revealed preferences in voting records, the study provides an empirically grounded template for transforming policy design from an intuitive art into a data-driven science. Findings indicate bipartisan majority support may be achievable even on contentious environmental issues when policy proposals are strategically tailored through data-driven insights into majority values and priorities (Homsy, 2018).

The combined predictive accuracy exceeding 89.5% demonstrates significant promise for generalizing this approach across contexts and supports the notion that political processes in democratic systems follow quantifiable rules driven by internal values and priorities rather than complete randomness. This supports Keeney's (1996) theory on the role of intrinsic values in decision-making. From a practitioner standpoint, the ability to predictively model politicians' voting decisions on proposed policies could revolutionize policy design by enabling recommendations to be empirically crafted to align with majority values uncovered through quantitative analysis rather than relying on intuitions alone. The cluster analysis unveiling four distinct voting blocs reflecting a left-to-right ideological spectrum, observed through revealed preferences in voting records rather than stated party affiliations, provides the most significant finding. While conservative members exhibited the highest skepticism of environmental proposals, they were often still amenable, highlighting the study's nuanced insights. Quantitative modelling surfaces such obscured thematic patterns by harnessing the untapped potential of open government data combined with advanced predictive analytics techniques (Puussaar et al., 2018). This model provides an accountability mechanism quantifying political responsiveness to evidenced priorities versus other influences (Mergel et al., 2016).

The methodology demonstrated public policy development can be transformed from an art to more of a science through data-driven insight into multifaceted political priorities and leanings. The study also illustrates the vast possibilities of open government data to drive accountability and transparency when intelligently mined (Puussaar et al., 2018). For scholars, the methodology provides a template for conceptually sound quantification of the many interacting facets shaping policy decisions. Quantitative rigour is complemented by a philosophically grounded framing of motivations and values. Further refinements to improve predictive accuracy and generalization represent a promising new research domain at the intersection of data science and governance.

The findings further indicate that strategic policy design grounded in data-driven comprehension of politicians' values and priorities may overcome barriers to bipartisan solutions for collective challenges. Even in polarized contexts, majorities are achievable when proposals are tailored to substantiated priorities and constituencies. This provides a template for developing public policies with the urgency which current environmental threats demand. While focused on a municipal environmental case study, the method can be generalized to other non-party affiliated democratic institutions. It provides scholars with an analytical framework to complement ideological assumptions with empirical validity regarding motivations. For practitioners, it offers a mechanism to pre-emptively discern proposals' likelihood of majority support so that policies achieve outcomes aligned with public interests.

This study did not directly explore the applicability of this method to partisan councils where party affiliation could be straightforwardly used as an input variable. However, the findings suggest a promising avenue for future research. The delineation of four distinct voting behaviour groups within the Toronto City Council, which to some extent align with broader political spectrums, indicates that the methodology could be adapted to partisan environments. In such contexts, party affiliation might be a strong predictor, potentially simplifying the model's complexity while retaining its predictive power. Future studies could explore incorporating party affiliation as an input variable in random forest models to examine its impact on predictive accuracy in partisan councils.

References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Busuioc, M. (2021). Accountable artificial intelligence: Holding algorithms to account. *Public Administration Review*, *81*(5), 825–836. https://doi.org/10.1111/puar.13293

Desouza, K. C., & Jacob, B. (2017). Big data in the public sector: Lessons for practitioners and scholars. *Administration & Society*, *49*(7), 1043–1064. https://doi.org/10.1177/0095399714555751

Fischer, F. (2019). Knowledge politics and post-truth in climate denial: On the social construction of alternative facts. *Critical Policy Studies*, *13*(2), 133–152. https://doi.org/10.1080/19460171.2019.1602067

Hardy, K., & Maurushat, A. (2017). Opening up government data for big data analysis and public benefit. *Computer Law & Security Review*, *33*(1), 30–37. https://doi.org/10.1016/j.clsr.2016.11.003

Herndon, J. M. (2018). Air pollution, not greenhouse gases: The principal cause of global warming. *Journal of Geography, Environment and Earth Science International*, *17*(2), 1–8. https://doi.org/10.9734/JGEESI/2018/44290

Homsy, G. C. (2018). Unlikely pioneers: Creative climate change policymaking in smaller U.S. cities. *Journal of Environmental Studies and Sciences*, *8*(2), 121–131. https://doi.org/10.1007/s13412-018-0483-8

Keeney, R. L. (1996). Value-focused thinking: Identifying decision opportunities and creating alternatives. *European Journal of Operational Research*, *92*(3), 537–549. https://doi.org/10.1016/0377-2217(96)00004-5

Luderer, G., Bertram, C., Calvin, K., De Cian, E., & Kriegler, E. (2016). Implications of weak near-term climate policies on long-term mitigation pathways. *Climatic Change*, *136*(1), 127–140. https://doi.org/10.1007/s10584-013-0899-9

Mergel, I., Rethemeyer, R. K., & Isett, K. (2016). Big data in public affairs. *Public Administration Review*, *76*(6), 928–937. https://doi.org/10.1111/puar.12625

Probst, P., Wright, M., & Boulesteix, A.-L. (2019). Hyperparameters and tuning strategies for random forest. *WIREs Data Mining and Knowledge Discovery*, *9*(3). https://doi.org/10.1002/widm.1301

Puussaar, A., Johnson, I. G., Montague, K., James, P., & Wright, P. (2018). Making open data work for civic advocacy. *Proceedings of the ACM on Human-Computer Interaction*, *2*(CSCW), 143:1-143:20. https://doi.org/10.1145/3274412

Vogl, T. M., Seidelin, C., Ganesh, B., & Bright, J. (2020). Smart technology and the emergence of algorithmic bureaucracy: Artificial intelligence in UK local authorities. *Public Administration Review*, *80*(6), 946–961. https://doi.org/10.1111/puar.13286

Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, *70*, 356–365. https://doi.org/10.1016/j.jbusres.2016.08.009

Group Decision-Making Process for Cyberattack Mitigation

Andrej Bregar[1], Anas Husseis[2], and Jose Luis Flores[2]

[1] Informatika d.o.o., Vetrinjska ulica 2, 2000 Maribor, Slovenia
[2] IKERLAN, Po. J. Ma. Arizmendiarrieta, 2 – 20500 Arrasate-Mondragón, Gipuzkoa, Spain
*andrej.bregar@informatika.si, ahusseis@ikerlan.es, jlflores@ikerlan.es

Abstract

This paper addresses the strategic necessity of proactively managing cyber threats through effective mitigation measures, considering their benefits and costs. Cen-tral to our approach is a group decision-making process that navigates the com-plex interplay between different assets and the varied perspectives of stakeholders at multiple organizational tiers. Our process unfolds in two stages: assessing the impacts of cyber threats on assets and then identifying and selecting optimal miti-gation measures. This method integrates the Delphi technique for the consensus of decision-makers and a versatile Multi-Criteria Decision-Making (MCDM) model, which accommodates both quantitative and qualitative evaluations and in-cludes a veto function to ensure the effectiveness of chosen mitigations. The model adheres to standards from the CIA, ISO/IEC 15408, and NESCOR frameworks. Further, we develop a comprehensive mitigation framework, draw-ing on methodologies from MITRE ATT&CK, CIS Critical Security Controls, and ACSC strategies, and implement a decision support system that encompasses the entire decision process, enhancing analytics through sensitivity analysis and integration with external data sources.

Keywords: Multi-Criteria Decision Analysis; Group Decision-Making; Decision Support Systems; Decision Processes; Value Theory; Mitigation Strategies; Cybersecurity.

1. Introduction

Cybersecurity's growing importance in IT and OT systems, critical infrastructures, and enterprises is driven by the intensifying digitalization, sophisticated cyber-attacks, and geopolitical risks [21, 28]. Strategies to bolster IT and OT resilience include both proactive risk, threat, and vulnerability estimation and reactive approaches using systems like SIEM (Security Information and Event Management) and SOAR (Security Orchestration, Automation and Response) for continuous monitoring and response. The challenge lies in selecting and implementing optimal, cost-effective mitigation measures, considering infrastructure characteristics, technology maturity, incident severity, and organizational objectives. As Staves [29] points out, in the assessment of the effectiveness of current guidance for ICS (Industrial Control Systems) cybersecurity response and recovery, relevant requirements must first be identified. As a result, four primary phases are identified, of which the planning is the first step involving roles and responsibilities, response planning, criticality assessment, threat analysis, and risk management. In addition to this, Smith [30] also considers the necessity to include disparate business elements, such as engineers or business analysts, in the incident response team to facilitate knowledge exchange between cybersecurity professionals. This implies that the criticality and diversity of various aspects that need to be considered necessitates a group decision-making approach which address-es multiple levels: information system, business process, and organization, involving roles from Chief Information Officers (CIOs) and Chief Information Security Officers (CISOs) to business analysts.

To enhance the analysis of cyber vulnerabilities and mitigation measures, it's crucial to integrate best practices and recommendations from recognised sources such as MITRE ATT&CK [19], CIS [7], ACSC [26], EPRI & NESCOR [2], and NVD [20]. Our goal is to develop a comprehensive mitigation framework and a systematic decision-making methodology for cost-effective strategy selection based on multi-criteria analysis. This approach aims to address the limitations of current methodologies, which often overlook the complexities of group decision-making and multi-criteria considerations.

Existing methods for impact assessment, like the risk matrix [23] and CVSS [10], are either too simplistic or rigid. Optimization models [18, 22, 25] focus on cost and performance but fail to consider multiple criteria impacts. Approaches like EPRI's cybersecurity insights [2] and Bayesian networks [18] provide valuable frameworks, yet they still don't facilitate group decision-making across organizational levels and roles. Alshawish & de Meer's game theory-based method [1] offers strategic decision-making support but lacks a holistic MCDM approach and necessary decision-support tools.

This paper introduces a systematic, comprehensive group decision-making methodology built on multi-criteria analysis, aiming to effectively respond to cyber threats in complex IT/OT systems. Originating from the Horizon 2020 CyberSEAS project [11] and extending previous work by Bregar et al. [6], this methodology is coupled with a decision support system, operationalizing the group MCDM process. It's de-signed primarily for the energy sector, but adaptable for various ecosystems, addressing the need for an integrated, multi-criteria, and collaborative decision-making framework in cybersecurity.

The rest of the paper is organized as follows. Section 2 presents the mitigation framework and the integrated external information sources. In Section 3, we propose the group decision-making model for the impact assessment and mitigation selection. Section 4 defines the MCDM methodology which is based on the value theory, the Delphi technique, and a standardized scoring system to equivalently support cardinal and ordinal scales. Section 5 discusses some implications and limitations of the introduced approach. Finally, Section 6 concludes the paper.

2. Mitigation Framework

The Mitigation Framework forms the cornerstone for selecting and executing effective cyber defence measures. MITRE ATT&CK is the primary source in our approach, offering a comprehensive, routinely updated online repository for Tactics, Techniques, and Procedures (TTPs) [19]. It provides standardized mitigations to counteract TTPs and bolster cyber resilience. However, given the diversity of assets in complex infrastructures and varying providers, it's essential to integrate other sources like CIS Critical Security Controls (CSCs) [7], ACSC mitigation strategies [26], vendor recommendations, and internal countermeasures for more exhaustive coverage.

CSCs offer 18 distinct controls to fortify cybersecurity, addressing areas like patch management and incident response planning. Similarly, ACSC mitigation strategies provide detailed defence recommendations based on expertise in dealing with real-world cyber-attacks and the analysis of the latest techniques and trends utilized by attackers. The ACSC framework investigates five distinct mitigation categories that deal with different aspects of cybersecurity and provides detailed recommendations for effective defence implementation.

Our comprehensive approach merges these recognized sources into a tailored classification of mitigation countermeasures, ensuring broad and relevant coverage for the infrastructure under protection. At the same time, the approach remains fully aligned with the infrastructure we are protecting. For instance, Table 1 demonstrates mitigations against phishing and malware, derived from ACSC strategies, with each mitigation assessed for its impact and cost-effectiveness. A mitigation measure is appropriate for implementation only if 1) it limits the estimated or measured impact of a detected cybersecurity incident, and 2) its impact achieves or exceeds a prede-fined level because its cost would not justify the benefits otherwise.

Table 1. Examples of mitigations for the prevention of phishing and malware.

Mitigation	Mitigation description and use	Impact	Effect
Full visibility of installed software	Obtain a detailed visibility of what software is installed on computers by using a standard operating environment, maintaining an inventory of software installed, and implementing a robust change management process	Mediu m	Improved visibility of potentially compromis ed software
Endpoint protection and anti-malware	Use endpoint protection and anti-malware software, with added application control functionality	High	Blocked delivery

In the decision process, we analyse potentially suitable mitigation measures based on vulnerable assets and the identified cyber incidents or threats, respectively. The mitigation framework therefore defines a set of rules and mappings to obtain the initial set of mitigations from which the decision-makers can select the most efficient ones for implementation. The set of mitigation countermeasures, attack techniques, detected cyber-attacks, and mappings of assets hence represents the knowledge repository of the decision support system.

Figure 1 shows the flow of actions and information to build the initial set of all potentially useful mitigation countermeasures according to the compromised assets and detected cyber-attacks targeting these assets. In the next phase, we apply the decision support system and the underlying MCDM methodology to assess the efficiency of available mitigations and select them for implementation.

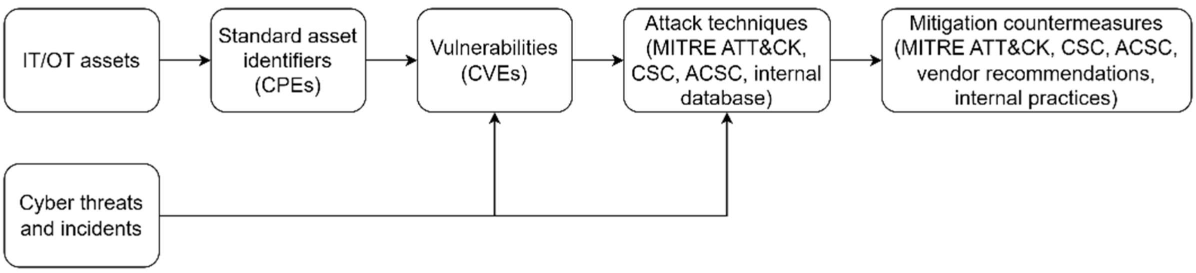


Figure 1: Identification of potential mitigation countermeasures based on compromised assets.

The identification process starts with defining the Common Pattern Enumeration (CPE) of a vulnerable IT/OT asset [8], facilitating the search for known vulnerabilities in the NVD database [9]. Most vulnerabilities are assigned CVSS scores [10], a stand-ard method for assessing severity based on base, temporal, and environmental metrics. However, due to its generality, CVSS scores alone are inadequate for directly reflecting the specific ecosystem characteristics. Hence, we map identified vulnerabilities and incidents to standard attack techniques in the MITRE ATT&CK database, which are then aligned with mitigation countermeasures. These countermeasures, evaluated by the mitigation selection decision model, include additional mappings from complementary sources like CSC and ACSC, ensuring a thorough and context-specific selection process.

3. Group Decision-Making Process

The group decision-making process in our approach is a synthesis of the decision-making process and the MCDM methodology, supported by a decision support system. This process adheres to the four-phase general decision-making model of intelligence, design, choice, and implementation [27], with the intelligence phase grounded in our mitigation framework for data collection and problem identification. The other phases are covered with the MCDM methodology.

The proposed decision-making process focuses on countermeasures to mitigate cybersecurity incidents and threats. Therefore, it must also align with two other gen-eral processes. The first is the risk management process as standardized by ISO 31000 [15]. It covers risk identification, analysis, evaluation, and treatment. Second-ly, because the decision-making process results in the selection of countermeasures that are triggered to mitigate detected incidents, it also impacts the overall incident response process. It can be noticed that it aligns well with the OODA (Observe, Orient, Decide, and Act) loop [13]. In this incident response loop, the observe phase pertains to the use of security monitoring to identify anomalous behaviour, the orient phase evaluates what is going on in the cyber threat landscape, the decide phase refers to the selection of the best tactic for minimal damage according to observations, and the act phase triggers and improves the appropriate incident response procedures.

Figure 2 illustrates the alignment of four processes. It underscores the need for harmonious incorporation of the incident impact assessment MCDM model, the mitigation selection MCDM model, and the mitigation framework, along with external in-formation sources that support the intelligence phase. Figure 3 details the overall decision-making process, emphasizing the need for collaboration across organizational levels to ensure effective decision-making.

	Decision-making process	Risk management process	IR process	Implemented process
Intelligence	Define requirements Data collection Analyze information Problem identification	Establishing the context Risk identification	Observe	Cybersecurity data collection Enrichment and analysis of attack indicators Identification of incidents Identification of compromised assets and vulnerabilities
Design	Identify alternatives Set criteria for choice Formulate model Identify criteria importance	Risk analysis	Orient	Identification of mitigations ==> Alternatives Criteria specification for impact assessment Criteria specification for mitigation assessment Criteria weighting for impact assessment Criteria weighting for mitigation assessment
Choice	Solution to the model Validation and tests Sensitivity analysis Plan for implementation	Risk evaluation	Decide	Assessment of impacts of incidents ==> Impact scores Assessment of mitigations ==> Mitigation scores Aggregation of impact, severity, and risk scores Sensitivity analysis for impact assessment Sensitivity analysis for mitigation assessment Mitigation selection
	Implementation of solution	Risk treatment	Act	Mitigation implementation

Figure 2. Alignment of processes.

Central to the decision-making process are cybersecurity events, characterized as groups of attack techniques from the MITRE ATT&CK framework, such as account theft using various discovery and cracking techniques. The process involves evaluating the impact of these events on compromised assets and the efficiency of potential countermeasures, considering both individual actions and comprehensive remediation strategies.

The proposed decision-making process adopts a multi-tiered risk management approach suggested by NIST [24], encompassing organization, mission/business pro-cess, and information system levels. In the first phase, security experts evaluate countermeasures' technical requirements, dependencies, and impact on information/service confidentiality, integrity, and availability (CIA), following ISO/IEC 15408 Common Criteria [16]. The second phase involves business analysts assessing the impact of mitigation strategies on organizational metrics like RTO (Recovery Time Objective) and RPO (Recovery Point Objective). Figure 4 shows the two-phase evaluation model, integrating business process and information system levels. In this pro-cess, the maturity level of the organization will have a significant influence on the results. This aspect is emphasized in preliminary processes, specifically in the preparation step for incident response plans (NIST [31] or SANS [32]). In those organizations where the maturity level is very high, the second phase can be partially carried out in an intertwined way accelerating the process, whereas in immature organizations, the process will be practically sequential. In other words, the greater the maturity, the greater the acceleration of the process.

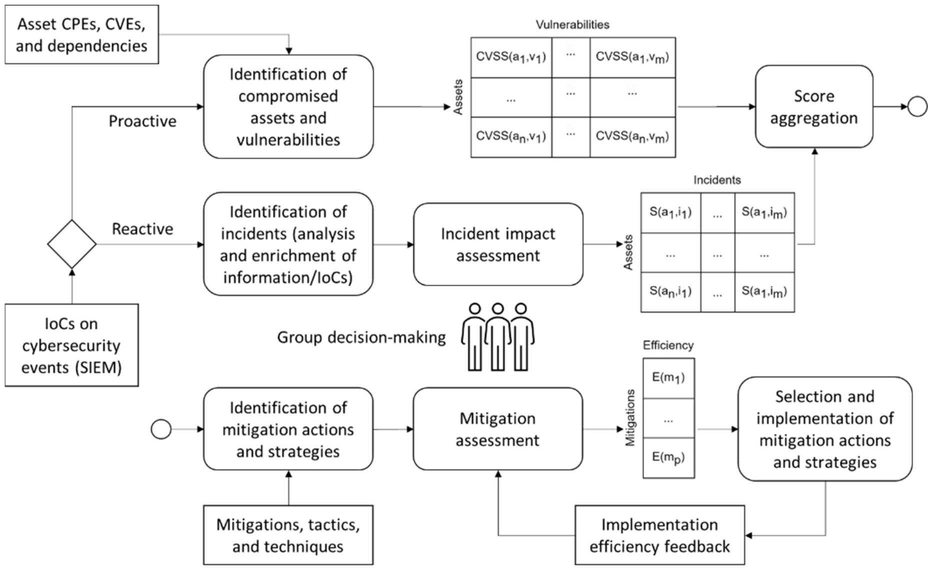


Figure 3. Decision-making process.

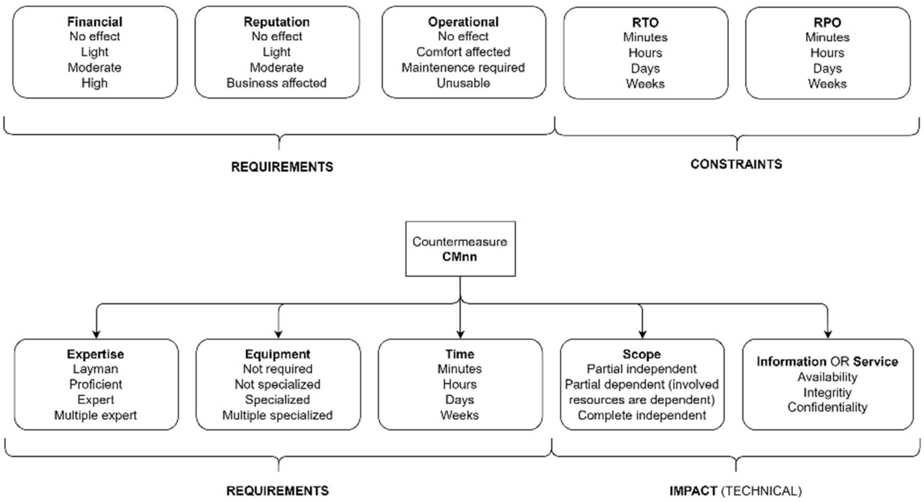


Figure 4. Two-phase evaluation model incorporating the business process level (above) and the

The mitigation selection model, detailed in Figure 5, incorporates NESCOR criteria for high-risk failure scenarios [2], asset criticality, and impact assessment based on CVSS scores and SIEM severity magnitude. This structured approach ensures systematic decision-making and involves collaboration at all organizational levels, leading to better, more implementable decisions. Senior executive staff play a key role in ap-proving decisions and allocating resources for implementing mitigation measures.

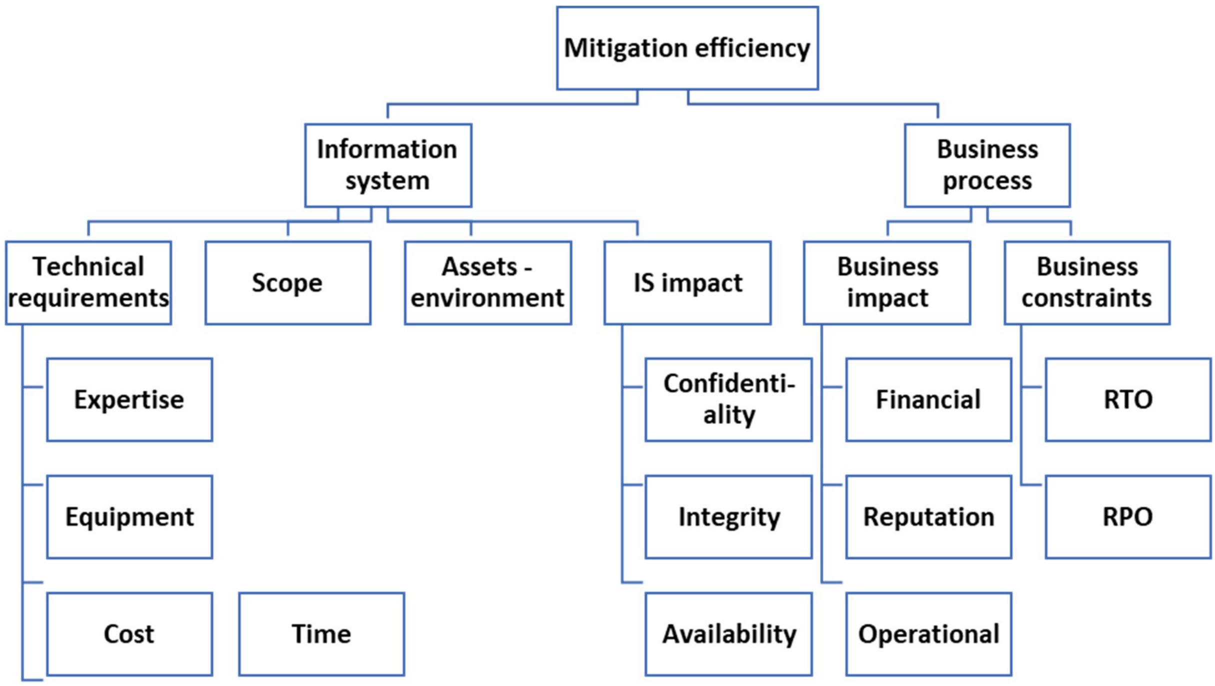


Figure 5. Mitigation selection criteria.

4. MCDM Analysis

The MCDM methodology is designed to be as suitable as possible for cybersecurity experts. For this purpose, we introduced a uniform Countermeasure Scoring System (CMSS), which is aligned with the standard CVSS system and puts cardinal and ordi-nal MCDM approaches on a common denominator. It uses a colouring scheme and several impact levels as defined in Table 2. The interpretation is that the higher the score, the worse it is because it reflects a stronger impact of a cyber threat or incident. A mitigation measure can reduce this impact. Its score is then interpreted as the im-pact of the cyber threat or incident that remains after the mitigation is implemented.

Table 1. Standard countermeasure scoring system.

None	Negligible	Low	Medium	High	Critical
0.0	0.1 to 1.0	1.1 to 4.0	4.1 to 7.0	7.1 to 9.0	9.1 to 10.0

The MCDM methodology and DSS support, in a transparent way, both the quantita-tive assessment (e.g., MAVT/MAUT [17]) and the qualitative assessment (e.g., DEXi [12]). In the quantitative approach, the additive value function is used as a simple aggregation method to make the decision model appropriate and comprehensive for security experts without a background in decision theory. The overall impact score of the lth incident and the overall efficiency score of the kth mitigation, respectively, are calculated with the weighted sum:

$$s^I(A_l) = \sum_{j=1}^{n} w_j s_j^I(A_l) \tag{1}$$

$$s^E(M_k) = \sum_{j=1}^{n} w_j s_j^E(M_k) \tag{2}$$

Here, $s^I(A_l)$ denotes the impact score of the lth incident, $s^E(M_k)$ is the efficiency score of the kth mitigation countermeasure, and w_j represents the weight of the jth criterion. This approach integrates the correlation between incident impact and mitigation efficiency, suggesting that the effectiveness of a mitigation countermeasure is inherently linked to the severity of the incident it aims to address.

Several forms of correlation may be used: mitigation suitability threshold, discordance/veto, or dynamic derivation of criteria weights. Of the three approaches, we model the veto function according to our previous methodology [3]. It opposes the efficiency of the mitigation countermeasure and reduces it if at least one of two conditions is met:

1. The mitigation does not make a sufficient improvement because the impact of a security-related event after its implementation is higher than the initial impact reduced according to the Mitigation Improvement Acceptability Threshold (MIAT). The effect of veto starts at this threshold and then increases linearly to the point of equality, where there is no improvement at all, so a strict veto occurs.

2. The mitigation performs very poorly on at least one assessment criterion, such that the scores of $C_{\max} - C_\delta$ or less are acceptable, the scores of $C_{\max}$ or more are unacceptable (strict veto), and the scores in-between result in a linear increase of veto (weak veto). Usually, $C_{\max}$ is set to 10 and C_δ is set to 2.

The effects of these two conditions are aggregated, such that their impact is distributed according to the parameter $\gamma \in [0,1]$. The definition of the veto function is hence:

$$v^E(M_k) = \gamma \begin{cases} 0 & , \quad s^E(M_k) \leq \mathrm{MIAT} \cdot s^I(A_l) \\ \dfrac{s^E(M_k) - \mathrm{MIAT} \cdot s^I(A_l)}{(1-\mathrm{MIAT})s^I(A_l)} & , \quad \mathrm{MIAT} \cdot s^I(A_l) < s^E(M_k) < s^I(A_l) \\ 1 & , \quad s^E(M_k) \geq s^I(A_l) \end{cases} \tag{3}$$

$$+(1-\gamma) \begin{cases} 0 & , \quad s^E(M_k) \leq C_{\max} - C_\delta \\ \dfrac{s^E(M_k) - (C_{\max} - C_\delta)}{C_\delta} & , \quad C_{\max} - C_\delta < s^E(M_k) < C_{\max} \\ 1 & , \quad s^E(M_k) \geq C_{\max} \end{cases} \tag{4}$$

In eq. (4), $C_{\max}$ is the veto threshold at which the measured mitigation efficiency for any criterion becomes unacceptable regardless of the performance on other criteria. A weak veto can also be opposed to the mitigation countermeasure starting to take effect at the value of $C_{\max} - C_\delta$, where C_δ is the length of the interval of the increasing veto degree.

Based on the veto function, the effective mitigation score is now calculated:

$$S^E(M_k) = s^E(M_k)\left(1 - v^E(M_k)\right) \tag{5}$$

Several types of sensitivity and robustness analyses are applicable [5]. One of them is the LP-based multi-dimensional robustness analysis. It searches for a minimal change of the weight vector for which an observed mitigation countermeasure gets reassigned into a chosen better or worse category according to the standard CMSS limits. The Euclidean distance minimizes the deviation. The definition of LP is as follows:

$$\Delta_w = \min \frac{\left[\sum_{j=1}^{n} |w_j - \widetilde{w}_j|^P\right]^{1/P}}{\Delta_w^{\max}} \text{ subject to} \tag{6}$$

$$s^E(M_k) = \sum_{j=1}^{n} w_j s_j^E(M_k) = C_q \tag{7}$$

$$\sum_{j=1}^{n} w_j = 1,0 \leq w_j \leq 1, \forall j = 1, \dots, n \tag{8}$$

Scores of individual decision-makers are aggregated into group assessments. As described in Section 4, there are two complementary cases of group decision-making:

1. Criteria on the organizational level are addressed by chief executives, such as CIO, CISO, and CTO. Business analysts make evaluations at the business process level. Criteria of the information system level are relevant primarily for security experts and technical IT/OT staff. However, CIOs, CISOs, and CTOs might also express their opinions regarding criteria at the information system and process levels.

2. In case dependent assets are compromised, several different stakeholders may own or manage these assets. These stakeholders should all provide assessments according to impact and mitigation criteria. Their opinions should be unified.

Because the proposed MCDM methodology unifies quantitative and qualitative decision-making, the Delphi technique [4] is used for group decision-making. It can deal both with numerical scores and ordinal categories. If a qualitative method (such as DEXi) is applied, individual and group assessments are expressed only in terms of categories. On the other hand, if we follow the above-defined MAVT-based quantitative approach, numerical scores on the $[0 \dots 10]$ scale are the primary result, but they are also directly mapped to corresponding ordinal categories based on the uniform CMSS scoring system.

In the Delphi process, judgments of individual group members are aggregated over several consecutive iterations (rounds) so that participants can modify and unify their opinions based on the provided feedback. Compiled statistical information allows each group member to analyze, reconsider, and improve personal judgements. In general, the Delphi process is continuously iterated until consensus is determined to be achieved, but in practice, it can be stopped after the third round because at this point the payoff usually begins to diminish.

One of the key characteristics of the Delphi method is statistical group response. Group opinion is defined as a statistical tendency of opinions provided by individual participants. Total unanimity of group members is not required, and a spread of individual opinions is a normal outcome of the final iteration. Because we apply Delphi to multi-criteria decision analysis based on numerical and qualitative preferences, statistical metrics present deviations of alternatives – mitigation countermeasures and cybersecurity events – according to the group of decision-makers. In this way, it is possible to indicate the common opinion of the group and the direction in which the group is heading. For both types of preferences – quantitative and qualitative –, the following measures are calculated:

$$\forall (CSE_i, A_j) \colon (C^I_{\min}, C^I_{\text{median}}, C^I_{\max}) \text{ for } \{DM_1, \dots, DM_m\} \tag{9}$$

$$\forall (CSE_i, A_j, M_k) \colon (C^E_{\min}, C^E_{\text{median}}, C^E_{\max}) \text{ for } \{DM_1, \dots, DM_m\}, M_k \in \{MA_k, RS_k\} \tag{10}$$

For each pair of the cybersecurity event CSE_i and the asset A_j compromised by this event, the best, worst, and median categories representing the impact of the event on the asset are calculated and presented to the group according to eq. (9) by considering the opinions of all decision-makers. The cybersecurity event CSE_i is characterized as a group of techniques T as follows:

$$CSE_i = \{\text{T1087.001}, \text{T1087.002}, \text{T1087.003}, \text{T1087.004}, \text{T1110.002}\} \tag{111}$$

In this example, all techniques are taken from the MITRE ATT&CK framework, such that the cybersecurity event refers to account theft, for which the attacker employs five complementary attack techniques: local, domain, email, and cloud account discovery, as well as password cracking. Due to cascading effects, an attack may compromise several connected assets, so the decision-makers should consider dependencies between these assets. Hence, CSE_i impacts the compromised assets $A_p, \dots, A_q$, where for each two assets it may hold that they are dependent (i.e., connected) or not:

$$(A_j D A_k) \vee \neg(A_j D A_k), \forall p \leq j < k \leq q, j \neq k \tag{112}$$

Different dependent or independent assets $A_p, \dots, A_q$ affected by CSE_i may be owned and managed by different stakeholders. In case there is more than one stakeholder, we are facing a group decision-making problem, such that all stakeholders must contribute to the assessment. The impact of CSE_i on A_j is assessed by each decision-maker individually. When the quantitative weighted value function is used, the impact score is expressed with $s^I(A_j)$ as defined in eq. (1). This score can be mapped into one of six qualitative categories $C_1, \dots, C_6$ (i.e., None to Critical) specified in Table 2. The Delphi statistics from eq. (9) presents to the decision-making group the lowest and the highest categories assigned by the two most discordant decision-makers and the median category calculated according to the entire group. In the case a qualitative method is applied, such as DEXi, the categories $C_1, \dots, C_6$ are directly obtained as the sole impact assessments.

A similar statistic is compiled in eq. (10) for the tuple where the mitigation countermeasure M_k is applied to remediate the impact of the event CSE_i on the asset A_j. The countermeasure can be either an independent mitigation action MA_k or a complex remediation strategy $RS_k = \{MA_1, \dots, MA_r\}$ comprising of a sequence of related mitigation actions. Again, categories are aligned with the uniform scoring system shown in Table 2. They are obtained based on the mapping of numeric mitigation efficiency scores $s^E(M_k)$ or directly with a qualitative method.

In the quantitative approach, we calculate and present to decision-makers additional measures of central tendency dealing with the highest, lowest, and average impact scores and mitigation efficiency scores, respectively:

$$\forall(CSE_i, A_j): (s^I_{\min}, s^I_{\text{avg}}, s^I_{\max}) \text{ for } \{DM_1, \dots, DM_m\} \tag{13}$$

$$\forall(CSE_i, A_j, M_k): (S^E_{\min}, S^E_{\text{avg}}, S^E_{\max}) \text{ for } \{DM_1, \dots, DM_m\}, M_k \in \{MA_k, RS_k\} \tag{14}$$

5. Implications and Limitations

Our methodology covers a wide range of aspects. In consequence, it has many implications in the organizations, policies, and other aspects. In the following sections, we will describe these implications, limitations, and challenges.

5.1. Implications

The methodology has implications for the policies, standards, and industry practices. Regarding cybersecurity policies, our methodology addresses two areas: business and technical. Without any of them, it is not possible to estimate the best response strategy. This aspect is completely aligned with NIST [31] and SANS [32] incident response plans. In consequence, several aspects emerge:

- The need emerges for a real and practical implementation of different cybersecurity policies which require the involvement of the company's business managers. Managers become an essential part of the response instead of being the customers of the actions of the IT team. Indeed, Smith [29] puts a finger on the wound by claiming that discipline is necessary when a remediation strategy is carried out.

- Business managers are now partially responsible for the mitigation strategy to increase the resilience of the company. IT personnel do not have a comprehensive knowledge of the business impacts and can make mistakes. The guidance of business managers cannot only help but also provide more security when assessing technical remediations.
- When implementing cybersecurity policies that involve quantitative or qualitative metrics these are approximate because in some cases, there is no previous history for estimating them. Our methodology requires that every metric involved is computed as accurately and realistically as possible.

In summary, an active involvement of all essential parties when implementing cyber-security policies is a must and not an option. It is a mechanism to measure the real involvement of the organization.

The implications of the framework for cybersecurity practice will also depend on the use of the information after applying the methodology. We explain how to carry out the process with many sources of information, but it is extremely relevant to col-lect the impacts of the selected strategy for improving the system. That is, this meth-odology can be used to show the volume of information processed, the type of in-formation, the set of strategies, the selected strategy, and the results after applying the strategy. This will generate not only the best practices but also the context of eve-ry strategy, that is, a more comprehensive best practice.

Our methodology can have a positive influence on different standards, in both ge-neric and specialized forms by providing specific practical methods and metrics. As such, we will show some exemplary implications for two standards:

- Specific standards: ISO/IEC 62443-4-1. This standard specifies the process requirements for the secure development of products used in industrial au-tomation and control systems. Specifically, in Practice 6 (Management of security-related issues) there is a process DM-4 (Addressing), which states that a process shall be employed for addressing security-related issues. The supplier shall establish an acceptable level of residual risk when determining an appropriate way to address each issue including different options, such as the creation of a remediation plan to fix problems and informing other pro-cesses of the issue or related issues.
- Generic standards: ISO 27035, Part 3: Guidelines for incident response oper-ations. This standard provides guidelines for efficient incident management, response, and IRT (Incident Response Team) operations. Specifically, there are different phases including the assessment and decision (5.2) and re-sponse (5.3). In the first phase, the proposed methodology can facilitate the process in two stages dealing with the assessment and initial decision through the PoC (Proof of Concept) and the assessment and incident confirmation by IRT. During the second phase, the methodology can provide specific da-ta to make decisions regarding immediate responses, later responses, and re-sponses to critical situations.

Industry practices can also take advantage of our methodology because it involves a comprehensive view of mitigation actions. In industry, it is a mistake to assume that in the selection of mitigation actions, there are typically two classes of profiles – IT personnel and business personnel. The methodology requires that the most suitable profiles are the ones who make decisions, and this, in the industrial field, involves the presence of OT personnel. The implications are manifold:

- In industry, there is a need for involvement of different profiles. Classical IT teams need to be advised by specialized industrial engineers for selecting and assessing any remediation action because they do not have specific knowledge of the operational technologies and impacts. In other words, a comprehensive team in terms of knowledge will necessarily involve IT personnel and OT personnel.
- A multidisciplinary team is required, involving especially OT personnel because they know not only the operational technology and its impact but also the business implications of interfering in OT processes.
- Incident response teams that are made up of IT personnel represent a great weakness, particularly when making decisions. As a result, the methodology cannot show its full potential.

5.2. Limitations and Challenges

There are several requirements and relevant aspects to be considered in the imple-mentation of the methodology. Firstly, as we have mentioned previously, maturity has a significant influence on the potential of the methodology but there is another essential pillar: the data. The proposed methodology performs numerical calculations based on data from multiple sources, such as vulnerability databases, IoC databases, configuration database management, risk analysis data, business continuity values, etc. This reflects the need for a high level of automation and integration to accelerate the process.

Secondly, the computation of the best strategy in complex environments requires scientific computation equipment to accelerate the process. When an incident has taken place, the response time and RTO is a very valuable aspect and therefore this kind of technology requirement must be considered. Finally, in the cybersecurity poli-cies, it is a "must" to assure the involvement of representatives from management, technical, legal, and communications disciplines. The partial absence of these repre-sentatives has a very significant influence on the quality of the results.

Furthermore, there are a set of challenges associated with resources and human factors when implementing this methodology due to its diversity and complexity. The implementation of the methodology will require the design and building of different APIs for facilitating the integration with existing services as shown in Fig. 6. From the SBOM (Software Bill of Materials) database, a continuous, reliable, and permanent connection to public databases (NVD, MITRE) is required to collect software com-ponent information, vulnerabilities, and weaknesses. At this point when collecting vulnerabilities, it is necessary to perform a triage process to verify that a software component is present because the presence of CPE information is not a warranty. Human interaction is therefore required in this process. The total or partial absence of these services involves an additional significant effort for automating the proposed methodology because it will require more time, which can be discouraging. However, the recent approval of the EU Cyber Resilience Act establishes the need to have a vulnerability management program in place. As a result, the organizations will be required to have some of the previously cited services, with special emphasis on in-dustrial and medical products.

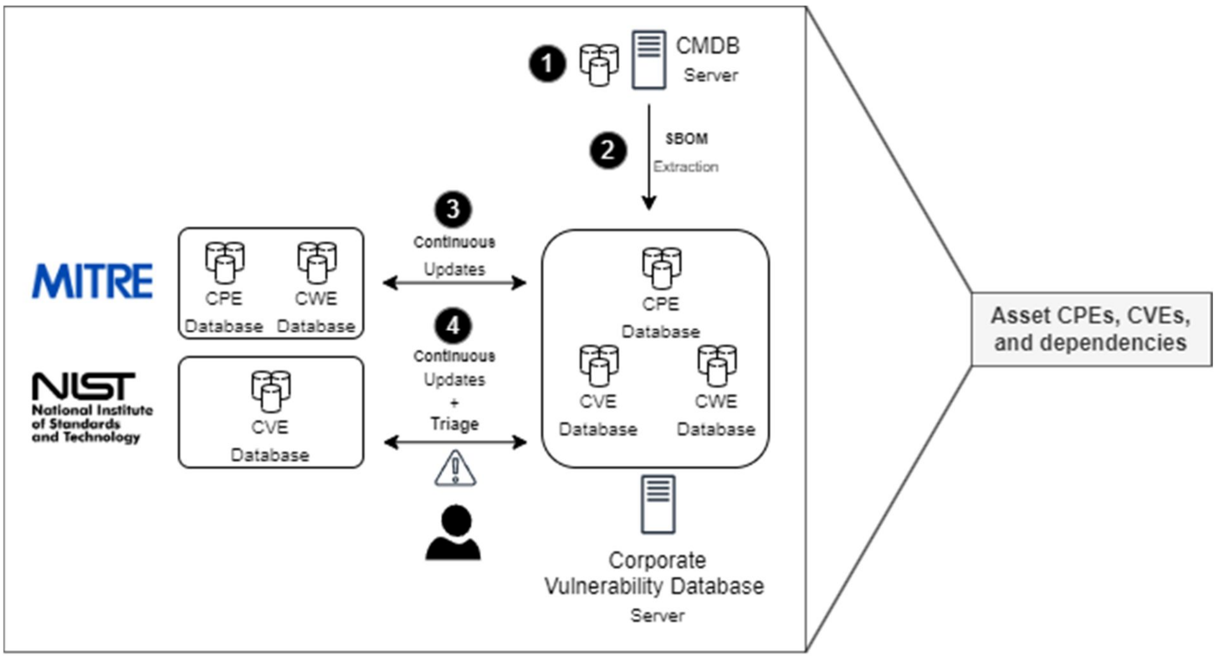


Figure 4: Integration of services.

The implementation of this methodology will also involve a recurrent training pro-cess for educating the users. Every effort invested in making processes and tools easi-er will facilitate the decisions and subsequent actions in key moments minimizing the resistance. It is necessary to pre-compute key values such as risk, RTO, and RPO to provide accurate information for decision-makers. The use of business process cata-loguing and classification/contextualization [33] jointly with machine learning can help to predict intentionality which becomes a key aspect for decision-makers. When receiving vulnerability information related to assets, it is necessary to perform the CVE triage process. If the personnel involved in this task is not committed to it, the mistakes produced in this part will be propagated to the rest of the process and will impact the decision process even when the personnel assigned to other tasks is fully committed.

6. Conclusion

Cybersecurity in complex ecosystems requires an advanced and proactive approach to managing mitigation measures. As a part of the Horizon 2020 CyberSEAS project, we hence introduced a systematic MCDM methodology to justify the costs and re-sources needed to implement such measures efficiently. The approach puts special emphasis on group decision-making because the implementation of mitigation measures is a collective effort that requires contribution on different organizational levels as well as from different stakeholders managing dependent IT/OT assets. We initially developed the methodology for the domain of critical energy infrastructures and power grids, but it is general enough to be relevant and suitable for application in a variety of complex ecosystems.

The group decision-making process is facilitated by the DSS. It integrates several external sources and established frameworks to provide decision-making infor-mation. This improves the efficiency of decision analysis and reduces the cognitive load of decision-makers. Furthermore, this methodology coherently and consistently merges technical efficiency and business resilience. This means that it is not only able to associate the computation of the countermeasure mitigation efficiency with an incident but also quantify the degree of integrating the information at the organiza-tional level with the technical efficiency information. With this strategy, it is hence possible to provide the optimal mitigation strategy information.

Within the scope of future research work, we will further enhance these aspects by developing appropriate machine learning models to automatically derive initial scores of impact assessment and mitigation selection criteria based on historical data. Deci-sion-makers will then be able to enhance these objective scores according to their subjective preferences and knowledge. The decision-making process can also be effi-ciently supplemented with the analysis of attack trees [14].

The evolution of the framework will provide more automation by utilizing a rein-forcement learning model to automatically estimate the risks, predict vulnerabilities, eliminate the need for CVE triage, and learn the best strategies. This will enhance the process of group decision-making by facilitating human decisions with correct infor-mation. A generative model based on MITRE ATT&CK can also be used to generate new kinds of attacks for training other models taking part in a proactive approach.

References

1.Alshawish, A., de Meer, H.: Risk mitigation in electric power systems: Where to start? Energy Informatics 2(34) (2019). https://doi.org/10.1186/s42162-019-0099-6.

2.Lee, A.: Analysis of Selected Electric Sector High Risk Failure Scenarios, Version 2.0. National Electric Sector Cybersecurity Organization Resource (NESCOR) & Electric Power Research Institute (EPRI) (2015).

3.Bregar, A.: Decision support on the basis of utility models with discordance-related preferential information: investigation of risk aversion properties. Journal of Decision Systems 27(sup1), 236–243 (2018). https://doi.org/10.1080/12460125.2018.1468170.

4.Bregar, A.: Application of a hybrid Delphi and aggregation–disaggregation procedure for group decision-making. EURO Journal on Decision Processes 7(1–2), 3–32 (2019). https://doi.org/10.1007/s40070-018-0094-3.

5.Bregar, A., Gyorkos, J., Jurič, M. B.: Robustness and visualization of decision models. Informatica 33(3), 385–395 (2009).

6.Bregar, A., Husseis, A., Flores, J. L.: A MCDM methodology for cyberattack mitigation. In: Drobne, S., Zadnik Stirn, L., Kljajić Borštnar, M., Povh, J., Žerovnik, J. (eds.) Proceedings of the 17th International Symposium on Operational Research in Slovenia SOR'23, pp. 417–420. Slovenian Society Informatika, Bled, Slovenia (2023).

7.CIS Controls. Center for Internet Security. https://www.cisecurity.org/controls, last accessed 2024/01/04.

8.Common Platform Enumeration Dictionary. National Institute of Standards and Technology. https://nvd.nist.gov/products/cpe, last accessed 2024/01/04.

9.Common Vulnerabilities and Exposures. The MITRE Corporation. https://www.cve.org/, last accessed 2024/01/04.

10.Common Vulnerability Scoring System. National Institute of Standards and Technology. https://nvd.nist.gov/vuln-metrics/cvss, last accessed 2024/01/04.

11.CyberSEAS – Cyber Securing Energy dAta Services. https://cyberseas.eu/, last accessed 2024/01/04.

12.DEXi: A Program for Multi-Attribute Decision-Making. Version 5.05. https://kt.ijs.si/MarkoBohanec/dexi.html, last accessed 2024/01/11.

13.Fritz, J.: Incident Response Methodology: The OODA Loop Explained. AT&T Cybersecurity (2019). https://cybersecurity.att.com/blogs/security-essentials/incident-response-methodology-the-ooda-loop, last accessed 2024/01/10.

14.Husseis, A., Flores, J. L., Bregar, A., Mazzeo, G., Coppolino, L.: Enhancing Cybersecurity Proactive Decision-Making through Attack Tree Analysis and MITRE Framework. In: 56th IEEE International Carnahan Conference on Security Technology ICCST 2023. Defence Institute of Advanced Technology, Pune, India (2023).

15.ISO 31000:2018: Risk management – Guidelines. International Organization for Standardization (2018). https://www.iso.org/standard/65694.html, last accessed 2024/01/10.

16.ISO/IEC 15408-1:2022: Information security, cybersecurity, and privacy protection, Evaluation criteria for IT security – Part 1. International Organization for Standardization (2022). https://www.iso.org/standard/72891.html, last accessed 2024/01/10.

17.Keeney, R. L., Raiffa, H.: Decisions with Multiple Objectives: Preferences and Value Tradeoffs. Cambridge University Press, New York (1993).

18.Mancuso, A., Zebrowski, P., Vieira, A. C.: Risk-based selection of mitigation strategies for cybersecurity of electric power systems. In: 28th Annual Conference of the Society for Risk Analysis Europe. SRA Europe, Potsdam, Germany (2019).

19.MITRE ATT&CK. The MITRE Corporation. https://attack.mitre.org/, last accessed 2024/01/04.

20.National Vulnerability Database. National Institute of Standards and Technology. https://nvd.nist.gov/, last accessed 2024/01/10.

21.Panetta, K.: How to Mange Cybersecurity Threats, Not Episodes. Gartner (2023). https://www.gartner.com/en/articles/how-to-manage-cybersecurity-threats-not-episodes, last accessed 2024/01/10.

22.Poolsappasit, N., Dewri, R., Ray, I.: Dynamic security risk management using Bayesian attack graphs. IEEE Transactions on Dependable and Secure Computing 9(1), 61–74 (2012). 10.1109/TDSC.2011.34.

23.Risk matrix. https://en.wikipedia.org/wiki/Risk_matrix, last accessed 2024/01/10.

24.Ross, R. S.: Managing Information Security Risk: Organization, Mission, and Information System View, Special Publication (NIST SP) – 800-39. National Institute of Standards and Technology (2011). https://www.nist.gov/publications/managing-information-security-risk-organization-mission-and-information-system-view, last accessed 2024/01/10.

25.Shameli-Sendi, A., Louafi, H., He, W., Cheriet, M.: Dynamic optimal countermeasure selection for intrusion response system. IEEE Transactions on Dependable and Secure Computing 15(5), 755–770 (2018). 10.1109/TDSC.2016.2615622.

26.Strategies to Mitigate Cyber Security Incidents. Australian Government, Australian Signals Directorate (2017). https://www.cyber.gov.au, last accessed 2024/01/04.

27.Turban, E., Aronson, J. E., Liang, T.-P.: Decision Support Systems and Intelligent Systems. 7th edn. Prentice-Hall, Upper Saddle River (2004).

28.What is Cybersecurity? Gartner. https://www.gartner.com/en/topics/cybersecurity, last accessed 2024/01/10.

29.Smith, R., Janicke, H., He, Y., Ferra, F., Albakri, A.: The agile incident response for industrial control systems (air4ics) framework. Computer Security (2021). 109:102398

30.Staves, A., Anderson, T., Balderstone, H., Green, B., Gouglidis, A., Hutchison, D.: A cyber incident response and recovery framework to support operators of ICS and critical national infrastructure. International Journal of Critical Infrastructure Protection (2022).

31.Cichonski, P., Millar, T., Grance, T., Scarfone, K.: Computer Security Incident Handling Guide. NIST SP 800-61 Rev. 2. Last accessed 2024/03/17.

32.Kraal, P. Incident Handler's Handbook. SANS. White Paper. Last accessed 2024/03/17.

33.Dimitriadis, A., Prassas, C., Flores, J. L., Kulvatunyou, B., Ivezic, N., Gritzalis, D. A., Mavridis, I. K.: Contextualized Filtering for Shared Cyber Threat Information. Sensors 21(14) (2021).

Social Loafing in Human-AI Collaboration

Edona Elshan[1], Triparna de Vreede[2], Dominik Siemon[3*], GJ de Vreede[2*], Phillipp Ebel[4]

[1] Free University Amsterdam, the Netherlands
[2] University of South Florida, USA
[3] LUT University, Finland
[4] University of St. Gallen, Switzerland
[*] gdevreede@usf.edu

Abstract

This study examined the dynamics of social loafing in AI-integrated team environments. Our aim was to understand how the inclusion of AI affects individual motivation and group performance. We employed a mixed-methods approach, combining quantitative data from controlled experiments with qualitative insights from interviews and surveys. Key findings indicate that AI presence can both mitigate and exacerbate social loafing tendencies, depending on factors like task complexity and AI capabilities. These results have important implications for designing AI systems and managing team dynamics, suggesting that tailored AI integration can optimize team efficiency and engagement.

Keywords: Artificial Intelligence; Human-AI Collaboration; Social Loafing

1. Introduction

Integrating Artificial Intelligence (AI) into organizations has significantly improved operational efficiencies and influenced workforce behavior. With the continual improvement in computational capacities, AI systems, such as conversational agents (CA) (Diederich et al., 2022), have transitioned from reactive problem-solving tools to proactive collaborators (Hevner & Storey, 2022). For instance, the role of CAs has moved from responding to questions (Vu et al., 2021) to actively participating in team creation activities (Dennis et al., 2023). Integrating these systems as active constituents in hybrid teams allows them to mitigate recognized teamwork limitations like inadequate probability handling, in-group favoritism, or confirmation bias (Peng et al., 2022). However, utilizing AI systems as teammates and allowing them to share the cognitive burden usually reserved for humans comes with its challenges.

On the one hand, it risks exacerbating one of the known challenges of group work, which is social loafing, where individuals exert less effort in group tasks compared to when working alone (Harkins, 1987). The ease and ostensibly coherent responses that the AI systems generate make it easier for the users to wholly depend on the AI for cognitive work. However, this perception of an AI taking over inherently human tasks also ignites uncertainty and mistrust toward the inclusion of an AI Teammate (Dennis et al., 2023). This dual impact of decreased effort and increased anxiety towards the AI systems may harm the organizations seeking to introduce AI teammates. Research shows that social loafing and anxiety towards human teammates have negative impacts on group work (Alnuaimi et al., 2010), but little research has examined the dynamics of social loafing with AI teammates. This inquiry is critical as AI becomes integrated into group interactions and dynamics. Examining the behavioral patterns in human-AI teams could reveal some unique differences with solely human-centric teams. This allows organizations and AI designers to develop more effective AI systems designed to counteract social loafing tendencies and foster a culture of accountability and engagement. Therefore, understanding the dynamics of human-AI groups is an essential first step to designing better AI teammates. Social loafing theory argues that individuals tend to disperse their efforts in group settings. This dispersion is often attributed to the diffusion of responsibility, where the accountability for the outcome is shared among all group members; thus, individuals might feel less personally responsible for exerting maximum effort. This leads to a motivational decline in group settings due to various factors, such as the perceived lack of contribution recognition, diminished concern about evaluation, or the assumption that others will compensate for one's reduced effort.

Thus, the central aspect of designing human-AI team dynamics is to understand the conditions in which human-AI collaboration is most optimal. The human-AI systems may then be designed such that there is accountability and engagement without evaluation apprehension or effort diffusion. The design could further incorporate individual differences such that people with varying characteristics could derive similar results from their interactions with the AI teammates. Against this backdrop, we theorize that the human-AI interaction is most impacted by the tendency of groups to engage in free-riding behaviors as well as the attitude of the team members towards the AI as well as each other. Our research questions are:

RQ1: Does the presence of an AI teammate result in social loafing behavior among the human counterparts?

RQ2: How does the attitude towards AI influence the effort exerted by humans in human-AI teams?

To answer each of these questions out, we conducted an experiment. For this, we created a human-AI collaborative environment where the subjects believed that they were interacting with an AI teammate and three human teammates. The results of these studies show that the degree of AI presence is a critical criterion for social loafing such that team members tend to engage more in social loafing behaviors if they perceive the AI presence. In addition, the presence of AI also resulted in decreased perceptions of knowledge self-efficacy among the team members. However, the sole presence of AI did not impact participants' attitudes toward AI. The theoretical model provides a foundation for future studies by showing that increasing the social presence of the AI by increasing the human likeness does not always have the intended effects on the team members.

2. Overview of Studies

Our research comprised three empirical studies exploring AI's impact on team dynamics and individual motivations; within this extended abstract we will report on our third study. The first online study focused on how AI affects individual contributions, interdependence, and psychological safety (H1-H4). The second, a controlled lab study, revisited these hypotheses and examined AI relatability (H5), integrating psychophysiological measurements. The final study, also in a lab, used neuroimaging techniques to further validate these hypotheses, adding depth to our understanding of the human-AI interaction. Each study progressively built on the previous, enriching our insights into the complex human-AI collaborative environment.

2.1. Hypothesis Development

Within our research, we explore the nuanced dynamics of social loafing in the context of human-AI collaboration, Despite its data processing strength, we propose that AI may inadvertently lead to reduced human effort in team tasks. We present five hypotheses: H1 examines AI's correlation with increased social loafing; H2 investigates if AI anxiety intensifies this effect; H3 considers whether a decrease in knowledge self-efficacy due to AI presence amplifies loafing behaviors; H4 explores the moderating role of team-oriented commitment; and H5 assesses how psychological safety might counteract social loafing in AI-integrated teams.

2.2. General Experimental Procedure

Each investigation operated on a bespoke digital interface, christened "Collaboration Space," tailored to our research objectives. Recent deliberations by Fink (2022) underscore the efficacy of digital empirical studies for human behavioral insights, lauding their pristine internal validity. The Collaboration Space interface paired participants with proxies, enacting either human or AI collaborators. This platform visualized a dispersed team quintet — including the participant — engaged in real-time idea-sharing sessions. All "collaborators," save for the actual participant, followed pre-set scripts, simulating interactive dialogues. These scripts drew parallels with generative AIs such as ChatGPT in terms of text production. To maintain neutrality, each simulated collaborator bore an androgynous moniker (e.g., Taylor, Kim, Alex, Sam).

2.3. EEG StudyMeasures

The surveys conducted both before and after the experiment encompassed 25 items. This selection drew from various research tools, such as the Social Presence Scale, a five-item adaptation inspired by Qiu and Benbasat (2009). Additionally, a revised version of the four-item Willingness to Depend on Others Scale, initially developed by McKnight et al.(1980), was incorporated. Van den Heuvel et al. (1995) provided the foundation for our seven-item Team-Oriented Commitment Scale. Furthermore, a version of Lin's (2006) five-item Motivation to Contribute Scale was adopted, along with questions to understand the demographic distribution of participants. Every metric in our survey was measured on a 5-point Likert-style scale. An exhaustive version of the questionnaire is available in the accompanying supplementary material.

Regarding physiological measurements, we incorporated electroencephalography (EEG) to capture neural activities and oscillations. EEG provides a non-invasive method to measure the brain's electrical activity, offering insights into the cognitive states, focus, and emotional arousal of participants. The decision to implement. First, the doing less (quantity) component of social loafing was addressed by evaluating each participant's length of contribution during the brainstorming session. This can be assessed by the number of ideas presented or the count of characters or words expressed (Alnuaimi et al., 2010; Chidambaram & Tung, 2005). In this study, the number of words written during the brainstorming process was used to measure the quantity component of social loafing as it reflects the effort exerted in generating each idea. Second, the doing poorly (quality) component of social loafing was captured through self-assessment questions. Numerous questionnaires exist that are tailored to measure social loafing, either about another teammate or oneself (George, 1992). Specifically for this research, the Social Loafing Tendency Questionnaire (SLTQ) formulated by Ying et al. (2014) was utilized. The SLTQ was chosen to evaluate the quality dimension of social loafing due to its proven capability to reliably predict individual performance in group task conditions (Ying et al., 2014). Additionally, to avoid redundancy, two out of the seven original items from the questionnaire were excluded as they closely resemble other items. Responses to each item were given on a scale ranging from 1 (strongly disagree) to 5 (strongly agree). Third, the distraction or lack-of-attention component of social loafing was assessed by evaluating participants' engagement levels during the simulation, as low task engagement was found to correlate strongly with distraction (Halin et al., 2014). The EPOC+ EEG neuroheadset, a commercial device, was used to capture real-time engagement metrics. This device continuously records mental performance stemming directly from the user's cognitive activity. Engagement, as stated by the manufacturer's website, represents a blend of attention and concentration. This is associated with heightened physiological arousal and beta wave activity, as well as diminished alpha wave presence. The EmotivPRO software processes this data, outputting a consistent and standardized engagement score on a scale from 0-1, with 0.5 representing the average engagement level for each participant. Since participants wore the EEG device before and after the simulation, engagement metrics during brainstorming were benchmarked against each individual's baseline state. Decreased values indicated diminished engagement, which implied greater distraction. The quantitative (word count), qualitative (questionnaire), and EEG engagement metrics were normalized to align with a 0-1 scale. Where necessary, the scales were inverted to ensure all three metrics aligned in the same direction. The combined weighted average of these three metrics culminated in a comprehensive social loafing score. Scores approaching 1 indicate elevated levels of social loafing.

Participants

Of the initial 79 participants, 74 completed the experiment. Technical issues, such as difficulties with the Collaboration Space platform or disruptions in the EEG measurements, prevented five participants from concluding the study. The age range of participants was between 18 and 35 years, with an average age of 23.65 years (M = 23.65, SD = 3.398). The gender distribution was 58.1% male (43 participants) and 41.9% female (31 participants). Regarding educational background, 43.2% (32 individuals) had a high school degree at the time of the experiment, 35.2% (26 participants) held a bachelor's degree, 20.3% (15 participants) a master's degree, and 1.3% (one person) had achieved a doctoral degree.

3. Results

The Tucker-Lewis Index (TLI) was greater than or equal to .95, TLI = 1.95, indicating that the model is a good fit for the data (Hooper et al., 2008). The CFI, despite being less than .90, CFI = 0.00, was accepted in light of other favorable fit indices. The RMSEA index was less than .08, RMSEA = 0.06, 90% CI = [0.00, 0.24], showcasing a good model fit (Hooper et al., 2008). The SRMR, landing between .05 and .08, SRMR = 0.06, further strengthens the argument that the model adequately fits the data (Hooper et al., 2008). Furthermore, a Chi-square goodness of fit test suggested that the analysis model fits the data comfortably. Despite the usual sensitivity of the Chi-square test to sample size, the test did not identify a significant misfit, $\chi2(2) = 2.51$, $p = 0.285$.

In the recent path analysis, certain pivotal relationships came to the forefront. Concerning the direct influence, Hypothesis 1 suggested that introducing an AI team member would lead to an increase in social loafing among human participants. Nonetheless, the findings indicate no significant difference in the levels of social loafing between subjects in the control group and those in the experimental group. This observation persisted despite addressing known limitations in prior studies on social loafing in hybrid teams, such as increasing the team size beyond two individuals (Onnasch & Panayotidis, 2020). Moreover, the study specifically addressed the shortcomings regarding the experimental procedure from Siemon & Wank's pre-study (2021). This was done by utilizing rule-based scripts to simulate the AI team member to reduce participants' perceived social presence of the AI and by incorporating a combination of quantitative and neurophysiological measures for a more robust assessment of social loafing rather than solely relying on self-reported data. The findings suggest that the levels of effort exerted by individuals when working with AI-based team members are comparable to those when collaborating solely with humans. This contradicts earlier studies on technology-supported team collaboration, introduced in the theoretical background of this study, which implied that perceiving other individuals as non-human entities would make it easier to target them with antisocial behaviors such as social loafing (Alnuaimi et al., 2010). Moreover, when considering the isolated results of the three different methods used to measure social loafing, participants collaborating with the AI team member wrote fewer words (-21 words) and self-reported slightly higher levels of social loafing (2.67) compared to the control group (2.59). However, the engagement levels, as measured by the EEG device, showed no difference between the groups. The role of presence was noteworthy as it revealed a negative association with knowledge self-efficacy. This translates to the observation that an increase of one unit in presence brings about a reduction of 0.20 units in knowledge self-efficacy. In simpler terms, individuals who sensed a heightened social presence were likely to experience a dip in their confidence in personal competencies within the given setting. The statistical relevance of this correlation is affirmed by a z-value of -2.42 and a p-value of 0.016.

Moving on to knowledge self-efficacy, it was discerned that this variable played a role in mitigating social loafing. When an individual's knowledge self-efficacy surges by a unit, there's an ensuing decline of 0.29 units in social loafing. This data point suggests a pattern: individuals who harbor a robust self-efficacy appear to shy away from tendencies of social loafing. The solidity of this association is highlighted by its z-value, -3.36, and a p-value that is less than 0.001. On the flip side, some connections failed to hold water statistically. Contrary to expectations, the study's results indicated no significant moderation by AI anxiety, suggesting that while AI anxiety negatively predicts AI adoption, it may not play a significant role in assessing an AI-based team member's competence within collaborative task with a z-value of 0.23 and a p-value of 0.819. A similar scenario played out with AI anxiety and social loafing, where a negative relationship was observed but lacked sufficient statistical weight, evidenced by a z-value of -0.64 and a p-value of 0.525. To encapsulate, Study 3 sheds light on the negative relationships between presence and knowledge self-efficacy and between knowledge self-efficacy and social loafing. However, connections involving AI anxiety, both with presence and social loafing, remained outside the realm of statistical significance.

4. Discussion

Understanding human-AI team dynamics is vital for effective system design, especially given AI's rapid workplace integration. This study investigates social loafing within such teams, focusing on AI presence and knowledge self-efficacy. Findings consistently show a negative relationship between AI presence and knowledge self-efficacy and its role in mitigating social loafing. However, AI anxiety's impact remains unclear, indicating a need for further research. The results suggest revisiting social loafing theories to incorporate AI dynamics. Knowledge of self-efficacy's negative correlation with social presence and its influence on team behavior demands more focused theoretical and empirical exploration. The inconclusive nature of AI anxiety calls for better theoretical understanding and measurement. For practice, these insights guide AI integration in team settings. Managing AI's social presence is crucial to avoid undermining human self-efficacy. Organizations should leverage AI as a complementary tool, focusing on enhancing human potential. Emphasizing knowledge, self-efficacy, and psychological safety is key to successful Human-AI collaboration. AI design should aim to augment, not replace, human capabilities, ensuring increased human participation and responsibility sharing.

5. Conclusion

The integration of AI systems as collaborators represents an impactful shift in group dynamics with profound implications. This research aimed to elucidate the influence of AI presence on critical aspects of teamwork like social loafing and individual motivations. The studies reveal a complex interplay between perceptions of AI presence, knowledge self-efficacy, and social loafing tendencies. The consistent negative relationship between presence and self-efficacy provides impetus for further examination of AI's impacts on human cognition and behavior. While this research expands the nascent understanding of hybrid human-AI collaboration, ample opportunities exist for additional theoretical and empirical endeavors. As AI transitions from automation to autonomy, developing insights into effective sociotechnical collaboration will only grow in urgency and importance. Both researchers and practitioners have pivotal roles to play in steering the future trajectory of human-AI coordination. This undertaking requires sustained multi-disciplinary investigations, designing human-centric systems while proactively addressing emerging sociotechnical challenges.

References

Alnuaimi, O. A., Robert, L. P., & Maruping, L. M. (2010). Team Size, Dispersion, and Social Loafing in Technology-Supported Teams: A Perspective on the Theory of Moral Disengagement. Journal of Management Information Systems, 27(1), 203–230. https://doi.org/10.2753/MIS0742-1222270109

Chidambaram, L., & Tung, L. L. (2005). Is Out of Sight, Out of Mind? An Empirical Study of Social Loafing in Technology-Supported Groups. Information Systems Research, 16(2), 149–168. https://doi.org/10.1287/isre.1050.0051

Dennis, A. R., Lakhiwal, A., & Sachdeva, A. (2023). AI Agents as Team Members: Effects on Satisfaction, Conflict, Trustworthiness, and Willingness to Work With. Journal of Management Information Systems, 40(2), 307–337. https://doi.org/10.1080/07421222.2023.2196773

Diederich, S., Brendel, A. B., Morana, S., & Kolbe, L. (2022). On the Design of and Interaction with Conversational Agents: An Organizing and Assessing Review of Human-Computer Interaction Research.

Fink, L. (2022). Why and how online experiments can benefit information systems research. Journal of the Association for Information Systems, 23(6), 1333–1346.

George, J. (1992). Extrinsic and intrinsic origins of perceived social loafing in organizations. Academy of Management Journal, 35(1), 191–202.

Halin, N., Marsh, J. E., Hellman, A., Hellström, I., & Sörqvist, P. (2014). A shield against distraction. Journal of Applied Research in Memory and Cognition, 3(1), 31–36.

Harkins, S. G. (1987). Social loafing and social facilitation. Journal of Experimental Social Psychology, 23(1), 1–18.

Hevner, A. R., & Storey, V. C. (2022). Research Challenges for the Design of Human-Artificial Intelligence Systems (HAIS). ACM Transactions on Management Information Systems.

Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural Equation Modelling: Guidelines for Determining Model Fit. Electronic Journal of Business Research Methods, 6(1), 53–60.

Lin, H.-F. (2006). Understanding Behavioral Intention to Participate in Virtual Communities. CyberPsychology & Behavior, 9(5), 540–547. https://doi.org/10.1089/cpb.2006.9.540

McKnight, R. D. (1980). Validation Results For The Sdx Cell Homogenization Code For Pin Geometry. Nuclear Science and Engineering, 75(1), 111–125. Scopus. https://doi.org/10.13182/NSE80-A20322

Onnasch, L., & Panayotidis, T. (2020). Social Loafing with Robots – An Empirical Investigation. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 64(1), 97–101. https://doi.org/10.1177/1071181320641026

Peng, A., Nushi, B., Kiciman, E., Inkpen, K., & Kamar, E. (2022). Investigations of performance and bias in human-AI teamwork in hiring. Proceedings of the AAAI Conference on Artificial Intelligence, 36(11), 12089–12097. https://ojs.aaai.org/index.php/AAAI/article/view/21468

Qiu, L., & Benbasat, I. (2009). Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems. Journal of Management Information Systems, 25(4), 145–182.

Siemon, D., & Wank, F. (2021). Collaboration With AI-Based Teammates–Evaluation of the Social Loafing Effect.

Van den Heuvel, H., Ellemers, N., & Seghers, B. (1995). Translated into English-(The importance of career-oriented and team oriented commitment among male and female employees: A study among PhD students). Sociale Psychologie En haar Toepassingen, 9, 168–186.

Vu, T. L., Tun, K. Z., Eng-Siong, C., & Banchs, R. E. (2021). Online FAQ Chatbot for Customer Support. In E. Marchi, S. M. Siniscalchi, S. Cumani, V. M. Salerno, & H. Li (Eds.), Increasing Naturalness and Flexibility in Spoken Dialogue Interaction (Vol. 714, pp. 251–259). Springer Singapore. https://doi.org/10.1007/978-981-15-9323-9_21

Ying, X., Li, H., Jiang, S., Peng, F., & Lin, Z. (2014). Group laziness: The effect of social loafing on group performance. Social Behavior and Personality: An International Journal, 42(3), 465–471.

joint conference group and negotiation DECISION support systems technology Porto 2024

24th International Conference on Group Decision and Negotiation &
10th International Conference on Decision Support System Technology

Human-Centric Decision and Negotiation Support for Societal Transitions

Design Considerations for a Crisis Communication Support System: Some Supporting Evidence from the Navy's Red Hill Fuel Leak at Pearl Harbor

Andrea Rivera[1], Tung Bui[1]

[1] University of Hawaii, Honolulu, HI 96822
[*] alrivera@hawaii.edu, tungb@hawaii.edu

Abstract

In a society increasingly influenced by social media, any crisis that captures public interest is susceptible to a rapid and widespread proliferation of posts, comments, and discussions throughout the course of the event. Meanwhile, authorities involved in such crises often struggle to effectively manage and contain the explosive flow of emotionally-charged information. The stark contrast between cautious and tardive institutional responses, and the continuous surge of disorderly and uncontrollable social media commentary, only exacerbates tension between the opposing sides. Analyzing the exchange of information and commentary surrounding a recent crisis—an extensive jet fuel leak at Pearl Harbor—we propose that authorities and their public relations officials examine the behavioral responses of social media users to effectively help the general public cope with the crisis by providing implementable institutional support. We propose the development of a Crisis Communication Support System (CCSS), aimed at assisting organizations in fostering a climate of trust and collaboration within the communities they serve.

Keywords: Crisis management and communications; crisis convergence behaviors; responsible information systems; crisis communication support and social media; generative AI for crisis management

1. Introduction

The sudden onset of a crisis situation often triggers varying emotional and behavioral responses among those directly or indirectly impacted by the event (Covello, 2011; Folkman et al., 1997; Kemp et al., 2021). Research has shown that regardless of the type of crisis (natural or man-made), the devastation and uncertainty that often accompany such an event lead to feelings of helplessness, anxiety, anger, and sadness among members of the community (Covello, 2011; Jin et al., 2007). Emotional reactions tend to drive behavioral intentions (Frijda, 1986), which are then exacerbated through online platforms, such as social networking sites (Eriksson, 2018). The ways in which people respond to a crisis have implications at both the individual and community levels, though governments and organizations often fail to provide the public with relevant and meaningful information following a crisis (Quarantelli, 1988).

Extant literature falls short in offering crisis managers communication strategies that are driven by real-time needs of the communities they serve. Lack of guidance or policy and inadequate staffing was cited as a barrier to effectively deal with the information overload generated by social media (Plotnick and Hiltz, 2016). There has been a recent call for research that explores responsible information systems (IS) design, development, deployment, and use (Davison et al., 2023). The objective of such research is to create theoretical frameworks and IS governance models that promote societal wellbeing, trust, and accountability with public stakeholders (Mikalef et al., 2022). However, current literature in crisis communication is inadequate at providing a framework that uses information systems to identify and characterize the affected community's real time emotional and behavioral responses to a crisis in order to effectively generate strategies that promote the public's wellbeing, organizational accountability, and trust. The absence of such a system results in missed opportunities to provide assistance to those in need.

In this paper, we propose a high-level architecture of a Crisis Communication Support System (CCSS), a more responsible solution to crisis communication. Drawing from existing theory in crisis communication and coping mechanisms, we consolidate known reactions to a crisis into categories designed to help crisis managers create tailored messages that meet the needs of the communities they serve. Our model uses real-time input from social media to evaluate the public's current emotional and behavioral responses to the event, and then uses that feedback to generate messages that address the physiological and psychological needs of the public. As such, our model is designed to promote the wellbeing of the public by reinforcing qualities of trust, accountability, and informed actions.

Our paper begins with an examination of a recent crisis, namely the Navy's fuel leak at Red Hill, Pearl Harbor, Hawai'i, highlighting the deficiencies in crisis communication and emphasizing the necessity of a crisis support system driven by public involvement. Subsequently, we provide an overview of prominent theories in the crisis communication literature, pinpointing their limitations and identifying gaps that need to be addressed. We then introduce a novel taxonomy of crisis convergence behaviors, and present our conceptual framework for the development of a Crisis Communication Support System (CCSS), outlining its six modules. Finally, we illustrate how the proposed CCSS could have drastically improved the Navy's response to the aforementioned crisis through a proof-of-concept analysis using generative AI.

2. The Navy's Red Hill Jet Fuel Leak in Pearl Harbor, HI

On November 28, 2021, the U.S. Navy reported in a Facebook post that they were investigating reports of a chemical smell in the drinking water at several military homes on Joint Base Pearl Harbor Hickam (JBPHH), Hawai'i. According to the Environmental Protection Agency (EPA), about 93,000 US Navy water system users were impacted by contaminated water. Soon after, the residents witnessed a surge of Facebook posts expressing confusion, worry, and anger. A crisis was underway - fuel storage tanks three miles inland of Pearl Harbor were leaking petroleum into the Red Hill aquifer, which supplies drinking water to military personnel and their families living in military housing (Britzky, 2021, 2022). Plagued with headaches, vomiting, diarrhea, and skin irritation, hundreds of military families were forced out of their homes and into military-funded hotel rooms (Britzky, 2021). Military families worried that the Navy was indolent in its campaign to identify contaminated drinking water, resulting in deficient contamination reports and false claims of safety (Liebermann, 2022). Many were appalled that the first lines of communication were via social media and that the nature of those messages seemed evasive. Local residents and the media did not forget an earlier jet fuel release incident that occurred just seven months earlier.

The fuel leak crisis spanned nearly three months, uprooting families from their homes and schools. The Navy employed several different crisis communication strategies during the course of fuel leak. Initially, they denied that drinking water was contaminated, though many residents were posting images and videos of oil-sheened water coming from their faucets. Four days later, the Navy acknowledged that drinking water was unsafe and offered actions people could take to protect themselves. However, many thought the Navy lacked genuine concern for those affected by the crisis, taking shortcuts in response measures to preserve their public image. During town hall meetings in which the Navy reported the status of the crisis and answered questions, Naval leadership appeared agitated and inconvenienced. In response to an onslaught of negative Facebook commentary, the commanding officer of the military base posted an apology on Facebook (one week after the institution initially denied the crisis), in which he expressed deep remorse for the pain and suffering caused by his words and actions. For the next two months, the Navy provided information on Facebook regarding the status of the fuel leak, cleanup efforts, and resources available to those impacted by the crisis. Table 1 provides excerpts from JBPHH Facebook postings at the onset of the crisis, and some of the public's responses to those postings (note that DOH stands for Department of Health).

Table 1. JBPHH Facebook postings and community responses at the onset of the fuel leak crisis.

Date	JBPHH Post	Public's Response
Nov. 28, 2021	*The Navy is investigating reports of a chemical smell in drinking water at several homes… There is no immediate indication that the water is not safe."*	*This is an inadequate statement and extremely underwhelming response.* *No one believes you that the water is safe. FIX IT!!! Don't just cover it up!!*
Nov. 29, 2021	*DOH and Navy tested water – no measurable quantities of fuel found.*	*Trying to cover this up and lie about it is pathetic.* *These families need solutions and CLEAN WATER not your gaslighting!*
Dec. 2, 2021	*The Navy has detected petroleum products in the Red Hill well… Based on these findings, the Navy will work with DOH to revise public health guidance.*	*You do know that absolutely NONE of us trust you hacks right?* *JBPHH starting taking some responsibility you also told us to drink, bathe, launder, and feed this to our families and children!*
Dec. 5, 2021	*We mistakenly felt the initial tests (that indicated no detection of contamination ppm) meant we may drink the water… I apologize with my whole heart that we trusted those initial tests.*	*We stand behind You Sir and confident we will overcome this issue.* *My faith in your leadership that day and EVERY day since has been LOST! You have failed us on all levels!*

This significant, life-threatening incident is an example of the repercussions stemming from inadequately developed crisis communication strategies employed by authorities. The failure to promptly address the public's response to the crisis carelessly amplifies the emotional and physical distress experienced by those who are already burdened by the crisis event. It is unfortunate to witness such a mishap within a prominent and resourceful organization that prides itself on a culture of responsibility and accountability.

Large crisis response organizations (such as those for state and federal governments) have a dedicated emergency response command center with a multidisciplinary team of professionals tasked with managing events such as the fuel leak. These emergency response centers typically have a trained group of responders whose sole purpose is to monitor the media (i.e., television news channels and social media platforms). These groups have their own physical workspaces equipped with televisions, phones, and computers to monitor various media channels. The standard procedure for these groups is to search the media for erroneous posts (specifically, those posted by other government officials) that may damage the organization's reputation. However, these task forces are not trained or even expected to use their resources to identify the public's needs following the crisis, and then use that information to draft public relations messages that are tailored to meet the community's needs. In other words, large crisis response organizations have the resources to assess the public's needs but, rather, choose to identify reputational risks to the organization. We argue that the primary reason why the Navy's crisis communication strategies were so poorly developed is related to deficiencies in existing academic literature in crisis communication.

3. Relevant Crisis Communication Literature

When a crisis occurs, the general public affected by the crisis may encounter various information-related challenges. Some common concerns include:

- Lack of timely, accurate information about the crisis, including details about its cause, severity, and potential impacts.
- Information overload and fragmentation from multiple sources, such as traditional media, social media, and official statements, can overwhelm the public and create confusion.
- Trust and transparency: The public may harbor doubts about transparency, honesty, and the credibility

of the information provided. Emotional support and empathy: During a crisis, individuals may require emotional support and empathy. They may seek information that acknowledges their emotional needs, provides reassurance, and offers guidance on managing stress and coping mechanisms.

- Community relevance and contextualization: Individuals often seek information that is personally relevant to their specific circumstances during a crisis and their convergence behaviors. They may require information tailored to their location, risk level, available resources, and appropriate actions to take. Lack of personalized and contextualized information can impede decision-making and preparedness efforts.
- Inadequate communication channels: The availability and accessibility of communication channels can pose challenges. Limited or disrupted communication infrastructure, overwhelmed hotlines, or inaccessible websites can hinder the public's ability to receive crucial updates and guidance.

There are three primary areas in which existing literature is insufficient at helping crisis managers effectively communicate with the public. First, existing crisis communication literature promotes strategies that are primarily designed to protect the organization's image, rather than protect the public. These theories include Corporate Apologia (Ware & Linkugel, 1973), Image Repair Theory (Benoit, 1997), and Situational Crisis Communication Theory (Coombs & Holladay, 2002). Corporate Apologia focuses on how organizations respond to criticism following a crisis, focusing on defensive strategies to protect the organization's image (Ware & Linkugel, 1973; Hearit, 2001). The Image Repair Theory is designed to help crisis managers repair and restore an organization's public image following a crisis, noting that public perception of the organization is more important than reality (Benoit, 1997). Similar to Image Repair Theory, Situational Crisis Communication Theory (SCCT) offers crisis managers a framework to preserve the organization's reputational capital following a crisis, offering crisis communication strategies that range from accommodative to defensive depending on how much blame the public places on the organization (Coombs & Holladay, 2002). While these theories offer practical crisis communication strategies, all three of the strategies are designed to protect the organization's image or reputation following a crisis. That is, none of the strategies are specifically designed to meet the needs of the communities impacted by the crisis.

Second, existing literature treats crisis situations and the public's responses to a crisis as static, failing to offer a system that uses real-time and evolving public sentiments to design crisis response messages. While some technology-based crisis support systems (e.g., S-HELP) are designed to help decision-makers assess the physical circumstances of the crisis, there are no such systems to help crisis managers assess the public's emotional/behavioral responses in real-time. Therefore, existing theoretical frameworks only hypothetically anticipate the public's response to a crisis based on various constructs that are evaluated at the onset of the crisis. However, crisis situations are dynamic and, as such, so are the public's responses to a crisis.

Third, existing literature that does take a more community-centered approach fails to effectively offer crisis communication strategies that address the public's specific needs following a crisis. For example, though the Centers for Disease Control and Prevention (CDC) created seven best practices for crisis risk communication and public protection in 2011, Parmer et al. (2016) found that media messages largely fail to incorporate these practices into their message-drafting. Other community-centered research in crisis management includes the Integrated Crisis Mapping (ICM) Model and literature on crisis convergence behaviors. The Integrated Crisis Mapping (ICM) Model, examines the public's emotional responses following a crisis, specifically, anger, anxiety, sadness, and fright. ICM employs two continuums to map these emotions: 1) coping mechanisms, ranging from conative (action oriented) to cognitive (thought oriented) coping strategies, and 2) the level of organizational engagement in the crisis, spanning from low to high. With this framework, crisis managers can anticipate how the public will emotionally respond to and cope with a crisis based on the type of crisis event and the level of organizational involvement in the crisis (Jin et al., 2007). Research on crisis convergence behaviors reveals that individuals directly or indirectly affected by the crisis tend to converge toward one of a number of different behaviors, referred to as convergence behaviors (Fritz & Mathewson, 1957). Existing literature has identified thirteen different types of convergence behaviors identified by the public. While the ICM model and convergence behavior literature both focus on how the public responds to a crisis, they fall short in providing crisis communication strategies based on those known behaviors. Moreover, the sheer number of thirteen different behaviors can be overwhelming and impractical for crisis managers to effectively respond to after a crisis event. In Section 4.2. of this paper, we reconceptualize these behaviors and offer a new taxonomy of convergence behaviors by consolidating them into five categories based on the informational needs and highly associated behaviors of each group.

To address these three issues with existing literature, we offer the Crisis Communication Support System (CCSS) which draws data on the public's current sentiment regarding the crisis, filters and categorizes the data according to group-level informational needs, and recommends crisis response messages that are tailored to meet those needs while promoting trust and accountability. The remainder of this paper, thus, focuses on reconceptualizing the communication theories in the context of social media, and the analysis and design of the CCSS.

4. Requirements Analysis for the CCSS

The purpose of the CCSS is to deliver effective and efficient crisis communication recommendations that align with the public's current and future needs. In the upcoming section, we present a CCSS that effectively fulfills these requirements. The CCSS architecture exhibits a more responsible approach to crisis management by prioritizing community-driven information needs, while protecting the institution's reputation. The approach aims to foster qualities of trust, accountability and community-engagement with the public. To achieve this goal, the system should meet a dual requirement: Firstly, the system should have an open architecture to address crisis commentary on social media in an inclusive manner; secondly, the system should differentiate various types of convergence behaviors among the general public to tailor specific messages to them.

4.1. Social Media Requirements

As a result of the previous discussion, the system should be functional in the following four areas:

- *Assessing public sentiment:* The system should be capable of evaluating the prevailing public sentiment regarding the crisis. It should be able to gather all information that is related to the crisis from all possible sources, and analyze the inputs to accurately gauge the public's emotional state, concerns, and reactions.
- *Identifying informational needs:* Based on the current state of public, the system should assess specific information gaps and knowledge needs of different groups of convergence behaviors.
- *Recommending crisis communication strategies:* The system should provide crisis spokespeople with tailored recommendations for communication messages that address the information needs of the public.
- *Determining the appropriate channels:* The system should determine the most appropriate channels through which information is disseminated to the public. It should consider selection factors such as accessibility, timeless, and targeted relevance to ensure effective deliveries.

In order for the system to meet these functional needs, the following general system requirements must be met:

- *Public-driven operation:* The system should generate outputs derived from public sentiments. It should prioritize the public's perspectives, ensuring that their true needs and concerns shape the system's recommendations.
- *Agile system:* The system should be flexible and adaptable, open-source whenever applicable, allowing crisis responders to run it periodically or on an as-needed basis.
- *Omnichannel architecture:* The CCSS involves integrating numerous systems. These systems may have different data formats, siloed, proprietary, or public. The overall architecture should be inclusive and transparent.
- *Strategic, responsible, and practical protocols:* The system should offer clear, responsible, fair and practical guidelines that crisis managers can employ to address the public in an accountable manner.

4.2. Convergence Behavior Requirements

Existing literature on convergence behaviors lacks discussions on crisis communication strategies based on behavioral responses and includes an excessive number of behaviors for response managers to handle effectively. However, even though these convergence behaviors may be mutually exclusive on an individual level, they are not mutually exclusive from a crisis communication perspective. To address these issues, we propose categorizing and consolidating the thirteen responses based on the informational needs and coping mechanisms of each group, thus, enabling crisis managers to tailor messages accordingly.

Sturges (1994) identifies two primary categories of crisis information: instructing information, which guides physical responses to the crisis, and adjusting information, which addresses the psychological needs of the group (Quarantelli, 1988). Duhachek (2005) describes coping as a set of cognitive and behavioral processes initiated in response to an emotionally arousing or stress-inducing situation. Coping efforts may be categorized as either problem-focused (taking action to change the circumstances of the stressful situation) or emotion-focused (reappraisal of the stressful situation) (Lazarus, 1991). Problem-focused coping includes two types of coping: 1) Action coping, which involves direct attempts to manage the source of stress, and 2) instrumental support-seeking, which involves engaging with others to help relieve the stress through action (Duhachek, 2005). Emotion-focused coping includes six types of coping, three of which are relevant to crisis response behaviors: 1) Positive thinking, which involves reconstruing the event so that it's less stressful; 2) rational thinking, which involves preventing one's emotions from driving behavior; and 3) emotional support-seeking, which involves engaging with others to improve one's emotional state (Duhachek, 2005).

Building on this research (including research in convergence behaviors and the ICM Model), we have developed five crisis response categories by grouping the convergence behaviors according to their informational needs and coping mechanisms:

Category 1: Knowledge Explorers

The first category of crisis response behaviors combines the following five convergers: the curious, the fans, the detectives, the spreaders, and the correctors. Members of these groups have a common need for factual information regarding the crisis. The curious are interested in the impact of and destruction caused by the event (Fritz & Mathewson, 1957); the fans look for information they can use to support crisis responders (Kendra & Wachtendorf, 2003); the detectives gather intelligence about the event so that they can watch over the situation (Subba & Bui, 2010); the spreaders look to share information regarding the crisis (Leonardi et al., 2021); and the correctors identify and clarify misinformation in online discussion boards (Vraga & Bode, 2017). Though these behaviors may manifest differently on an individual level, all of these convergers desire factual information about the crisis, including the circumstances leading up to the crisis, the current state of the crisis, and what is anticipated to happen next. Therefore, from a crisis manager's perspective, the strategy used to communicate to these five groups is relatively the same.

If a large percentage of the public falls under Category 1, crisis managers should employ subject-matter experts to provide detailed descriptions, pictures, and graphics of the crisis to assuage the public's need for factual information. Since Category 1 behaviors are not associated with strong emotional reactions to the crisis, we don't anticipate members of this category to engage in any coping mechanisms. Therefore, the most effective crisis communication strategy to address this category of response behaviors is to provide detailed, expert-level information regarding the crisis.

Category 2: Crisis Support Mobilizers

The second category of crisis response behaviors includes the returnees and the helpers. Members of this category have a common need for instructional information. The returnees are those who return to the crisis location to help victims, protect their properties, assess damage, or look for loved ones (Fritz & Mathewson, 1957). The helpers are those who look for ways to assist friends, relatives, and other victims in recovering from the crisis (Fritz & Mathewson, 1957). Both of these groups want to help, but may need guidance on ways to help that don't interfere with existing crisis response efforts. Crisis Support Mobilizers exhibit active involvement in returning to the crisis locations, providing assistance, and exercising an invaluable role in offering aid, protection, assessment, and search efforts, in addition, and independent of, official rescue responders.

If a large percentage of the public falls under Category 2, crisis managers should offer instructing information to encourage active coping. The nature of these convergence behaviors (the helpers and the returnees) implies that members of this category engage in action coping to manage the stress associated with the crisis. Members of this group will need specific guidelines on actions they can take; left to their own will, they may unintentionally engage in activities that harm themselves and others, or hinder relief efforts. Crisis managers should offer instructing information that encourages specific activities that members of this group can safely partake in to help themselves, help others, or help first responders.

Category 3: Anxiety Copers

The third category of crisis response behaviors consists of the anxious. The anxious are those who are worried about the crisis situation and the people impacted by the crisis (Fritz & Mathewson, 1957). In their theoretical development of the ICM model, Jin, Pang, and Cameron (2007) suggest that the anxious may feel overwhelmed by the crisis situation and look for ways to mitigate their anxiety through conative (action-oriented) coping, though they may not know exactly what actions to take (Jin et al., 2010; Lazarus, 1991). Under the uncertainty of how to resolve their anxiety, they may engage in instrumental support-seeking as they look toward others to help them identify actions to take. The Anxiety Copers may feel overwhelmed by the crisis and seek ways to alleviate their anxiety through action-oriented coping strategies. However, they may lack clarity on the specific actions to take. Consequently, they engage in instrumental support-seeking, turning to others for guidance and assistance in identifying appropriate courses of action to address their anxiety and contribute to the crisis resolution.

Crisis managers can support members of this category by offering instructing information (to advise people on actions they can take) along with some adjusting information (to help them psychologically deal with their anxiety). Crisis managers may also encourage safe ways for people to communicate with each other to further reduce their anxiety through social engagement.

Category 4: Emotional Reactors

The fourth category of crisis response behaviors consists of the frightened and the angry. Though these are clearly different constructs as experienced on an individual level, we group them together here because they harness similar informational needs and coping mechanisms. Smith and Ellsworth (1985) found that the two constructs differ primarily in qualities of certainty, control, and responsibility. Fear develops in response to the uncertainty of a threat (Jin et al., 2007) and is associated with a low level of control (Jin, 2010). Research has demonstrated that people may respond to a threat with avoidance (Duhachek, 2005) if they don't believe they can control the circumstances.

Anger, on the other hand, manifests when the public believes the organization holds a high level of attribution in the crisis (Coombs, 2007) or if the public feels the organization has caused or is failing to prevent harm to them (Jin et al., 2007). Anger is typically associated with a higher level of certainty and individual control (Smith & Ellsworth, 1985). Research in coping mechanisms demonstrates there is an action-tendency among those who are angry to move toward the source that's causing the anger (Smith & Ellsworth, 1985). The ICM model further suggests that anger involves conative coping, including active or passive forms of attack against the organization (Jin et al., 2007). As such, SCCT posits that when organizations hold a high level of responsibility in a crisis, they should respond with accommodative messages (including compensation and an apology) using adjusting information to reduce negative affect toward the organization (Coombs, 2007).

As fear and anger share similarities in terms of their informational needs and coping mechanisms, we propose crisis managers can support members of this category by offering instructing information to encourage positive activities that can alleviate the emotions experienced by individuals in this group To discourage avoidance or attack, crisis managers can promote positive thinking, rational thinking, and instrumental support to help people reframe their fear and anger and find others to support them in the process.

Category 5: Grief Processors

The fifth category of crisis response behaviors includes the mourners, which we relate to the emotional state of sadness proposed in the ICM model. The mourners are those who grieve the devastation of the crisis and mourn the loss of people or property (Kendra & Wachtendorf, 2003). Lazarus (1991) explains that sadness develops after the realization that the loss is irrevocable, resulting in a sense of helplessness. The ICM model places sadness under cognitive (emotion-oriented) coping as individuals try to understand and find meaning in the crisis event (Jin et al., 2007). We propose that crisis managers can best support this category by offering adjusting information and by encouraging emotional support coping mechanisms. Grief Processors engage in cognitive coping mechanisms to understand and find meaning in the crisis event. Crisis managers can help members of this category by offering empathy, psychological support, and by promoting emotional support through social engagement.

Two convergence behaviors, exploiters and manipulators, which both involve illegal activities, are intentionally excluded from these categories. Crisis managers should focus on addressing the needs of those in distress and leave the handling of illegal activities to local authorities as necessary. Table 2 summarizes the five crisis response categories, associated convergence behaviors, informational needs, and coping mechanisms for each category. By adopting this categorization approach, crisis managers can effectively address the diverse information needs of the public and tailor their communication strategies accordingly.

Table 2. A taxonomy of five crisis convergence response categories and their associated communication support and coping needs.

Response Category	Convergence Behaviors	Information Needs	Supportive Coping Mechanisms
Category 1 Knowledge Explorers	Curious, Fans, Detectives, Spreaders, & Correctors	Factual	None
Category 2 Crisis Support Mobilizers	Returnees & Helpers	Instructing	Action Coping
Category 3 Anxiety Copers	Anxious	Instructing & Adjusting	Instrumental Support
Category 4 Emotional Reactors	Frightened & Angry	Adjusting & Instructing	Positive Thinking Rational Thinking Instrumental Support
Category 5 Grief Processors	Mourners	Adjusting	Emotional Support

5. Architecture of the CCSS

Our proposed CCSS is modeled in Figure 1 and consists of six distinct modules, each of which is briefly described below. Due to concerns regarding the manuscript's length, we are not able to provide more technical details on each of the modules in this paper.

Social Media Commentary Scraper

The Social Media Commentary Scraper is a web scraping tool that extracts data from social media platforms, including social media sites such as Facebook, Twitter, Instagram, and news sites. During crisis events, social media platforms facilitate "many-to-many" peer communications, allowing people to exchange thoughts, crisis-related information, and guidance (Bui, 2019). They provide a rich source of data demonstrating the public's emotional and behavioral responses to a crisis. This type of tool is available for purchase through a variety of different platforms (such as Dripify or Octoparse), or it can be designed in-house by the organization. The overall objective of the Social Media Commentary Scraper is to pull real-time social media commentary so that it can be analyzed to detect emerging crisis convergence behaviors. The tool is agile and can scrape commentary along any prescribed periodicity or on an as-needed basis, offering crisis managers high levels of flexibility.

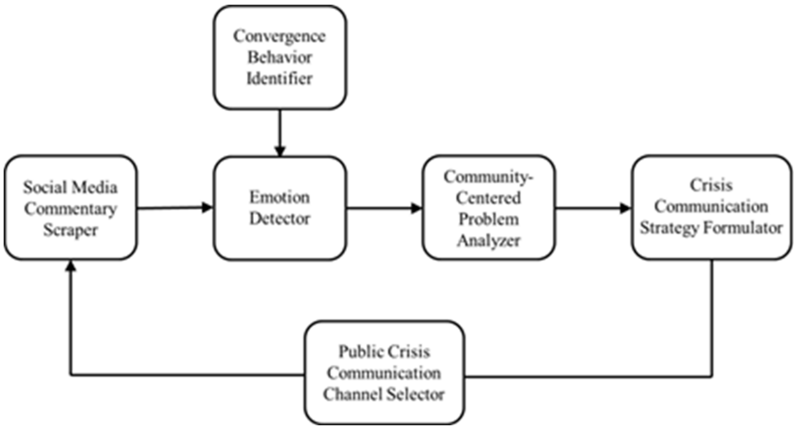


Figure 1. The Components of the Crisis Communication Support System

Social Media Commentary Scraper

The Social Media Commentary Scraper is a web scraping tool that extracts data from social media platforms, including social media sites such as Facebook, Twitter, Instagram, and news sites. During crisis events, social media platforms facilitate "many-to-many" peer communications, allowing people to exchange thoughts, crisis-related information, and guidance (Bui, 2019). They provide a rich source of data demonstrating the public's emotional and behavioral responses to a crisis. This type of tool is available for purchase through a variety of different platforms (such as Dripify or Octoparse), or it can be designed in-house by the organization. The overall objective of the Social Media Commentary Scraper is to pull real-time social media commentary so that it can be analyzed to detect emerging crisis convergence behaviors. The tool is agile and can scrape commentary along any prescribed periodicity or on an as-needed basis, offering crisis managers high levels of flexibility and control.

Convergence Behavior Identifier

The Convergence Behavior Identifier and Emotion Detector work together to analyze social media commentary, categorize the commentary according to the existing thirteen convergence behaviors, and quantify the results for use by the Community-Centered Problem Analyzer. In our CCSS model, the Convergence Behavior Identifier provides insight regarding how each convergence behavior manifests in social media commentary (based on those characteristics described in Table 2) so that the Emotion Detector can effectively identify those behaviors in the data.

Emotion Detector

Designed based on input from the Convergence Behavior Identifier, the Emotion Detector identifies convergence behaviors in the text and quantifies the proliferation of each behavior (i.e., how many of each of the convergence behaviors are detected in the social media text). Emotion detection (ED) software is a new technology developed as a byproduct of sentiment analysis tools. ED uses machine learning and natural language processing to analyze text (or, in other cases, facial expressions) to detect emerging human emotions. ED software is often offered as a cloud-based product that can be trained to identify specific emotions in unstructured text with a high degree of accuracy (Garcia-Garcia et al., 2017). After the Emotion Detector has analyzed the social media text, it then sends the results to the Community-Centered Problem Analyzer.

Community-Centered Problem Analyzer

This component uses the taxonomy developed in Table 3 to consolidate the convergence behaviors according to similarities in the informational needs and coping mechanisms of each group. It identifies the most prominent category (1-5) demonstrated in the social media text and aggregates the informational needs and coping mechanisms associated with that category. This component of the CCSS offers a concise and, therefore, more manageable set of public response behaviors crisis managers can use to design messages that are tailored to meet the needs of the communities they serve.

Crisis Communication Strategy Formulator

This module uses input from the Community-Centered Problem Analyzer to recommend communication styles and coping mechanisms that should be incorporated into a crisis response message, based on current public needs. The output of the Formulator includes: 1) A visual aid that uses color intensity to indicate the concentration of each response category; 2) the general informational needs and coping mechanisms associated with each category; and 3) a recommended press release script. The press release script is produced by generative AI (such as ChatGPT), using the aforementioned qualities (category type, informational needs, and coping mechanisms) associated with the most prevalent category. Figure 2 offers an overview of the deliverable product created by the Formulator, providing a clear and concise indication of the public's emotional and behavioral state, and the best way to address them given their informational needs.

Public Crisis Communication Channel Selector

The purpose of this module is to help crisis managers determine the best channel to disseminate information. Ideally, information would be disseminated across as many channels as possible, but time and cost constraints may limit the variety of channels an organization can use. This tool will gather information from the Social Media Commentary Scraper on which platforms (news and social media sites) seem to have the highest concentration of commentary regarding the crisis, and recommend that response managers promulgate information over those channels if resources are limited.

Leveraging its massive information resources, AI can be programmed to swiftly provide contextually relevant insights and offer recommendations on public relations (PR) with the identification and selection of communication channels and propose situation-dependent and channel-specific press releases. It should also be designed to deal with issues related to protecting privacy, and avoiding potential biases and prejudices that may be present in machine learning algorithms relying on open source data.

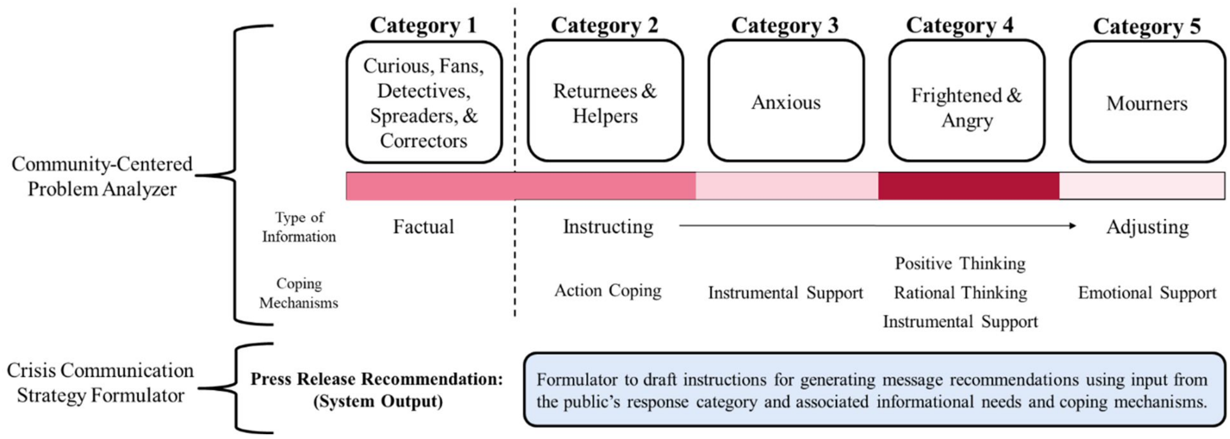


Figure 2. CCSS Functionalities Based on Target Convergence Behaviors' Information Needs

6. Proof-of-Concept: Fuel Leak Crisis Revisited

We conducted a feasibility test on our proposed CCSS to ensure its intended functionality. Of note, this is not intended to demonstrate a full empirical application of our CCSS; rather, it is meant to demonstrate what the output would look like in a real-world crisis situation. We applied the functional model (Figure 2) to the Fuel Leak crisis. Utilizing 210 Facebook posts collected over the first week after the incident, users were sorted into the five categories of response behaviors proposed in Table 3 (Scraper/Convergence Behavior Identifier/Emotion Detector/Crisis Communication Strategy Formulator). The percentage of each crisis response category demonstrated among the 210 posts is as follows, in descending order: Category 4 (40%), Category 1 (28%), Category 2 (20%), Category 3 (10%), and Category 5 (2%). Of note, the largest group of Facebook posts, comprising 40% of the total comments we analyzed, fell into Category 4, identified as "Anxiety Copers." This group exhibited fear and anger, which was detected by our Emotion Detector. The Communication Strategy Formulator next utilized generative AI technologies to recommend a press release message that demonstrated empathy (adjusting information), offered tasks people could do to overcome their distress (instructing information), and harnessed qualities of positive thinking, rational decision-making, and instrumental support. In Appendix A, we provide an example of this message to demonstrate the capability of the system and, in particular, the effectiveness of generative AI at producing eloquent crisis response messages that meet the public's needs. In contrast to the ineffective messages reported in Table 1, the recommended wording from the CCSS incorporates best practices from crisis communication theories discussed in Section 3. These practices aim to alleviate anxiety through action-oriented coping and by fostering institutional engagement with the public.

7. Summary

Social media has emerged as a powerful platform for the widespread dissemination of information during times of crises. In ideal circumstances, the practice of crowdsourcing information sharing has proven to be invaluable to the general public, enabling the timely posting of relevant content. Authorities face increased pressure to navigate a fact-confusing and emotionally-charged crisis situation, while facing competing interests to either protect the organization or protect the public. Institutions that bear responsibility and/or accountability during such crises must exercise extreme caution when formulating their communications. Our examination of the fuel leak incident at Pearl Harbor serves as a prime example of officials' failure to provide timely and accurate information to the general public. This failure can be attributed to poorly designed crisis communication strategies that focus on protecting the organization rather than addressing the public's needs, leading to widespread dissatisfaction.

This paper advocates for a responsible communication strategy that considers the convergence behaviors of the target audience and categorizes them based on their communication and information needs. Additionally, we introduce the design of a Crisis Communication Support System (CCSS) that aims to enhance interactions between authorities and their constituents. The CCSS incorporates empathetic, positive, accountable, and supportive messages that align with the emotions of specific social media users. We place significant emphasis on building a support system, as we advocate for the critical role of human involvement in Public Relations during times of crisis, where institutional trust is of paramount importance.

The feasibility analysis suggests that our proposed CCSS holds significant potential to enhance the effectiveness of communications and interactions between authorities and their constituents. However, to validate its functionality and effectiveness, an empirical test is still required. Of particular importance is verifying that the recommended crisis communication messages generated by the system do, in fact, genuinely meet the public's needs across the five categories. To initiate the testing process, we intend to gather reactions from residents at Pearl Harbor who have previously posted unsatisfactory messages and assess their responses to the messages recommended by our proposed framework. Should the CCSS prove to be a reliable system for designing crisis response messages, it has the power to revolutionize crisis communication practices. By shifting from hypothetical, static guesswork to generative, customized messaging intended to serve the public, it would ultimately benefit the organization and its stakeholders.

References

Benoit, W. L. (1997). Image repair discourse and crisis communication. Public Relations Review, 23(2), 177–186.

Bui, L. (2019). Social Media, Rumors, and Hurricane Warning Systems in Puerto Rico. Hawaii International Conference on Systems Sciences.

Coombs, T. (2007). Protecting Organization Reputations During a Crisis: The Development and Application of Situational Crisis Communication Theory. Corporate Reputation Review, 10(3), 163–176.

Coombs, T., & Holladay, S. (2002). Helping Crisis Managers Protect Reputational Assets. Management

Communication Quarterly, 16(2), 165.

Covello, V. T. (2011). Risk communication, radiation, and radiological emergencies: Strategies, tools, and techniques. Health Physics, 101(5), 511–530.

Davison, R. M., Majchrzak, A., Hardin, A., & Ravishankar, M. N. (2023). Special issue on responsible IS research for a better world. Information Systems Journal, 33(1), 1–7.

Duhachek, A. (2005). Coping: A multidimensional, hierarchical framework of responses to stressful consumption episodes. The Journal of Consumer Research, 32(1), 41–53.

Eriksson, Mats (2018), Lessons for crisis communication on social media: A Systematic review of what research tells the practice. International Journal of Strategic Communication. 12:5:526-551.

Fearn-Banks, K. (2002). Crisis Communication: A Casebook Approach. Lawrence Erlbaum.

Folkman, S., Moskowitz, J., Ozer, E., & Park, C. (1997). Positive meaningful events and coping in the context of HIV/AIDS. In Coping with Chronic Stress (pp. 239–314). Plenum Press.

Frijda, N. H. (1986). Studies in emotion and social interaction: the emotions. Cambridge University Press.

Fritz, C. E., & Mathewson, J. H. (1957). Convergence behavior in disasters: A problem in social control. National Academy of Sciences-National Research Council, No. 9.

Garcia-Garcia, J. M., Penichet, V. M. R., & Lozano, M. D. (2017). Emotion detection: A technology review. ACM International Conference Proceeding Series, Part F131194.

Jin, Y. (2010). Making sense sensible in crisis communication: How publics' crisis appraisals influence their negative emotions, coping strategy preferences, and crisis response acceptance. Communication Research, 37(4), 522–552.

Jin, Y., Pang, A., & Cameron, G. T. (2007). Integrated crisis mapping: Towards a public-based, emotion-driven conceptualization in crisis communication. Sphera Publica, 7, 81–96.

Jin, Y., Pang, A., & Cameron, G. T. (2010). The role of emotions in crisis response: Inaugural test of the integrated crisis mapping (ICM) model. Corporate Communications, 15(4), 428–452.

Kemp, E., Bui, M., & Porter, M. (2021). Preparing for a crisis: examining the influence of fear and anxiety on consumption and compliance. Journal of Consumer Marketing, 38(3), 282–292.

Kendra, J. M., & Wachtendorf, T. (2003). Reconsidering convergence and converger legitimacy in response to the World Trade Center disaster. In Terrorism and disaster: New threats, new ideas. Emerald Group Publishing Limited.

Lazarus, R. S. (1991). Emotion and adaptation. Oxford University Press.

Leonardi, S., Rizzo, G., & Morisio, M. (2021). Automated classification of fake news spreaders to break the misinformation chain. Information, 12(6), 248.

Mikalef, P., Conboy, K., Lundström, J. E., & Popovič, A. (2022). Thinking responsibly about responsible AI and 'the dark side' of AI. European Journal of Information Systems, 31(3), 257–268.

Newsom, D., & Carrell, B. J. (1986). Public relations writing: Form and style. Wadsworth.

Parmer, J., Baur, C., Eroglu, D., Lubell, K., Prue, C., Reynolds, B., & Weaver, J. (2016). Crisis and Emergency Risk Messaging in Mass Media News Stories: Is the Public Getting the Information They Need to Protect Their Health? Health Communication, 31(10), 1215–1222.

Plotnick, Linda and Hiltz, Starr Roxanne. (2016). Barriers to use of social media by emergency managers. Journal of Homeland Security and Emergency Management, 13(2):247-277.

Quarantelli, E. L. (1988). Disaster and crisis management: a summary of research. Journal of Management Studies, 25(4), 373–385.

Smith, C., & Ellsworth, P. (1985). Patterns of cognitive appraisal in emotions. Journal of Personality and Social Psychology, 48(4), 813–838.

Sturges, D. (1994). Communicating through Crisis: A Strategy for Organizational Survival. Management Communication Quarterly, 7(3), 297–316.

Subba, R., & Bui, T. (2010). An exploration of physical-virtual convergence behaviors in crisis situations. Proceedings of the 43rd Hawaii International Conference on Systems Sciences, 1–10.

Vraga, E., & Bode, J. (2017). Using expert sources to correct health misinformation in social media. Science Communication, 39(5), 621–645.

Group Decision and Negotiation - Network Analysis of Decisions in Groups -

Conference papers and extended abstracts

Research on Capability Resilience of Heterogeneous Information Combat Networks Based on Link Quality

Renjie Xu [1], Jiahao Liu [1], Hang Yang [1], Xueming Xu [1], Jiang Jiang [1] and Jichao Li [1*]

[1] National University of Defense Technology, Changsha, China
[*] ljcnudt@hotmail.com

Abstract

The combat system-of-systems (CSoS) in high-tech information warfare consists of multiple interconnected combat entities, which can be abstracted as a complex heterogeneous information combat network (HICN). Research on the capability resilience of HICN has substantial military value in terms of optimizing operational processes and enhancing network survivability during military confrontations. Accordingly, this paper presents an integrated framework called HICN capability resilience framework based on network percolation (HICNCR) for assessing the capability resilience of HICN. Specifically, first, a heterogeneous information network model of a combat network is established, considering the heterogeneous of entities and the diversity and quality of information flow. Based on this model, we propose an operational capability resilience index (OCRI) to measure the capability resilience of HICN when confronted with operational tasks. This index integrates networks structure and operational capability to identifies which CSoS is more resilient in when facing identical operational tasks. Finally, we conduct extensive experiments on a military maneuvers case to demonstrate the reliability and effectiveness of the proposed HICNCR.The results offer valuable insights to guide the operation and the design of more resilient CSoS.

Keywords: capability resilience; network percolation; heterogeneous information combat networks; kill chain

1. Introduction

Network-centric operations maximizes the potential of information technology in military operations, transforming information superiority into a competitive advantage and playing a crucial role in military reform. The functions of modern military forces have experienced a significant increase in complexity and diversity, surpassing the capabilities of traditional single-weapon platforms. Consequently, modern operational patterns have shifted from platform-centric warfare to system-of-systems confrontation, calling for a sophisticated and higher-level network architecture. The combat system-of-systems (CSoS), consisting of multiple interconnected combat entities, can be described as a heterogeneous information combat network (HICN) to meet the operational requirement of military confrontations. In this network, nodes represent the combat entities involved, while the links represent the information flows between them.

Network resilience has been a central focal point within the field of networks, spanning diverse domains including ecological networks, transportation network, and infrastructure networks. In the realm of military confrontations, the primary objectives involve safeguarding their own HICN while launching attack to the adversary's HICN. The concept of resilience summarizes and reflects the above objectives. The capability resilience of HICN refers to the ability to recover its operational capabilities and ensure sustained operational effectiveness when a few of their entities are damaged, is closely related to the functional robustness as well as operational capability and recoverability of CSoS. Enhancing the capability resilience of HICN has substantial military value in terms of optimizing operational processes and enhancing network survivability, and has gained increased attention in recent years.

The distinctive configuration and functional attributes of HICN give rise to many challenges when exploring the capabilities resilience of these networks. For general complex network, there exists a multitude of resilience indexes, encompassing natural connectivity, robustness, reliability, and global efficiency. However, most of them are designed for homogeneous networks, thereby restricting their direct applicability to heterogeneous networks. While heterogeneous networks have been used to represent CSoS, this approach often fails to provide a comprehensive understanding of CSoS due to its limitations in furnishing detailed information. As an example, both satellites and radar systems can be categorized as sensor entities. They may belong to different heterogeneous node, as there is a possibility that the degree of satellite nodes within the combat network is significantly higher compared to that of radar nodes. Moreover, although a heterogeneous network can represent the connections between satellite and command systems, it may not effectively differentiate the specific attribute information concerning different radars and command systems, such as the strength of their interconnectivity.

In view of this, this paper aims to investigate the issue of capability resilience in HICN from a system-of-systems perspective. Taking the advantages of the Boyd's mode, as well as kill chain (KC), an integrated framework called HICN capability resilience (HICNCR), is proposed based on network percolation. The major contributions of this study can be outlined as follows.

i) A HICN model is established to provide an accurate and comprehensive semantic information description of the operation process, considering the heterogeneous of entities and the diversity and quality of information flow.

ii) We present an index called operational capability resilience index (OCRI) to assess the capability resilience of HICN. This index directly identifies which CSoS is more resilient in when facing identical operational tasks.

iii) To demonstrate the reliability and effectiveness of the proposed HICNCR, we conduct a military maneuvers case study to assess the capability resilience of the CSoS under different loss scenarios. The results offer valuable insights to guide the operation and the design of more resilient CSoS.

2. Heterogeneous Information Combat Network Model

In modern military operations, the CSoS should integrate and deploy distributed combat entities strategically to maximize operational effectiveness. According the Boyd's model, the operational processes include observation, orientation, decision-making, and action. Accordingly, entities within the CSoS can be categorized into four types: sensor entities (S), decision entities (D), influence entities (I), and target entities (T). Furthermore, the information flows within the CSoS can be classified into six categories: target reconnaissance flow (T→S), reconnaissance sharing flow (S→S), reconnaissance command flow (S→D), command sharing flow (D→D), command influence flow (D→I), and influence target flow (I→T).

The kill chain (KC) represents the sequential stages an adversary or combat entities undergo to identify, track, locate, and eliminate a specific threat or target. A KC process involves sensor entities, like satellites, searching for enemy targets and transmitting intelligence information to decision entities, such as command systems. These decision entities analyze the information and issue orders to influence entities, such as electronic warfare aircraft and self-propelled artillery, to disrupt or damage enemy targets, thereby accomplishing operational tasks. The above description belongs to basic KC, while there is also a generalized KC that includes reconnaissance information sharing within sensor (decision) entities.

The quality of a link is determined by aggregating multiple indicators, where a higher value signifies a stronger correlation between entities. Similar to other dynamic complex networks, combat networks will changes in link quality as a result of attacks or recoveries. To simplify our analysis, we focus solely on the presence and removal of links, the quality of the link can be expressed as

$$C\left(e_k\right)=\begin{cases} w_k, v_k \xrightarrow{yes} v_{k+1} \\ 0, v_k \xrightarrow{no} v_{k+1} \end{cases} \tag{1}$$

3. Operational capability resilience measure of HICN

Traditionally, evaluating the operational capability of KC involves combining combat entity capabilities to determine the overall capability of the combat network. However, combat networks aim to integrate diverse capabilities using information technology to efficiently transfer material, energy, and information flows for optimal benefits. Thus, the emphasis is on the flow rather than entities. Additionally, Hu suggests that links play a critical role in combat networks, often exceeding the significance of nodes. Therefore, we propose a link-focused approach to calculate the operational capability of KC.

Before calculating the operational capability, two assumptions are formulated. Assumption 1: In the case of KC involving multiple reconnaissance sharing flows or command flows, their individual capabilities are considered as the cumulative sum of flows of the same type. Assumption 2: KC with shorter lengths is more effective than those with longer lengths due to their faster execution.

1) Resilience process: During the execution of operational tasks in the CSoS, it is inevitable potential attacks like fire strikes or electromagnetic interference occur. To counter these, measures are taken to recover the CSoS, such as reconstructing the Crimean Bridge or repairing communication lines. These actions can be represented by the removal and recovery of links within the HICN. The occurrence of attacks or recovery processes can lead to changes in the operational capability of the HICN as depicted in Fig. 1. These changes can be categorized into four different phases:

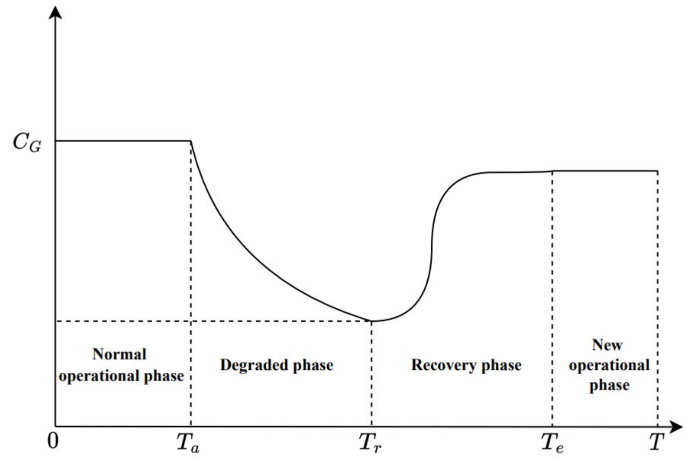


Figure 1: Schematic diagram of HICN resilience process

Normal operational phase (0 ~ T_a): The phase where the HICN has not yet been attacked or is about to face an attack. The HICN will takes actions to defense. If these actions are successful, the operational capability of the HICN remains unaffected. Otherwise, the operational capability of the HICN starts to decline.

Degraded phase (T_a ~ T_r): The phase where operational capability of the HICN begins to decline due to being attacked. We define the resilience at this phase as degraded resilience and where stronger degraded resilience leads to lesser cumulative capability losses.

Recovery phase (T_r ~ T_e): The phase where operational capability of the HICN begins to rise due to be recovered. We define the resilience at this phase as recovery resilience and where stronger recovery resilience leads to more effective restoration of capability.

New operational phase (T_e ~ T): The phase where recovery process concludes, and the operational capabilities remain stable. In this phase, the capability may be lower, equal, or higher in comparison to the initial operational phase.

Based on this, the capability resilience of HICN can be specifically categorized into two dimensions: degraded resilience and recovery resilience. The propose index for evaluating the capability resilience of HICN will focus on these two dimensions, enabling a comprehensive evaluation of overall resilience.

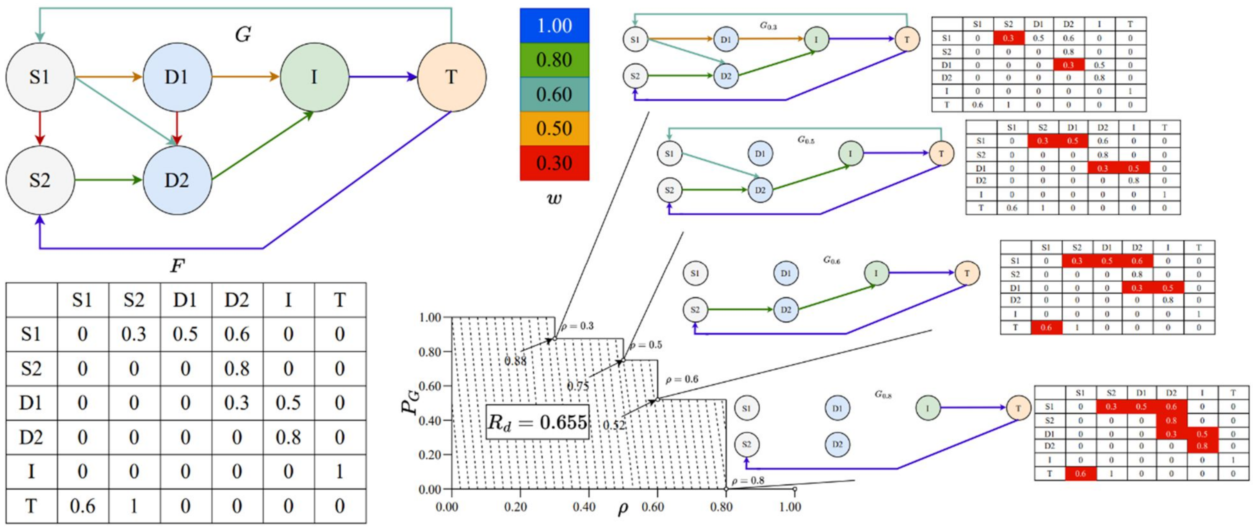
2) Degraded resilience: Degraded resilience in HICN is directly associated with a decline in their operational capabilities. During the execution of operational tasks, adversaries can potentially launch attacks that result in a reduction in the operational capabilities of HICN. In this paper, links removal during the percolation process can be viewed as a hypothetical attack procedure. This procedure involves constructing a set of sub-HICN, denoted as G_ρ, where $\rho \in (0, 1]$, which retains all the links from the original HICN G, except for links with quality $w \leq \rho$. By gradually increasing ρ and removing links with the lowest quality at each step, a collection of sub-HICN, can be extracted. Each sub-HICN represents a distinct operational capability level within the actual HICN. The properties of these sub-HICN can be studied to gain insights into the impact of varying G_ρ on network behavior.

To facilitate subsequent calculations, we employ a normalization process for the operational capabilities of HICN, calculated as

$$P_G = \frac{C_G}{\max(C_G)} \tag{2}$$

where PG represents the operational capability of HICN G.

We define the operational capability at threshold ρ of the percolation process as $P_{G\rho}$. It is instuctive exmaine how $P_{G\rho}$ varies with increasing ρ on the example network shown in Fig. 2, where the original/maximum operational capability is 0.307 by Equation (2-3) and $P_{G0} = 0.307/0.307 = 1$ initially. When $\rho < 0.3$, this does not affect the links between any pair of nodes. When $\rho = 0.3$, two links of the lowest quality are removed and the proportions of affected operational capabilities sum up to 0.038/0.307, thus the P_G drops to $P_{G0.3} = 0.88$.



	S1	S2	D1	D2	I	T
S1	0	0.3	0.5	0.6	0	0
S2	0	0	0	0.8	0	0
D1	0	0	0	0.3	0.5	0
D2	0	0	0	0	0.8	0
I	0	0	0	0	0	1
T	0.6	1	0	0	0	0

Figure 2: Percolation on an example HICN. This HICN G with $n = 6$ nodes and $m = 10$ links, where quality w of each link e is color-coded. Matrix F is used to quantify both the presence and strength of flow between nodes within a network. The percolation process is simulated by gradually raising a threshold ρ, while simultaneously removing any link e_m with weight $w_m \leq \rho$. Sub-HICN G_ρ is visualized at different threshold ρ with its affected(red) flow color-coded in matrix F.

3) Recovery resilience: The recovery resilience of the HICN relies on the improvement of operational capability through link recovery. In general, it is common for attacks to impact multiple links within a HICN. In the case of multiple attacked links, the recovery strategy of HICN depends on the specific sequence in which the faulty links are restored. Thus, the recovery strategy of HICN can be represented as a set of distinct link recovery sequences.

The recovery resilience of HICN can vary depending on the specific sequence of link recovery. Analyzing different fault recovery strategies for the same HICN can be used to determine the most resilient response, but acts as a roadblock when comparing HICN. Hence, we proposed a recovery resilience index to standardize recovery resilience to avoid the above roadblock.

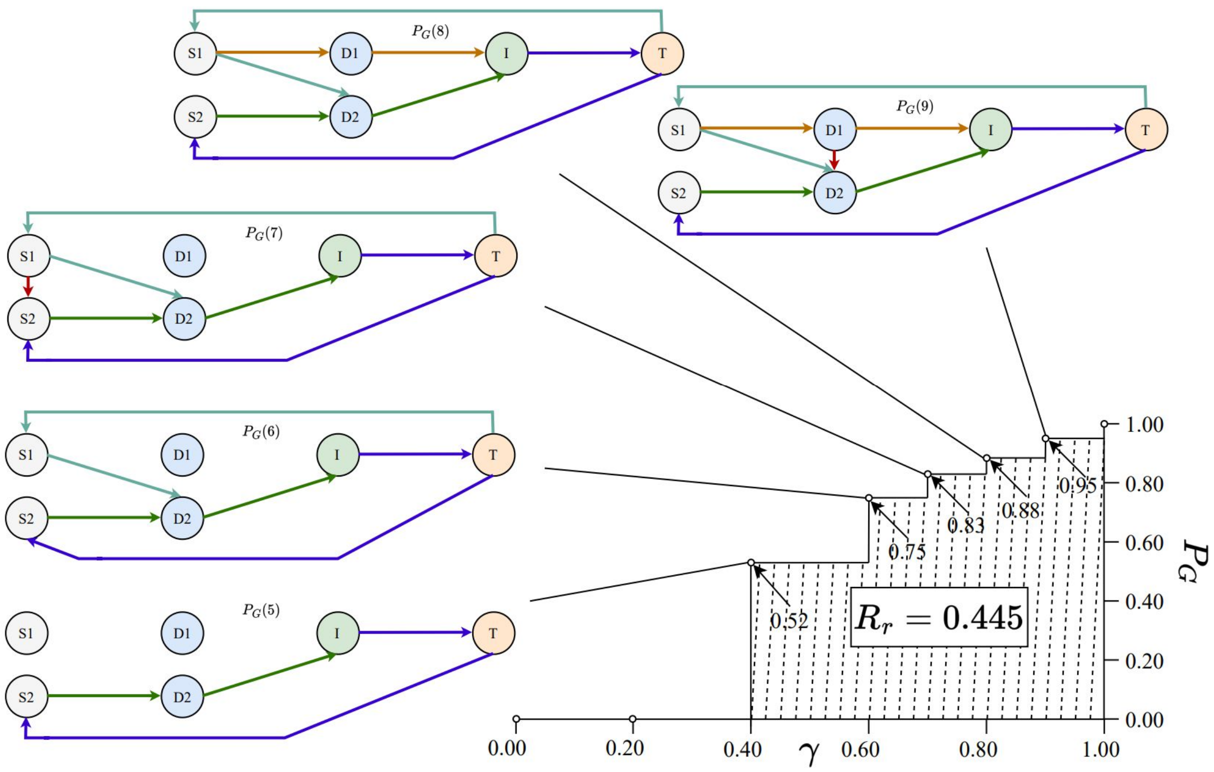


Figure 3: Recovery resilience on an example HICN. This HICN G with n = 6 nodes and m = 10 links. It is assumed that at the beginning, all links were attacked and removed, that is, when $\gamma = 0$, $P_G = 0$. The recovery phase is simulated by gradually raising a recovery strength γ, while simultaneously recovering $E^{\circ} = \gamma \times m$ links and select the maximum value to represent PG at γ.

4. Case Study

To verify the reliability and effectiveness of HICNCR in evaluating the capability resilience of CSoS, a comprehensive series of experiments was undertaken within the context of a HICN. The experimental procedure of the HICNCR is illustrated in Fig. 4. In this procedure, the HICN model of a case of military maneuvers based on heterogeneous information network is first established. Next, we examined three different loss scenarios, namely, node loss, link loss, and quality loss, and then comprehensive experiments were subsequently carried out for each scenario. Finally, we investigated the correlation between the capability resilience of HICN and the network's structure.

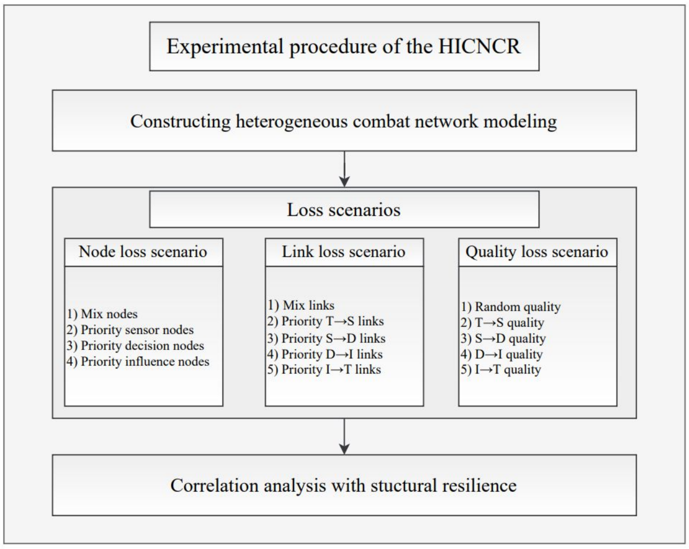


Figure 4: The experimental procedure of the HICNCR

5. Conclusions

Modern operational patterns have shifted from platform-centric warfare to system-of-systems confrontation, calling for a sophisticated and higher-level network architecture. Research on capability resilience of HICN is related to the functional robustness as well as operational capability and recoverability of CSoS. To identify more resilient CSoS when facing identical operational tasks, we present an integrated framework called HICNCR to assess the capability resilience of an HICN.

HICNCR exhibits the following merits: i) HICNCR takes into account the heterogeneity of entities, the diversity and quality of information flow when abstracting a CSoS into a combat network. ii) HICNCR integrates network's structure and operational capability to evaluate the capability resilience. iii) When calculating the capability resilience of HICN, HICNCR effectively avoids the influence of various disturbance and recovery strategies on quantifying resilience.

Consequently, several insightful and valuable conclusions have been drawn from the analysis of the results. For example, the destruction of the network's structure does not necessarily equate to the elimination of the network's capability resilience, and vice versa. The decision entity plays a crucial role similar to that of the human brain, while the T→S link symbolizes the initial step in searching for enemy targets, constituting the first phase in the formation of operational capabilities. Correspondingly, the attackers should emphasize targeting both elements, while the defenders should implement measures to safeguard these crucial components.

When confronted with measures link battlefield environmental changes resulting in a decrease in link quality, attackers should strive to take mix quality loss measures to the greatest extent possible. resource allocation should be tailored to the specific need of reducing link quality to 60%, thus avoiding resource wastage. Likewise, defenders should exert maximum effort to prevent the link quality from declining to 60% threshold.

However, this study is only a basic step in addressing the capability resilience of HICN. In the next step of work, there is a need to further investigate the inherent mechanism of combat network collapse. In addition, exploring methods to enhance capability resilience based on the proposed resilience index is also a challenging topic for future research.

24th International Conference on Group Decision and Negotiation &
10th International Conference on Decision Support System Technology
Human-Centric Decision and Negotiation Support for Societal Transitions

Fair Representation Learning for Enhanced Recommendations: From the Perspective of Users and Items

Mingxing Zhang [1,2], Xiaoxiong Zhang[1,2*], Bingfeng Ge[3], Hui Zhang[1,2]

[1] The Sixty-third Research Institute, National University of Defense Technology, Nanjing, Jiangsu, China
[2] Laboratory for Big Data and Decision, National University of Defense Technology, Changsha, Hunan, China
[3] College of Systems Engineering, National University of Defense Technology, Changsha, Hunan, China
[*] zxxandxx@163.com

Abstract

In the era of artificial intelligence, recommendation systems are widely applied across various domains, and ensuring the fairness of recommendations is crucial. The predominant approach for implementing fairness in existing algorithms involves mitigating the impact of sensitive information. However, striking a balance between recommendation accuracy and fairness remains a challenge. In this work, our main contribution is from the perspective of user and item fairness, striving to minimize the unfairness caused by long-tail items and user group differences while maintaining recommendation accuracy. Specifically, we accomplish the recommendation task with the help of a pre-training fine-tuning paradigm. First, the node degree is adjusted to increase the probability of long-tailed items being exposed during the pre-training phase of information aggregation on the user-item graph. Second, it mitigates the unfairness caused by the differences in user groups by referring to the word-embedding debiasing method in the fine-tuning phase. The orthogonality of the bias vectors is maintained to achieve decoupling between features for further fairness. Experimental validation on the ml-100k and ml-1m datasets demonstrates the superiority of our model.

Keywords: Recommendation Systems; Fairness; Sensitive Information; Long-Tail Items

1. Introduction

With the advent of information overload, recommender systems are widely used in various platforms such as e-commerce (Feng et al., 2020), employment (Gutiérrez et al., 2019), and healthcare (Gong et al., 2021). Although these systems can help us quickly get useful information from a sea of data, they also create problems of unfairness and bias. The unfairness of recommender systems can be divided into two aspects. On the one hand, there is a long-tail distribution problem of recommended items, different items are unfairly subjected to the recommender system due to their popularity. However, most items have only relatively low popularity, which is often ignored in the subsequent recommendations. On the other hand, users may be influenced by their gender, age, cultural background, and other factors. When choosing and accepting recommended content, thus forms a tendency. This influence is particularly prominent in employment selection (Islam et al., 2021). The design of recommender systems should consider how to balance the individualized needs of users and the diversity of recommendations. Ensure that the recommender system is fair to all users and avoids bias or unfair treatment of certain groups.

The lifecycle of a recommender system is abstracted as a feedback loop between three key components: users, data, and models. Data is collected from the user, including information such as user-item interactions, followed by learning a recommendation model based on the collected data; and the recommendation results are returned to the user to fulfill the user's needs. The recommended results will in turn influence the user's future behavior and decisions. As shown in Figure 1, when bias occurs, the exposure to non-popular items will gradually decay or even disappear in the subsequent recommendation process. From the perspective of the items, the recommender system is a feedback loop that can exacerbate this unfairness and bias problem. Similarly, from the user's point of view, the bias towards users' sensitive attributes will be gradually amplified in the recommendation process, thus affecting users' satisfaction with the recommendation results. The study of fairness in recommender systems has become an important area. In addition, the study of fair recommender systems brings significant value to the application areas of group decision-making and negotiation as well as decision support systems technology, where fair recommendations can ensure that the advice provided to a group is fair and balanced for all members, thus promoting a fairer group decision-making process.

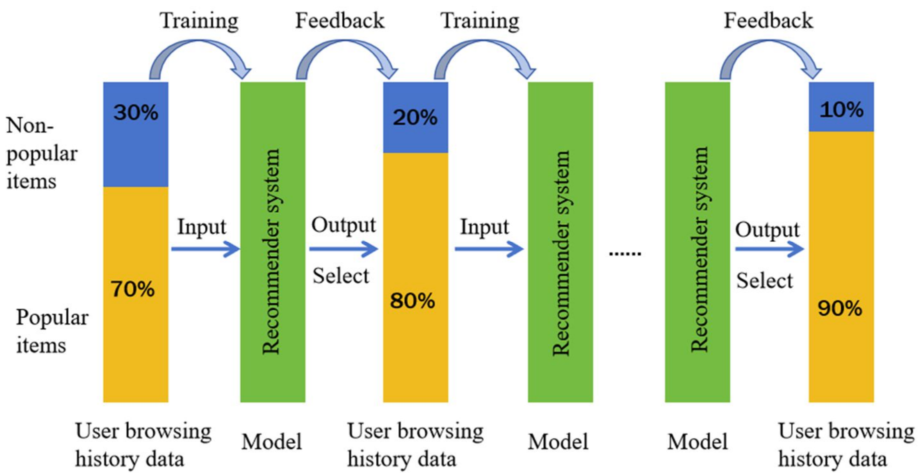


Figure 1: Bias Amplification Issues in Recommender Systems

Nowadays, machine learning algorithms are indispensable when mentioning recommender systems. Methods to achieve fairness in machine learning can be divided into three categories: pre-processing, in-processing, and post-processing (Li et al., 2023). Pre-processing refers to the processing of the data to be trained, in which the data is de-biased to remove data that causes unfair recommendations The model is trained on corrected unbiased data, (Alessandro et al., 2017), so it is naturally fair. This approach specifies processing for different datasets and may not eliminate unfairness in test data. Post-processing methods refer to the reordering of the recommended results to achieve fairness, and generally require consideration of increasing the weight of a few sample exposures; excessive post-processing may result in a loss of personalization of the recommended results. In-processing methods refer to those that achieve fairness from the perspective of model training by incorporating the fairness metric into the objective function of the main learning task, e.g., (Wu et al., 2021) (Wang et al., 2023). The main difficulty of this strategy lies in balancing accuracy and fairness, with higher flexibility through suitable recommendation algorithms and tuning of parameters.

Compared to pre-processing and post-processing, the advantage of in-processing to achieve fairness in recommender systems is that it can integrate fairness constraints more flexibly and deeply into the model training process, thus realizing more fine-grained performance trade-offs and adaptations to ensure comprehensive fairness of the recommendation results. Therefore, in this paper, we realize the fairness of recommender systems by in-processing. We achieve both user-side and item-side fairness and maintain the accuracy of fair recommendations, Specifically, we work as follows:

a) The representation vectors are pre-trained using the neighborhood aggregation mechanism of the user-item graph, and attention is given to long-tail items by adjusting the degree distribution of nodes in the interaction matrix. Thereby increasing the exposure probability of non-popular items.

b) The influence of user-sensitive attributes is eliminated by employing the orthogonalization debiasing method. This helps avoid introducing unnecessary correlations in the fine-tuning stage of the model, ensuring the accuracy of the model.

c) The effectiveness of the model in this paper was verified through experiments on the ml-100k and ml-1m datasets, achieving a balance between accuracy and fairness.

2. Related Work

In this section, we first introduce the definition and formulaic representation of recommendation tasks, and then briefly highlight related work on fair recommendation algorithms and common fairness metrics, which are the basis of our research.

2.1. Definition of Task

Given a set of users $U = \{u_1, u_2 \ldots u_m\}$ and a set of items $I = \{i_1, i_2 \ldots i_n\}$ in a recommender system, let M and N denote the number of users and the number of items. The user-item interaction matrix $Y \in R^{M \times N}$ is defined based on the implicit feedback $Y = \{y_{ui} \mid u \in U, i \in I\}$, and if u has clicked i, $y_{ui} = 1$ otherwise $y_{ui} = 0$.

The task of a recommender system is to predict the probability of generating interaction between a user and an item and to give a list of recommended items of specified length K based on the predicted probability score. The fair recommendation system aims to realize user fairness and item fairness, ensure that the recommendation results are not affected by the sensitive attributes of users, and at the same time, ensure that all kinds of items have equal opportunities to be recommended, to provide users with fair and unbiased recommendation services.

2.2. Recommendation Fairness

Fairness recommendation algorithms aim to ensure equal treatment among different groups. Current algorithms that promote fairness mainly formulate various fairness criteria as constraints to guide model optimization. (Yao et al., 2017) designed four fairness metrics and incorporated them into collaborative filtering algorithms aiming to ensure fairness in the context of the recommendation function. (Li et al., 2021) generated a fairness-aware recommendation list based on a fairness-constrained reordering method. This not only improves the group fairness of users but also achieves better overall recommendation performance. (Cai et al., 2022) used a multi-objective optimization algorithm to reasonably trade off fairness, accuracy, and interpretability. The challenge with this type of approach is that the training process is unstable and it is difficult to balance fairness and recommendation performance. Another way to achieve fair recommendations is adversarial training (Goodfellow et al., 2020). (Beigi et al., 2020) trained both a discriminator and a filter using adversarial training. The generator strives to produce a representation that does not contain sensitive information, while the discriminator strives to identify sensitive information from the generated representation. (Bose et al., 2019) introduced an adversarial framework to strengthen the fairness constraints of graph embedding, which can flexibly adapt to different combinations of fairness constraints. In addition, methods such as causal analysis (Chiappa et al., 2019) (Wang et al., 2021), and reinforcement learning (Jabbari et al., 2017) (Ge et al., 2022) have been applied to fair recommendation.

2.3. Fairness Metrics

Fairness metrics are used to assess fairness performance in recommender systems. These metrics can help researchers and practitioners understand whether algorithms are unfair in distributing recommendations among users or groups. In terms of item fairness, one approach utilizes the Gini Index (Charles et al., 2022) or measures the unequal distribution of certain users or items. A lower Gini Index indicates a relatively even distribution of resources (Sun et al., 2019). And user fairness is often associated with sensitive attributes. Value Unfairness is a concept used to describe the unfairness in a sensitive attribute such as gender. When it comes to unfairness in sensitive attributes, this may include different treatment of different attribute values. Value inequity occurs when a class of users is consistently predicted above or below their true preferences. If the errors in the predictions are balanced between overestimation and underestimation, then the inequity of values will be small.

$$U_{value} = \frac{1}{n}\sum_{j=1}^{n} \left| \Delta \mathrm{E}_g[\mathrm{y}]_\mathrm{j} - \Delta \mathrm{E}_{\neg g}[\mathrm{y}]_\mathrm{j} \right| \tag{1}$$

where $\Delta \mathrm{E}_g[\mathrm{y}]_\mathrm{j}$ and $\Delta \mathrm{E}_{\neg g}[\mathrm{y}]_\mathrm{j}$ denote the difference between the actual and predicted scores of item j in the different groups.

Absolute unfairness, which measures the inconsistency of absolute estimation errors across user types, is calculated as

$$U_{absolute} = \frac{1}{n}\sum_{j=1}^{n} \| \Delta E_g[y]_j | - | \Delta E_{\neg g}[y]_j \| \tag{2}$$

If this value is large, it indicates a large absolute unfairness. To measure value unfairness or absolute unfairness between multiple groups, the difference in unfairness between each pair of groups is calculated and the largest of these differences is taken as the measure of unfairness. Other methods such as popularity which calculates the ratio of popular items in the recommended list to the total number of items in the list are considered as a fair measure of popularity level (Ge et al., 2021). Underestimation unfairness measures the inconsistency of predictions underestimating true ratings (Yao et al., 2017). Differential fairness (Islam et al., 2021) focuses on whether the system behaves consistently under different attributes or conditions and whether different groups can be treated similarly under similar circumstances.

3. The Framework

We give the framework proposed in this paper. The new contribution of this section is the proposal of guaranteeing recommendation performance while achieving fairness at both user and item ends.

3.1. Overall Architecture

Most of the fairness recommender systems are considered from one side of users or items, this paper tries to design a model that can consider the unfairness of two sides at the same time. The overall framework is shown in Fig. 2, which is realized using pre-training and fine-tuning. The pre-training stage is realized by neighborhood aggregation through the binomial graph formed between users and items. To solve the unfairness of recommended items, we add the process of neighborhood aggregation to re-adjust the degree distribution of the nodes and increase the user's attention to the long-tail items. In the fine-tuning stage, we implement the debiasing of user-sensitive attributes, avoiding the introduction of unnecessary correlations in the model fine-tuning stage by orthogonalization debiasing to maintain the accuracy of the model. The specific details are explained in detail in the subsequent subsections.

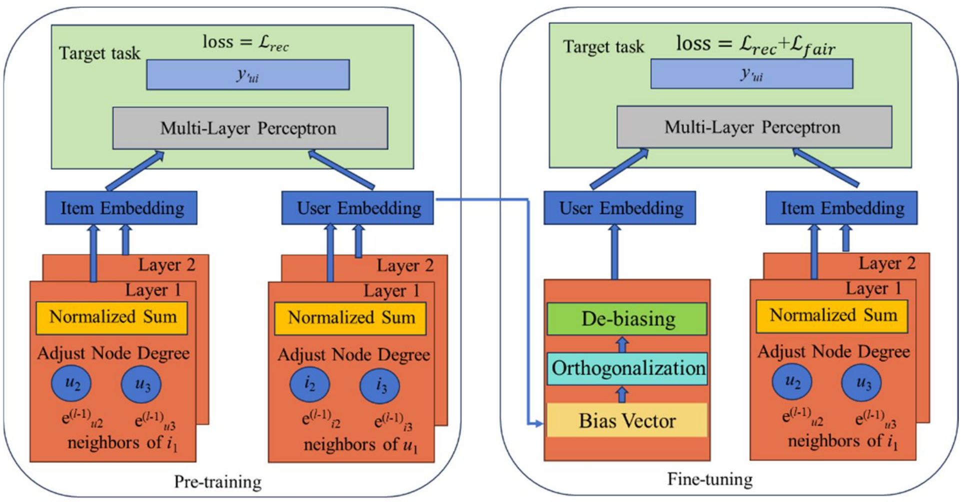


Figure 2: The Overall Architecture of Our Model

3.2. Pre-training Stage

Among recommender systems, users and items are binomial graphs, and a Graph Convolutional Network (GCN) is a neural network structure that has been gaining popularity in recent years. This kind of network is capable of deep learning on graph data, processing data with graph structure and deeply discovering its features and patterns. Therefore GCN can be considered to improve the richness of node representations by capturing more levels of neighbor information. Its core idea is to iteratively perform graph convolution to aggregate the features of neighbors into a new representation of the target node. The LightGCN model (He et al., 2020) verified that the feature transformation or nonlinear activation in the process of neighbor aggregation may be a burden for collaborative filtering. Removing these two operations can simplify the GCNs to a great extent therefore, this model implements the pre-training process based on the LightGCN model. The details are as follows:

Let the user-item interaction matrix be $Y \in R^{M \times N}$, where M and N denote the number of users and items, respectively. The adjacency matrix of the user-item graph is then obtained as

$$A = \begin{pmatrix} 0 & \mathrm{R} \\ \mathrm{R}^{\mathrm{T}} & 0 \end{pmatrix} \tag{3}$$

Let the layer 0 embedding matrix be $\mathrm{E}^{(0)} \in R^{(\mathrm{M+N}) \times \mathrm{T}}$, where T is the embedding size. Then we can get the matrix equivalent form of the embedding vector of the Kth layer as:

$$\mathrm{E}^{(k+1)} = (\mathrm{D}^{-\frac{1}{2}} \mathrm{A}^{\beta} \mathrm{D}^{-\frac{1}{2}}) \mathrm{E}^{(k)} = \tilde{\mathrm{A}} \mathrm{E}^{(k)} \tag{4}$$

where A is the adjacency matrix and D is the degree matrix (the diagonal elements are the degrees of each node). $\tilde{\mathrm{A}}$ is the normalized transpose matrix and β is the exponential weight parameter. When 0<β<1, the exponential operation strengthens the smaller elements and reduces the larger ones. We add a fixed value to each element of the adjacency matrix A to make β effective. This helps to balance the larger and smaller values in the adjacency matrix, and for nodes with a more uniform degree distribution, this adjustment may be more beneficial for long-tailed item fairness.

$$\mathrm{E} = \alpha_0 \mathrm{E}^{(0)} + \alpha_1 \mathrm{E}^{(1)} + \ldots + \alpha_{\mathrm{K}} \mathrm{E}^{(\mathrm{K})} = \alpha_0 \mathrm{E}^{(0)} + \alpha_1 \tilde{\mathrm{A}} \mathrm{E}^{(1)} + \ldots + \alpha_{\mathrm{K}} \tilde{\mathrm{A}}^{(\mathrm{K})} \mathrm{E}^{(\mathrm{K})} \tag{5}$$

The MLP (Multi-Layer Perceptron) structure is relatively simple and easy to implement and integrate into existing recommender systems, so the embedding vectors of users and items are subsequently fed into the MLP to adapt to the recommendation task and complete the pre-training. Of course, other complex machine learning models can also be an option.

3.3. Fine-tuning Stage

The fine-tuning stage focuses on debiasing user attributes to ensure that the model’s predictions or recommendations for different user groups are fair and non-discriminatory in the modelling and decision-making process. De-biasing helps to deal with potential bias or discrimination to ensure that the model is fair to all users. (Dev & Phillips, 2019) showed that a very simple linear projection of all words based on vectors captured by common names is an effective and general way to significantly reduce bias in word embeddings. We take inspiration from the word embedding debiasing direction of natural language processing and apply the method to user debiasing in recommender systems, taking the gender attribute of users as an example. Since some users may not have uploaded their gender information, by dividing the users into male and female groups, as well as an uncertain group, the average vector for the male user group can be calculated as

$$v_m = \frac{1}{N_m} \sum_{i=1}^{N_m} \mathrm{e}[i] \tag{6}$$

where $\mathrm{e}[i]$ is the pre-trained user vector, N_m is the number of male users.

A similar operation yields the mean vector v_f for the female group and the mean vector v_{un} for the uncertain group. Finally, we compute the overall gender bias vector

$$v_{\mathrm{bias\_1}} = v_f - v_{un} \tag{7}$$

$$v_{bias\_2} = v_{un} - v_m \tag{8}$$

$$v_{bias\_3} = v_m - v_f \tag{9}$$

If the distribution of user embeddings is changed only by adjusting their mean values, but decoupling between features is not explicitly considered. This may result in the features learned by the model not being sufficiently independent in space, thus affecting its generalization ability. For this reason, we use the Gram-Schmidt orthogonalization method to orthogonalize the generated bias vectors. Then we were able to get the final user representation

$$u[i] = e[i] - \alpha(\sum_{j=1}^{3} \langle e[i], v_{\text{bias\_j}} \rangle \cdot v_{\text{bias\_j}}) \tag{10}$$

where α is an adjustable parameter, $v_{\text{bias\_j}}$ is the orthogonalized bias vector, and $\langle e[i], v_{\text{bias\_j}} \rangle$ is the inner product of the pre-trained user vectors and bias vectors.

3.4. Objective Function

The model is realized through pre-training and fine-tuning. The purpose of pre-training is to build an initial representation of the model so that it can be better adapted to the recommendation task. We use the cross-entropy loss L_{rec} as the target for the recommendation task. Calculate y'_{ui} using the inner product function. For the case where $y_{ui} = 1$, the predicted value y'_{ui} approaches 1, then the value of the loss function should approach 0.

$$\mathrm{L}_{rec} = -\frac{1}{N}\sum_{i=1}^{N} y_{ui} \cdot \log(y'_{ui}) + (1 - y_{ui}) \cdot \log(1 - y'_{ui}) \tag{11}$$

In the fine-tuning stage, the main recommender system loss function can be maintained and a fairness regularization loss can be added, which can be used to constrain the model to be fairer in its recommendation results for different users or items. We use a differentiated fairness calculation (Islam et al., 2021) based on the Dirichlet distribution.

$$e^{\varepsilon} \geq \left| \frac{P_{M,\theta}(M(\mathrm{x}) = y \mid g_i)}{P_{M,\theta}(M(\mathrm{x}) = y \mid g_j)} \right| \tag{12}$$

where $M(\mathrm{x})$ is the recommendation algorithm mechanism that inputs personal data x and outputs a predictive score y for the recommendation. θ is the hyperparameters of the model, the characteristics of the input data, the configuration of the algorithm, etc. $P_{M,\theta}(M(\mathrm{x}) = y \mid g, \theta)$ indicate the user prediction scores for different groups. Suppose all of the $P_{M,\theta}(M(\mathrm{x}) = y \mid g, \theta)$ probabilities are equal for each group g_i or g_j, $\varepsilon = 0$, otherwise $\varepsilon > 0$. This calculation can be extended to multiple user groups by calculating the average value between any two groups. We penalize the mean of the n-item.

$$\mathrm{L}_{fair} = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i \tag{13}$$

Therefore, the overall loss in the fine-tuning phase can be $\mathrm{L} = \mathrm{L}_{rec} + \lambda \cdot \mathrm{L}_{fair}$ and λ is an adjustable parameter.

4. Experiments

4.1. Experimental Setup

Experiments were conducted on two public datasets: ml-100k and ml-1m, which are two movie recommendation datasets of different sizes containing sensitive information such as the user's gender, age, etc., as well as the user's rating data for different movies. There are a large number of long-tail items and different user groups in these two datasets.

Table 1. Statistics of The Datasets

Dataset	User	Item	Interaction
ml-100k	943	1682	100000
ml-1m	6040	3706	1000208

For experiments, we divided each dataset in the ratio of 80%:10%:10% to obtain the training, validation, and test sets. When the rate does not improve for 20 epochs on the validation set, the training is terminated.

4.2. Baseline Model

Compare our model with the following baseline models, which include NFCF (Islam et al., 2021), FairGo (Wu et al., 2021), FOCF (Yao et al., 2017), and PFCN (Li et al., 2021).

- NFCF developed neural fairness collaborative adjustment to mitigate gender bias when recommending sensitive items and enhance bias correction techniques.
- FairGo used adversarial learning to embed the original embedding space for each user and each item into a filtered embedding space based on a sensitive feature set.
- FOCF investigates fairness in collaborative filtering recommender systems, optimized by adding a fairness term to the learning objective.
- PFCN achieved counterfactual fairness in recommendation through adversarial learning. The framework allows recommender systems to achieve user-personalized fairness.

4.3. Evaluation Metrics

For each user in the dataset, the items that the user has interacted with are considered as positive sample items and the rest are considered as negative sample items. In the recommendation task, our goal is to select the highest-rated items to form a Top K recommendation list to fulfill the user's needs. In the experiment, the common hit@K is chosen to measure the performance of recommendation, which indicates whether the recommendation list contains the user's favorite items or not. The larger the value the better the recommendation performance. Three metrics are used to measure the fairness of recommendations, whereas the Gini index is used to measure the inequality of the distribution of items in the recommendation list. Value unfairness is a concept used to describe the unfairness in a sensitive attribute (e.g., gender). Absolute unfairness measures the inconsistency of the absolute estimation error across different user groups. The smaller the value of the three metrics, the fairer the recommendation algorithm proves to be.

4.4. Parameters Setting

The models are implemented by PyTorch, and the RecBole2.0 (Zhao et al., 2022) framework was borrowed for baseline model evaluation. To ensure a fair comparison, we uniformly set the embedding size to 64, using the Adam optimizer with a batch size of 1024 and a learning rate of 0.001, and for other parameters provided in the original paper for tuning on the two datasets. For the model, if not specified, the number of neighborhood aggregation layers is 2, the fairness weight is 0.1, the hidden layer is [128,64], and the gender is chosen as the user's sensitive information for the experiment.

4.5. Performance Comparison

We will present the main results of our experiments in Table 2, including a comparison of the recommended performance and fairness of all baseline models.

Our model is fair at both the user and item ends, while achieving excellent performance on hit@5. This performance excellence is due to our pre-training and fine-tuning process. We train by adjusting the node degree during pre-training instead of adding additional non-interactive items, which maintains the recommended hit rate to some extent. In the fine-tuning stage, we also avoid introducing additional correlations between user vectors that can affect the recommendation results by orthogonalizing the bias vectors. From the above two aspects, the model achieves better recommendation performance. This is consistent with our research goal of achieving fairness at both the user and program ends, while ensuring the accuracy of the recommender system.

Table 2. Overall Performance Comparison

Dataset	Model	hit@5↑	Gini↓	Value↓	Absolute↓
ml-100k	FairGo	0.5387	0.9808	0.2280	**0.1607**
	FOCF	0.4443	0.9782	0.2500	0.1854
	PFCN	0.5875	0.9810	0.2379	0.1908
	NFCF	0.8526	0.9371	0.2269	0.1654
	Our Model	**0.8632**	**0.9324**	**0.2254**	0.1625
ml-1m	FairGo	0.5045	0.9879	0.1335	0.1206
	FOCF	0.5050	0.9879	0.1299	0.1028
	PFCN	0.6581	0.9872	0.1271	0.1264
	NFCF	0.7901	0.9670	0.1249	0.0969
	Our Model	**0.8018**	**0.9659**	**0.1179**	**0.0965**

Overall, FOCF performs poorly on both datasets, mainly because traditional collaborative filtering algorithms are weaker than neighborhood aggregation-based embedding methods in terms of recommendation performance. In addition to the fact that the incorporation of fairness into the main task for simultaneous optimization leads to the training results appearing to be at a local optimum.

The FairGo model can achieve excellent fairness in the case of small datasets of ml-100k, PFCN has better fairness under large dataset of ml-ml. However, both models sacrifice the hit rate of recommendations and performance varies greatly between datasets. This is because such adversarial training may introduce training instability, causing convergence to become more difficult.

The NFCF model, its performance is better compared to other baseline models, the model is also realized through the process of pre-training plus fine-tuning. In contrast, the interaction information between users and items is ignored in the pre-training process, which is very helpful for the improvement of the recommendation performance. At the same time, the process of users' de-biasing inevitably loses the recommendation performance, and the NFCF model has no compensatory measures in this area.

Then, we mainly observe the changes in the metrics in the ml-100k dataset under different K values in Figure 3. The values of the three non-fairness metrics can always be at lower values when the hit rate of the model in this paper is kept optimal.

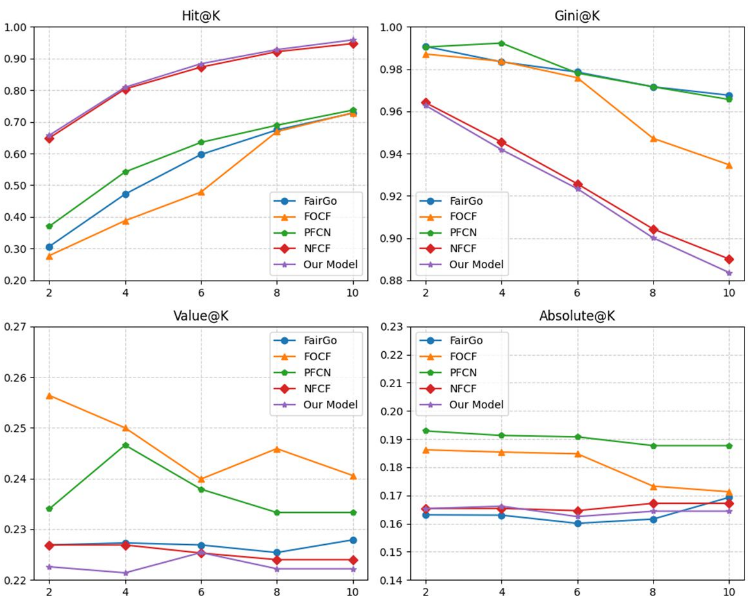


Figure 3: Performance Comparison at Different K Values

4.6. Ablation Experiment

In ablation experiments, we focus on measuring the impact of the item fairness and user fairness modules on performance. For this purpose, we set up three variants of the model, variant 1 is the performance in the pre-training stage. Variant 2 is the pre-training stage with the addition of a long-tailed item distribution solution. Variant 3 is the fine-tuning process with the addition of gender orthogonalization debiasing. Theoretically, when we add item debiasing, the Gini index, an indicator of item fairness, decreases to some extent. When we add user gender debiasing, both value unfairness and absolute unfairness decrease. Our model combines the above two items to achieve a better hit rate and three unfairness metrics.

Table 3. Performance Comparison of Our Model and Variants On ml-1m

Model	hit@5↑	Gini↓	Value↓	Absolute↓
variant 1	0.8171	0.9753	0.1262	0.1065
variant 2	0.8077	0.9698	0.1285	0.1068
variant 3	0.8003	0.9708	0.1184	0.0961
Our Model	0.8018	0.9659	0.1179	0.0965

The results of the experiments are as follows: variant 1 has several fairness metrics higher than the final model, variant 2 has a lower Gini coefficient than our model, and variant 3 achieves better user fairness with less value unfairness and absolute unfairness. In addition, the hit rate metrics decrease slightly in the process of achieving fairness, which is consistent with the theoretical analysis.

5. Conclusions

In this paper, we propose a way to consider fairness from both the user and item ends. It can be easily applied to recommender systems through pre-training and fine-tuning. We achieve fairness in recommendations by adjusting the node degree and keeping the bias vectors orthogonal. Moreover, these two operations can minimize the cost of accuracy. Extensive experiments on two real-world datasets show that the recommendation system outperforms existing baseline models in terms of accuracy and also maintains good fairness. In addition, the impact of user fairness and item fairness modules was demonstrated through ablation experiments. In this paper, although an effective fairness recommendation method is proposed, the model still has room for improvement in feature extraction and embedding generation. Although we used advanced graph convolution techniques, there is still potential to further improve the quality of embeddings by introducing more contextual information, dynamic factors, or knowledge graphs. Future research will consider exploring new and innovative embedding generation methods to improve the ability of recommender systems to model users and items.

Acknowledgments

The research was supported by the National Natural Science Foundation of China (71901215), the Young Elite Scientists Sponsorship Program in Specific Field (2021-JCJQ-QT-050).

References

Alessandro, B., Neil, C., & LaGatta. (2017) Conscientious Classification: A Data Scientist's Guide to Discrimination-Aware Classification," *Big Data*, 5(2), Pages. 120-134.

Beigi, G., Mosallanezhad, A., Guo, R., Alvari, H., Nou, A., & Liu, H. (2020). Privacy-aware recommendation with private-attribute protection using adversarial learning. *In Proceedings of the 13th ACM International Conference on Web Search and Data Mining. Association for Computing Machinery*. Houston, TX, USA. Pages. 34–42.

Bose, A. J., & Hamilton, W. L. (2019). Compositional fairness constraints for graph embeddings. *In Proceedings of the International Conference on Machine Learning.*

Cai, X., Xie, L., Tian, R., Cui, Z. (2022). Explicable Recommendation Based on Knowledge Graph. *Expert Syst.* Appl., 200(8), Pages. 1-11.

Charles, V., Gherman, T., & Paliza, J. C. (2022). The Gini Index: A Modern Measure of Inequality. *In V. Charles & A. Emrouznejad, Modern Indices for International Economic Diplomacy*. Springer International Publishing. Pages 55–84.

Chiappa, S., & Isaac, W. S. (2019). A causal Bayesian networks viewpoint on fairness. *In Privacy and Identity Management. Fairness, Accountability, and Transparency in the Age of Big Data: 13th International Summer School*, Vienna, Austria, Pages. 3–20.

Dev, S., & Phillips, J. (2019). Attenuating bias in word vectors. *In Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*. Pages 879–887.

Feng, Y., Hu, B., Lv, F., Liu, Q., Zhang, Z., & Ou, W. (2020). ATBRG: Adaptive Target-Behavior Relational Graph Network for Effective Recommendation. *In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. Association for Computing Machinery*. New York, NY, USA. Pages 2231-2240.

Ge, Y., Liu, S., Gao, R., Xian, Y., Li, Y., Zhao, X., Pei, C., Sun, F., Ge, J., Ou, W., & Zhang, Y. (2021). Towards long-term fairness in recommendation. *In Proceedings of the Fourteenth ACM International Conference on Web Search and Data Mining. Association for Computing Machinery*. Virtual Event, Israel. Pages 445–453.

Ge, Y., Zhao, X., Yu, L., Paul, S., Hu, D., Hsieh, C.-C., & Zhang, Y. (2022). Toward Pareto efficient fairness-utility trade-off in recommendation through reinforcement learning. *In Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining. Association for Computing Machinery*. Virtual Event, AZ, USA. Pages 316–324.

Gong, F., Wang, M., Wang, H., Wang, S., & Liu, M. (2021). SMR: Medical Knowledge Graph Embedding for Safe Medicine Recommendation. *Big Data* Research, 23, 100174.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). *Generative adversarial networks. Communications of the ACM*, 63(11), 139–144.

Gutiérrez, F., Charleer, S., De Croon, R., Htun, N. N., Goetschalckx, G., & Verbert, K. (2019). Explaining and exploring job recommendations: A user-driven approach for interacting with knowledge-based job recommender systems. *In Proceedings of the 13th ACM Conference on Recommender Systems. Association for Computing Machinery*. New York, NY, USA. Pages 60-68.

He, X., Deng, K., Wang, X., Li, Y., Zhang, Y., & Wang, M. (2020). LightGCN: Simplifying and powering graph convolution network for recommendation. *In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. Association for Computing Machinery*. Virtual Event, China. Pages 639–648.

Islam, R., Keya, K. N., Zeng, Z., Pan, S., & Foulds, J. (2021). Debiasing career recommendations with neural fair collaborative filtering. *In Proceedings of the Web Conference 2021 Association for Computing Machinery*. New York, NY, USA. Pages 3779-3790.

Jabbari, S., Joseph, M., Kearns, M., Morgenstern, J., & Roth, A. (2017). Fairness in reinforcement learning. *In Proceedings of the 34th International Conference on Machine Learning.* JMLR.org. Sydney, NSW, Australia. Pages 1617–1626.

Li, Y., Chen, H., Fu, Z., Ge, Y., & Zhang, Y. (2021). User-oriented fairness in recommendation. *In Proceedings of the Web Conference 2021*. Association for Computing Machinery. Ljubljana, Slovenia. Pages 624–632.

Li, Y., Chen, H., Xu, S., Ge, Y., Tan, J., Liu, S., & Zhang, Y. (2023). Fairness in Recommendation: Foundations, Methods, and Applications. *ACM Transactions on Intelligent Systems and Technology*, 14(5), Pages 1-48.

Li, Y., Chen, H., Xu, S., Ge, Y., & Zhang, Y. (2021). Towards personalized fairness based on causal notion. *In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. Association for Computing Machinery*. Pages 1054–1063.

Sun, W., Khenissi, S., Nasraoui, O., & Shafto, P. (2019). Debiasing the human-recommender system feedback loop in collaborative filtering. *In Companion Proceedings of the 2019 World Wide Web Conference. Association for Computing Machinery.* San Francisco, USA. Pages 645–651.

Wang, J., Ma, W., Jiang, C., Zhang, M., Zhang, Y., Li, B., & Jiang, P. (2023). Measuring item global residual value for fair recommendation. *In Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval. Association for Computing Machinery*. Taipei, Taiwan. Pages 269–278.

Wang, W., Feng, F., He, X., Zhang, H., & Chua, T.-S. (2021). Clicks can be cheating: Counterfactual recommendation for mitigating clickbait issue. *In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. Association for Computing Machinery*. Virtual Event, Canada.

Pages 1288–1297.

Wu, L., Chen, L., Shao, P., Hong, R., Wang, X., & Wang, M. (2021). Learning fair representations for recommendation: A graph-based perspective. In Proceedings of the Web Conference 2021. *Association for Computing Machinery*. Ljubljana, Slovenia. Pages 2198-2208.

Yao, S., & Huang, B. Beyond Parity: Fairness Objectives for Collaborative Filtering. (2017). *In Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, California, USA, Pages 2925-2934.

Zhao, W. X., Hou, Y., Pan, X., Yang, C., Zhang, Z., Lin, Z., Zhang, J., Bian, S., Tang, J., Sun, W., Chen, Y., Xu, L., Zhang, G., Tian, Z., Tian, C., Mu, S., Fan, X., Chen, X., & Wen, J. R. (2022). RecBole 2.0: Towards a more up-to-date recommendation library. *In Proceedings of the 31st ACM International Conference on Information & Knowledge Management. Association for Computing Machinery*. Pages 4722–4726.

Group Decision and Negotiation - Responsible NSS in the age of Generative AI-

Conference papers and extended abstracts

24th International Conference on Group Decision and Negotiation &
10th International Conference on Decision Support System Technology
Human-Centric Decision and Negotiation Support for Societal Transitions

Can AI with High Reasoning Ability Replicate Human-like Decision Making in Economic Experiments?

Ayato Kitadai [1*], Sinndy Dayana Rico Lugo [1], Yudai Tsurusaki [1], Yusuke Fukasawa [1], and Nariaki Nishino [1]

[1] The University of Tokyo, Tokyo, Japan
[*] a.kitadai@css.t.u-tokyo.ac.jp

Abstract

Economic experiments offer a controlled setting for researchers to observe human decision-making and test diverse theories and hypotheses; however, substantial costs and efforts are incurred to gather many individuals as experimental participants. To address this, with the development of large language models (LLMs), some researchers have recently attempted to develop simulated economic experiments using LLMs-driven agents, called generative agents. If generative agents can replicate human-like decision-making in economic experiments, the cost problem of economic experiments can be alleviated. However, such a simulation framework has not been yet established. Considering the previous research and the current evolutionary stage of LLMs, this study focuses on the reasoning ability of generative agents as a key factor toward establishing a framework for such a new methodology. A multi-agent simulation, designed to improve the reasoning ability of generative agents through prompting methods, was developed to reproduce the result of an actual economic experiment on the ultimatum game. The results demonstrated that the higher the reasoning ability of the agents, the closer the results were to the theoretical solution than to the real experimental result. The results also suggest that setting the personas of the generative agents may be important for reproducing the results of real economic experiments. These findings are valuable for the future definition of a framework for replacing human participants with generative agents in economic experiments when LLMs are further developed.

Keywords: large language models; economic experiment; multi agent simulation; ultimatum game; decision making

1. Introduction

Economic experiments are a relevant approach in economics. They provide researchers with a controlled environment to observe and analyze human behavior in response to various economic conditions, test theories and hypotheses, assess the impact of policy changes, and understand decision-making processes. However, gathering numerous individuals as participants is necessary to obtain sufficient experimental data that is considered valuable and reproducible (Latuszynska & Borawska, 2015) for such testing. This commonly implies considerable costs and effort to provide participants with sufficient motivation to participate and provide serious responses (Kunze, 2006).

Considering this problem, a novel simulation approach called “Large Language Models-driven multi-agent simulation (LLMs-driven MAS)” has recently received the attention of some researchers in economics and behavioral fields. In an LLMs-driven MAS, each agent—called a generative agent—makes decisions using LLMs. This differs from a conventional MAS, where agents follow completely predefined rules. The most recent LLMs-driven MAS, which can generate human-like responses using natural languages, has become increasingly attractive (Park et al., 2023).

Some researchers have recently worked on utilizing LLMs-driven MAS in economic experiments, but these remain challenging. On the one hand, some studies have succeeded in observing the same tendency as the well-known classical results of some economic experiments in LLMs-driven MAS (Aher et al., 2022; Horton, 2023). On the other hand, some studies have reported that the LLMs-driven MAS show both consistent and inconsistent data with real economic experiments (Brookins & DeBacker, 2023; Tsuchihashi, 2023; Guo, 2023; Phelps & Russell, 2023). In addition, some researchers affirm that the results of LLMs-driven MAS for economic experiments depend on the setting of the simulation such as the persona of generative agents and instructions to agents as participants (Horton, 2023; Phelps & Russell, 2023; Han et al., 2023). Thus, the application of LLMs-driven MAS in economic experiments has not been established, and further study on this topic is necessary.

As a first step to establishing a framework for such a new methodology, in Kitadai et al. (2023), we earlier conducted LLMs-driven MAS to explore the reproducibility of the outcome of the economic experiment on the one-shot ultimatum game published by Lin et al. (2020). The ultimatum game is a well-known classical theme of economic experiments that suggests the gap between theoretical equilibrium and actual human behavior (Güth et al., 1982). In the study, we used various prompts and parameter settings to analyze a general proper configuration of LLMs-driven MAS for economic experiments. Consequently, we found a suitable setting to reproduce the distribution of the original data for the proposer side but could not do so for the responder side (Kitadai et al., 2023).

Therefore, this study presents the next step toward establishing a general method for applying LLMs-driven MAS in economic experiments. As explained in detail in the section 2.4, the reasoning ability of the generative agents in our previous simulation was considered insufficient to reproduce the original result on the responder side. Therefore, this study aimed to clarify whether an LLMs-driven MAS can reproduce the results of economic experiments by improving the reasoning abilities of generative agents. Specifically, we conducted LLMs-driven MAS, using an updated version of the framework used by Kitadai et al. (2023), and compared the results of simulations through various methods to improve the reasoning ability of agents to reproduce the outcome of the one-shot ultimatum game.

The contribution of this research will be valuable for utilizing LLMs-driven MAS in economic experiments, even in the future, when LLMs with higher reasoning abilities are widely used. With the rapid development of the computer science, LLMs with higher reasoning abilities are being developed and will prevail. Considering this trend, the level of reasoning ability of LLMs is important for establishing the application of LLMs-driven MAS in economic experiments.

The remainder of this paper comprises the theoretical background presented in Section 2, the research methodology in Section 3, the results shown in Section 4, and the discussion and conclusions presented in Section 5.

2. Theoretical Background

2.1. Ultimatum Game

The ultimatum game involves a sequential decision-making process in which two players determine the distribution of a divisible good. Although there are both one-shot and repeated variations of this game, our investigation specifically concentrated on the one-shot ultimatum game. It comprises the following two steps (Kitadai et al., 2023):

1. The player making the offer suggests a partitioning the divisible the goods. For example, with 100 coins in play, the offering player may propose taking 70 coins for themselves, leaving the responding player with 30 coins.
2. The player responding has the choice to either accept or reject the proposed division. Upon acceptance, the suggested distribution is implemented; however, in the case of rejection, neither players receives anything.

Mathematically, the divisible good is defined as $R \in \mathbb{R}$. The strategies of players are: 1 is $s_1 \in [0,1]$ for

Player 1 and $s_2 \in$ {accept,reject} forPlayer 2. The payoffs of players are as follows:

$$(u_1, u_2) = \begin{cases} (\mathrm{R}(1 - \mathrm{s}_1), \mathrm{Rs}_1) & \text{if } s_2 = \text{accept}, \\ (0,0) & \text{if } s_2 = \text{reject}, \end{cases}$$

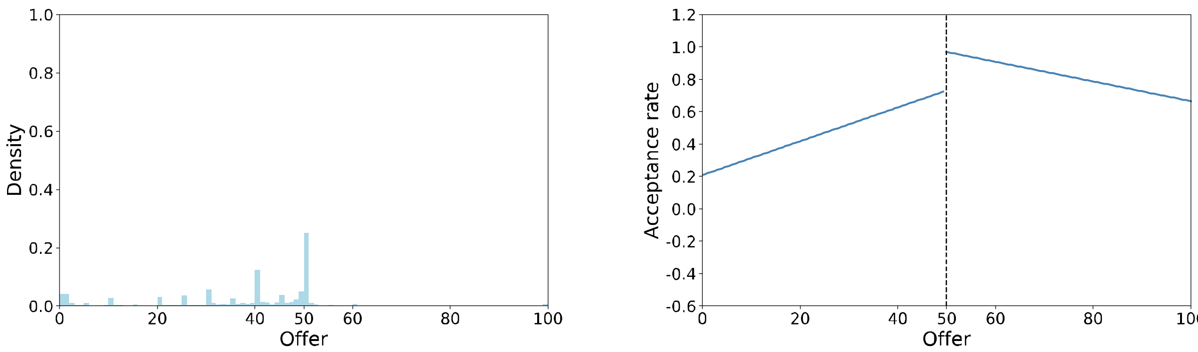
where u_i is the payoff of player $i \in \{1, 2\}$.

Although $s_2 =$ accept is the weakly dominant strategy of Player 2, in economic experiments, it is common to observe decisions of $s_2 =$ reject when the value of s_1 is low (Güth et al., 1982). A commonly proposed explanation for this phenomenon is that responders may choose to sacrifice their payoffs to resist perceived unfairness, particularly when their potential earnings are lower than those of the proposers.

To analyze economic experiments conducted globally, Lin et al. (2020) employed data from MobLab, a well-known educational platform. They focused on double auction and ultimatum games, revealing consistent insights across different cultures and regions. The current study focuses on two outcomes derived from the one-shot ultimatum game, which are different from the theoretical solution—subgame perfect equilibrium—of this game: $(s_1, s_2) = (0,\text{accept})$.

We plot Figure 1(a) based on a sample of 1000 experimental data points for the proposer side published by Lin et al. (2020). The horizontal axis shows the proposer values, and the vertical axis displays their normalized frequency. The distribution peaks at a 50 offer, with few proposals beyond.

Similarly, in Figure 1(b), the horizontal axis represents the proposer values, and the vertical axis indicates the corresponding acceptance rate. The straight line represents the outcomes of the piecewise linear regression, and the shaded area indicates the confidence interval. A notable abrupt increase was evident for the 50 offer. Moreover, there was a discernible upward trend in the acceptance rate of up to 50 as the offer value increased.



(a) Proposer side (b) Responder side

Figure 1: Results of 1000 Sample Data of Economic Experiments for One-shot Ultimatum Game in Lin et al. (2020)

2.2. Generative Pretrained Transformer

Generative Pretrained Transformer (GPT) is a type of LLMs developed by OpenAI®, which was the most famous and prevailing type at the time of study. After the announcement of GPT-3, a type of GPT, by Brown et al. (2020), the dizzying development of the computer science field was accelerated and is still rapidly developing. Currently, GPT is a widely used standard of LLMs in both research and business field. We also used a model of GPT in our study.

Various types of GPT exist. Kitadai et al. (2023) used "gpt-3.5-turbo-0613" which is the available version of GPT-3.5 as of June 23, 2023. Although GPT-4 is known as a higher performance model than GPT-3.5, Kitadai et al. (2023) did not use it because "gpt-4-0613" was more time-consuming in terms of generating a response. However, "gpt-4-1106-preview," which is the available version of GPT-4 as of November 11, 2023, responds quicker for a large sample of simulation and has higher reasoning ability than GPT-3.5.

In addition to these models, diverse prompting techniques are being developed to increase the reasoning ability of LLMs. Owing to the large number of currently available methods, selection is necessary. The following two prompting methods were selected for this study because they are widely known and established as valid methodologies:

- Few-shot prompting (Brown et al., 2020):
  Putting some examples of outputs into the prompt.
- Chain of Thoughts (CoT) (Wei et al., 2022):
  Putting some examples of outputs with their corresponding reason into the prompt.

In comparison to few-shot prompting, usual prompts that include simple instructions without any examples are called zero-shot prompting. To analyze the improvement in reasoning ability, the prompting methods used in this study are, in the reasoning ability order: CoT, few-shot prompting, and zero-shot prompting.

2.3. Literature Review

Numerous researchers have been studying the application of generative agents in social science, including economics (Li et al., 2023; Brand et al., 2023; Bybee, 2023; Lorè & Heydari, 2023; Akata et al., 2023) and psychology (Dillion et al., 2023; Hagendorff, 2023; Harding et al., 2023; Argyle et al., 2023; Binz & Schulz, 2023). Aher et al. (2022) utilized generative agents as participants for economic experiments. They proposed "Turing Experiments (TEs)" as an evaluating method to measure how much LLMs can simulate various human behavioral aspects. TEs comprises four classic social scientific experiments that include the ultimatum game's responder side.

Some researchers have reported that LLMs-driven MAS can reproduce the results of real economic experiments. The most representative is Horton (2023), who selected three classical themes related to the preferences of people and examined the effect of the presence of a minimum wage (Horton, 2023).

By contrast, other scholars have argued that caution should be exercised when replacing human participants with generative agents in such experiments. Brookins and DeBacker (2023) made generative agents play the dictator game and the prisoner's dilemma. In both games, generative agents showed fair decision-making more frequently than did human participants. Guo (2023) focused on the finitely repeated ultimatum and prisoner's dilemma games, finding that generative agents can partly exhibit behavior consistent with human behavior by using well-crafted prompts (Guo, 2023).

Phelps and Russell (2023) also simulated a finitely repeated prisoner's dilemma, using four types of bots as opponent players of the generative agents: always cooperating bot, always defecting bot, and bots who cooperate/defect in the first round and then imitate the choice of the generative agent in the previous round. They found that generative agents could not behave appropriately in response to the different strategies of the opponent. Tsuchihashi (2023) simulated a first-price auction (FPA) and second-price auction (SPA) using generative agents. Although human participants tend to bid higher than their willingness to pay in both the FPA and SPA, the study found that generative agents did not show such trends (Tsuchihashi, 2023).

Considering this background, the application of LLMs-driven MAS in economic experiments has not been established, and further study is necessary on this topic. Kitadai et al. (2023) made the first contribution toward the establishment of a methodology to reproduce the results of economic experiments. Thus, this study is the next step in exploring the influence of the reasoning ability of generative agents to replicate the real economic experiments.

2.4. Our Previous Study

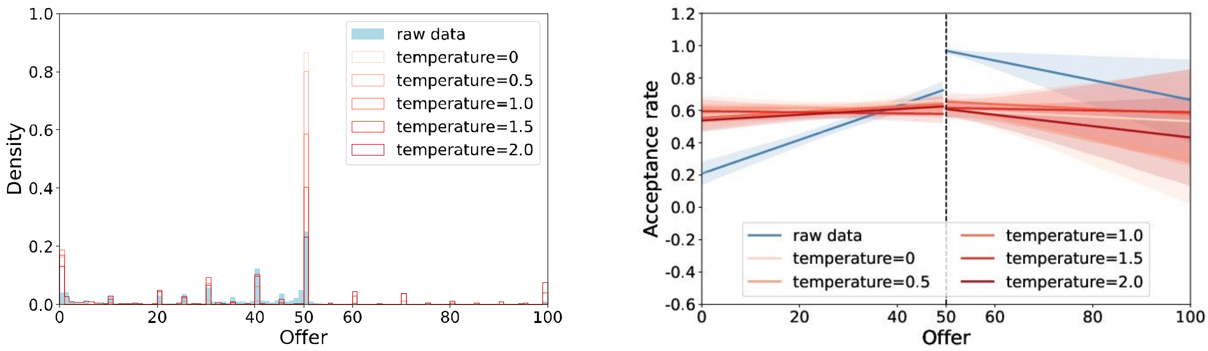
In our previous study, Kitadai et al. (2023), we attempted to reproduce the result of the ultimatum game of Lin et al. (2020) by LLMs-driven MAS using GPT focusing on the effects of the following three aspects:

- How the amount of money that the subjects will receive at the end of the experiment is determined.
- Whether explicit instructions aimed at maximizing rewards are provided.

- The setting of the parameter "temperature" in GPT.

Herein, "temperature" is a parameter in GPT that influences the randomness and creativity of its output, ranging from 0 to 2. Lower temperatures lead to more predictable, conservative outputs by favoring higher probability words, and higher temperatures increase creativity and variability, encouraging the model to choose less likely words, which can result in more novel and diverse responses but also increase the likelihood of irrelevant or nonsensical content.

A set of 1000 simulations was performed based on $4 \times 2 \times 5 = 40$ variations at each of the points listed above. Consequently, for the proposer side, it was found that not giving explicit instructions to maximize profits, specifying the amount to be received by the participants at the end of the experiment, and a higher value of "temperature" made the simulation result closer to the original result in Lin et al. (2020). Simultaneously, however, any simulation settings reproduced the original result for the responder side. Figure 2 shows the results obtained when no instruction to maximize profits was given, indicating that each coin obtained in the game would be redeemed at 100 dollars (Kitadai et al., 2023). In the figure, the data from MobLab used by Lin et al. (2020) are plotted in blue, whereas the simulation results for the five temperature settings (0, 0.5, 1.0, 1.5, 2.0) are depicted in different colors.



(a) Proposer side (b) Responder side

Figure 2: A Result of Simulations in Kitadai et al. (2023)

Figure 2(a) presents the results for the proposer side; the horizontal axis corresponds to the offer value and the vertical axis indicates the density. A peak appeared at the offered value of 50, which is consistent with the actual experimental results. Figure 2(b) presents the results for the responder side; the offer value is indicated on the horizontal axis and the acceptance rate is on the vertical axis. In this case, the responses of generative agents are almost insensitive to variations in the proposed values. In particular, low acceptance rates of offers greater than half are not expected and tend to be irrational. Therefore, we considered that the reasoning ability of the agents was not sufficiently high to achieve consistency with the actual economic experiment.

Regarding the setting of persona, Kitadai et al. (2023) generated the information of gender, age, and personality at random by using "gpt-3.5-turbo-0613". However, such random generation of personas can act as noise in revealing the effect of the reasoning ability of generative agents because the result of LLMs-driven MAS for economic experiments depends on the persona configuration (Horton, 2023; Phelps & Russell, 2023; Han et al., 2023). Therefore, the step of the persona setting for each agent in the simulation of the current research was eliminated.

3. Methodology

The prompt of simulations conducted in this study followed the framework below, which is an updated version of that used by Kitadai et al. (2023).

1. Explanation of the ultimatum game.
2. Description of the situation in which the agent makes a decision and instructions for decision-making. Each prompting method (Few-shot, CoT, or Zero-shot) is introduced here.

3. Specification of the output format.

All the prompts used in this study were written in English. This choice was made because of known reductions in LLM performance when prompts are written in languages other than English (Etxaniz et al., 2023). Moreover, the descriptions for the ultimatum game and the scenario for decision-making were the same as in our previous study.

Regarding the second item in the framework, based on the insights obtained by Kitadai et al. (2023), in the prompt, agents were tasked with deciding on the distribution of a non-real-world currency of 100 coins. It was further explained that each coin would be redeemed as 100 dollars at the end of the game, however, the instruction to maximize their profit was not mentioned; these are the settings which resulted in the most similar result with the actual experimental result of the ultimatum game in our previous study.

Using this prompt structure, we conducted the four simulation patterns described in Table 1. In Pattern A, which has the same combination of settings as used by Kitadai et al. (2023), we introduced the older GPT model, gpt-3.5-turbo-0613, as a benchmark to compare the reasoning abilities of the generative agents. Regarding the temperature parameter, we considered five variations—$0, 0.5, 1.0, 1.5, 2.0$—as used by Kitadai et al. (2023) for Patterns A, B, and C. However, for Pattern D, we did not consider the value 2.0. This is because, even with the latest model, only one response per hour was obtained by using CoT with a temperature of 2.0, making it an excessively time-consuming simulation setting as an alternative to economic experiments.

Table 1. Four Patterns of Simulation Settings

Pattern	GPT Model	Prompting Method	Temperature Range
A	gpt-3.5-turbo-0613	Zero-shot	$0, 0.5, 1.0, 1.5, 2.0$
B	gpt-4-1106-preview	Zero-shot	$0, 0.5, 1.0, 1.5, 2.0$
C	gpt-4-1106-preview	Few-shot	$0, 0.5, 1.0, 1.5, 2.0$
D	gpt-4-1106-preview	CoT	$0, 0.5, 1.0, 1.5$

To use few-shot and CoT promptings, some examples of prompts must be prepared. For few-shot prompting, we used 10 data points selected from the raw data of Lin et al. (2020). For CoT, we made a reason and added it to each sample dataset used in few-shot prompting, and placed them into the prompt. For each combination of prompting method, temperature, and proposer or responder side, 1000 agents were generated and made independent decisions.

4. Results

4.1. Proposer Side

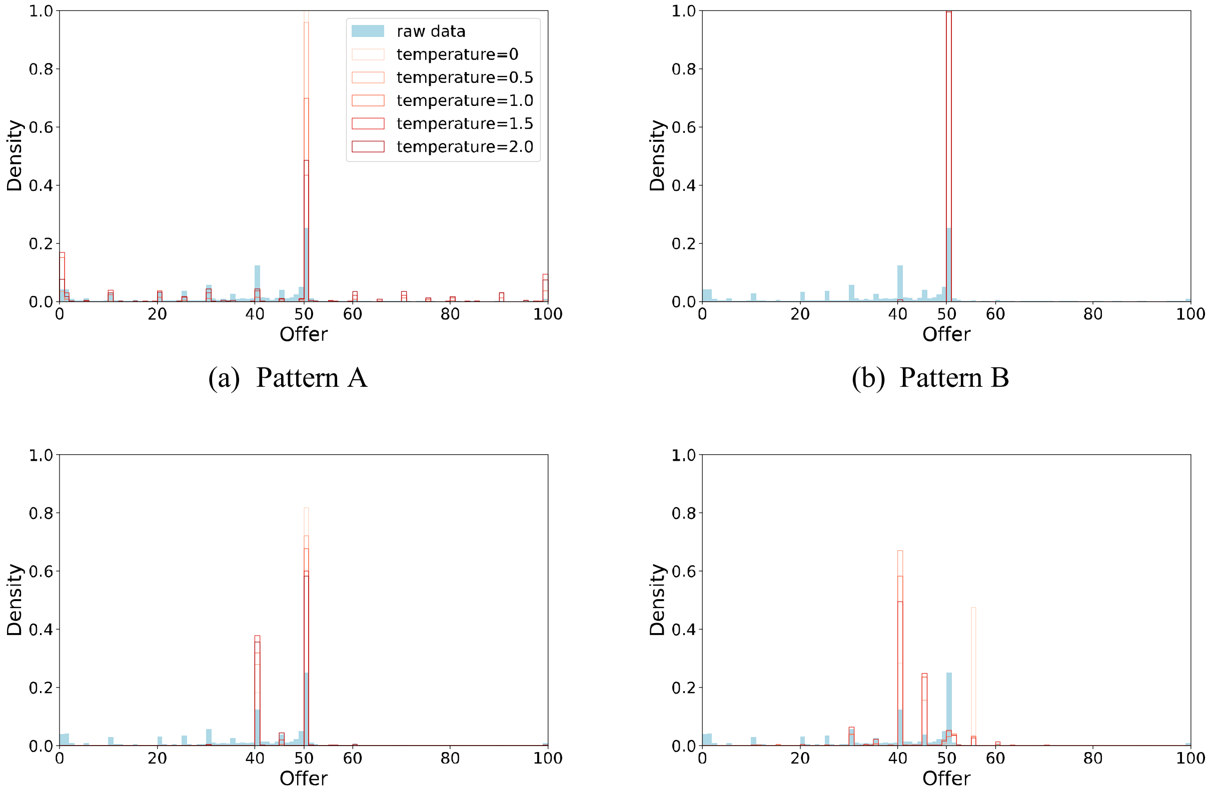
Figure 3 presents the results of the proposing player obtained from the simulations. The horizontal axis represents the offered value, and the vertical axis represents the normalized count of agents offering each value. The blue histogram shows the data from MobLab used by Lin et al. (2020) and remains consistent across all graphs. The simulation results for the five (or four) temperature settings are depicted using histograms outlined in different colors. This figure shows four graphs, each one for each pattern setting.

Figure 3(a) shows a similar but a slightly different tendency from that of our previous study (see Figure 2(a)). In the case of $\mathrm{temperature} = 0\ to\ 1.5$, the peak appears at an offer value of 50, and the higher the temperature setting, the closer the distribution is to the raw data. However, when $\mathrm{temperature} = 2.0$, the distribution did not become closer, showing a higher peak at an offered value of 50, compared with $\mathrm{temperature} = 1.5$. This implies a larger gap between the raw data and the previous one in Figure 2(a).

Figure 3(b) shows that at any temperature settings, almost all agents offered $50 - 50$ division of coins.

This is the most frequent offer in the raw data; however, the disappearance of other offers in the simulation is different from the real economic experiment.

Other tendencies could be identified by introducing other prompting methods. When utilizing few-shot prompting, Figure 3(c) shows the most frequent peak at an offer value of 50 and the second most frequent peak at 40, whereas there is almost no other offer value. When adapting the CoT prompting, in Figure 3(d), the peak at an offer value of 50 disappears. Moreover, in the case of $\text{temperature} = 0.5–1.5$, the most frequent peak is at an offer value 40, and the second most frequent peak appears at 45. However, in the case $\text{temperature} = 0$, the most frequent peak appears at 55, and the second and third most frequent peaks are at 40 and 45.



(a) Pattern A (b) Pattern B

(c) Pattern C (d) Pattern D

Figure 3: Proposer-side Simulation Results

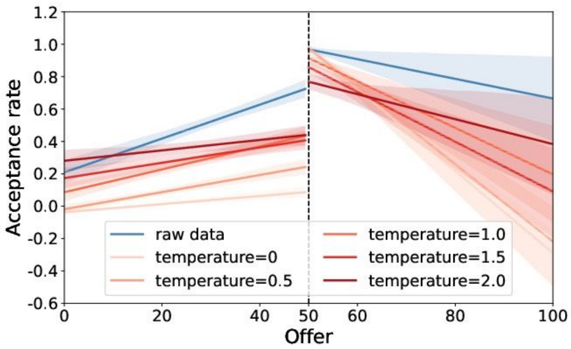
4.2. Responder Side

Figure 4 presents the responding player results of the simulations, with one graph for each settings pattern. The horizontal axis of each graph indicates the value offered by the proposer, and the vertical axis denotes the acceptance rate for each offer. The blue line represents data from 1000 samples out of those used by Lin et al. (2020) and remains consistent across all figures. The simulation results for five (or four) temperature settings are depicted using distinct colored graphs.

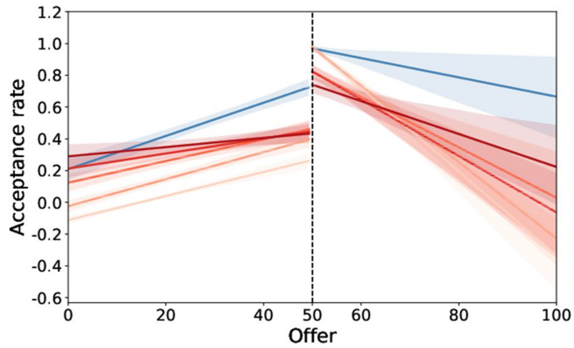
Figures 4(a) and 4(b) show similar tendencies. In all temperature patterns, the higher the offered value from 0 to 50, the higher the acceptance rate. A discontinuous jump appears at 50, and the higher the offered value from 50 to 100, the lower the acceptance rate. Although this tendency is consistent with the original results, some differences were observed. On the left side of these figures, the increase in the acceptance ratio is smaller than that of the blue line. At the offer value of 50, the acceptance rate of the left blue line was approximately 0.7, and the rate of the other colored lines was approximately 0.4 or less. On the right side of these figures, the acceptance rate of the simulations was lower than that of the raw data. In particular, at the offer value of 100, the rates were lower than those at 50. This implies a discrepancy in the experimental results.

Figures 4(c) and 4(d) show similar tendencies but differ from the other graphs. In each case, there were slight differences among the simulation settings for various temperature values. On the left side of both graphs, the simulation lines show a larger increase in the acceptance rate against the value of the offer than that of the raw data line. On the right side, the simulation lines are horizontal, and the acceptance rate is equal to 1.

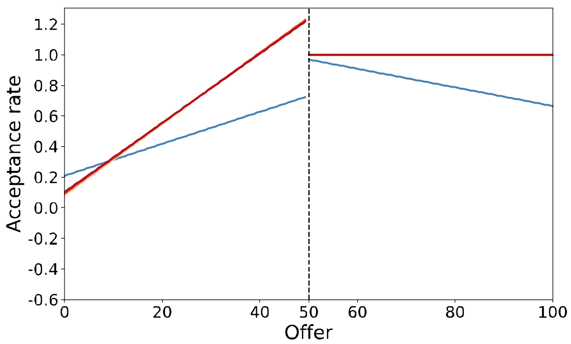
The reason these lines were depicted can be understood by Figure 5. This figure presents bubble plots derived from responder-side simulations. The axes and color schemes are consistent with those depicted in Figure 4. The center of each bubble represents the acceptance rate corresponding to a specific offer value, while the bubble size reflects the quantity of data points associated with that offer value. These figures reveal that the distributions of data points are skewed. This skewness acounts for the anomalies observed in the regression lines of Figure 4, where acceptance rates exhibited extreme values, either exceeding 1.0 or falling below 0.



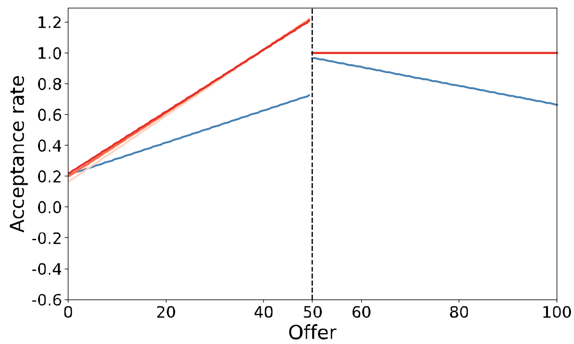
(a) Pattern A



(b) Pattern B

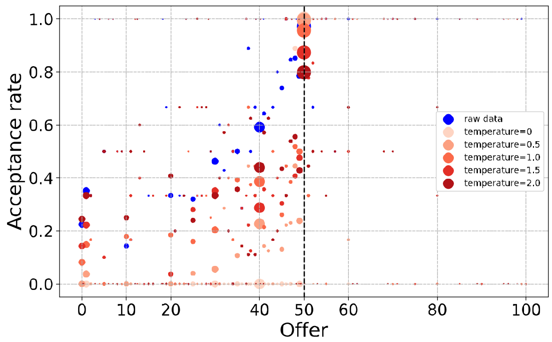


(c) Pattern C

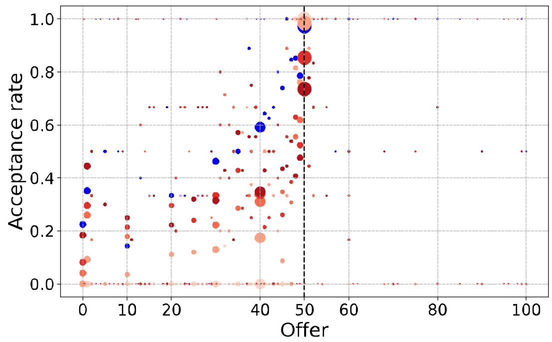


(d) Pattern D

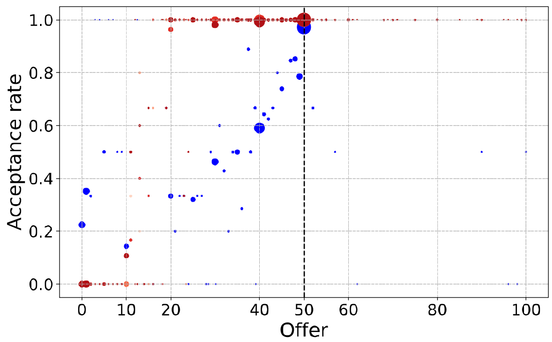
Figure 4: Responder-side Simulation Results



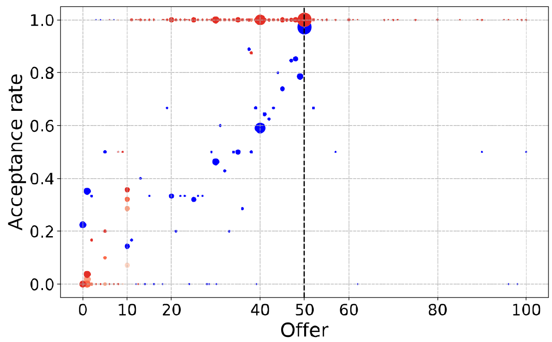
(a) Pattern A



(b) Pattern B



(c) Pattern C



(d) Pattern D

Figure 5: Bubble Plots of Responder-side Simulation Results

5. Discussion and Conclusion

5.1. Effects of Improving Reasoning Ability

In this study, we conducted simulations using four patterns as variations of the GPT model and prompting method. Regarding the GPT model, we used two types of models, gpt-4-1106-preview and gpt-3.5-turbo-0613, being known that gpt-4-1106-preview is superior to gpt-3.5-turbo-0613 in reasoning ability. For the prompting method, in decreasing order of reasoning ability, the three methods used were CoT, few-shot prompting, and zero-shot prompting. Thus, the four patterns in descending order of the reasoning ability of the generative agents are as follows: D, C, B, and A.

Considering this, it can be said that the higher the reasoning ability of the generative agents, the closer their decision-making becomes to the theoretical solution. In the theoretical solution of the ultimatum game, the subgame perfect equilibrium is $(s_1, s_2) = (0, \text{accept})$. Regarding the proposer side, from Figures 3(a) to 3(d), the distribution changed toward the offer 0. On the responder side, by utilizing prompting methods to improve the reasoning ability of the generative agents, the fluctuation of the acceptance rate in response to an offer value became stable, and the generative agents almost always decided $s_2 = \text{accept}$ when the offer value was larger than 20 and $s_2 = \text{reject}$ when the offered value was equal to or less than 10. Therefore, by improving the reasoning ability of generative agents, the result of LLMs-driven MAS does not align closely with the actual experimental result; rather, it aligns closely with the theoretical solution.

The motivation for improving the reasoning abilities of generative agents stemmed from insights indicating that those in our previous study were not sufficient to correctly understand the ultimatum game, which resulted in outcomes differing from those of the actual experiments. However, contrary to our expectations, such improvements brought the simulation results closer not to the experimental outcomes but to the theoretical solution. This finding is significant because it suggests that factors other than enhanced reasoning abilities need to be incorporated into generative agents to emulate human-like decision-making.

5.2. Toward the Replication of Human-like Decision Making

In this study, we attempted to reproduce the results of economic experiments found by Lin et al. (2020) through an LLMs-driven MAS, where the reasoning ability of generative agents was improved. Consequently, we found that the higher reasoning ability of generative agents does not indicate a result closer to the experimental result; rather, the simulation result becomes closer to the theoretical result. In the future, the reasoning ability of LLMs will further improve following the rapid development of computer science. Considering this expectation, this study suggests that further research is required to obtain intelligent generative agents for making human-like decisions.

Moreover, a comparison of this study with our previous research suggests the importance of persona settings. We attribute the difference between Figures 2(a) and 2(b) (the results of our previous study) and Figures 3(a) and 4(a) (the results of the simulation in this study) to the existence of persona settings. This is because the other settings of the simulations were the same. Comparing these results, the disappearance of the persona setting in this study worked negatively on the proposer side and positively on the responder side to reproduce the results of actual decision-making in Lin et al. (2020). This difference suggests that the settings of the personas of generative agents are essential for applying LLMs-driven MAS into economic experiments. This is consistent with previous research (Horton, 2023; Phelps & Russell, 2023; Han et al., 2023) and requires further investigation.

Therefore, how to set the personas of the agents is key to obtain generative agents that make human-like decisions. In this study, expanding on the work by Kitadai et al. (2023), we eliminated the step of the persona settings for each agent in the simulation, which generated a discontinuous jump at the offer value of 50 in Figure 3(a). However, when the reasoning ability of the generative agents improved, such an elimination caused a larger difference compared with raw data in the results for the proposer side. By setting an appropriate persona for agents, their decision-making tendency may align with that of actual people. Specifically, now we are planning to incorporate some measures of behavioral economics such as inequality aversion, risk aversion, endowment effect and so on into persona. Exploring the general framework to set such persona of generative agents is a future step toward establishing a general framework of LLMs-driven MAS to reproduce the results of economic experiments.

References

Aher, G., Arriaga, R. I., & Kalai, A. T. (2022). Using Large Language Models to simulate multiple humans and replicate human subject studies. Proceedings of the 40th International Conference on Machine Learning. 202:337-371. https://proceedings.mlr.press/v202/aher23a.html.

Akata, E., Schulz, L., Coda-Forno, J., Oh, S. J., Bethge, M., & Schulz, E. (2023). Playing repeated games with Large Language Models. arXiv:2305.16867.

Argyle, L. P., Busby, E. C., Fulda, N., Gubler, J. R., Rytting, C., & Wingate, D. (2023). Out of one, many: Using language models to simulate human samples. *Political Analysis, 31(3),* 337-351.

Binz, M., & Schulz, E. (2023). Using cognitive psychology to understand GPT-3. Proceedings of the National Academy of Sciences, 120(6), e2218523120.

Brand, J., Israeli, A., & Ngwe, D. (2023). Using GPT for market research. Harvard Business School Marketing Unit Working Paper No. 23-062, Available at SSRN: http://dx.doi.org/10.2139/ssrn.4395751.

Brookins, P., & DeBacker, J. M. (2023). Playing games with GPT: What can we learn about a large language model from canonical strategic games. Available at SSRN 4493398. http://dx.doi.org/10.2139/ssrn.4493398.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems, 33,* 1877-1901.

https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

Bybee, L. (2023). Surveying generative AI's economic expectations. arXiv:2305.02823.

Dillion, D., Tandon, N., Gu, Y., & Gray, K. (2023). Can AI language models replace human participants?. *Trends in Cognitive Sciences, 27(7),* 597-600. https://doi.org/10.1016/j.tics.2023.04.008.

Etxaniz, J., Azkune, G., Soroa, A., Lopez de Lacalle, O., & Artetxe, M. (2023). Do multilingual language models think better in English?. arXiv:2308.01223.

Guo, F. (2023). GPT agents in game theory experiments. arXiv:2305.05516.

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization, 3(4),* 367-388. https://doi.org/10.1016/0167-2681(82)90011-7.

Hagendorff, T. (2023). Machine psychology: Investigating emergent capabilities and behavior in large language models using psychological methods. arXiv:2303.13988.

Han, X., Wu, Z., & Xiao, C. (2023). “Guinea Pig Trials” Utilizing GPT: Anovel smart agent-based modeling approach for studying firm competition and collusion. arXiv:2308.10974.

Harding, J., D’Alessandro, W., Laskowski, N. G., & Long, R. (2023). AI language models cannot replace human research participants. *AI & Society, 1,* 1-3. https://doi.org/10.1007/s00146-023-01725-x.

Horton, J. J. (2023). Large language models as simulated economic agents: What can we learn from homo silicus? National Bureau of Economic Research, Working Paper No. w31122. https://doi.org/10.3386/w31122.

Kitadai, A., Tsurusaki, Y., Fukasawa, Y., & Nishino, N. (2023). Toward a Novel Methodology in Economic Experiments: Simulation of the Ultimatum Game with Large Language Models. In 2023 IEEE International Conference on Big Data (Big Data) (In Press).

Kunze, J. (2006). Design Problems of Complex Economic Experiments. In Spiliopoulou, M., Kruse, R., Borgelt, C., Nürnberger, A., & Gaul, W. (Eds.), From Data and Information Analysis to Knowledge Engineering. *Studies in Classification, Data Analysis, and Knowledge Organization* (pp. 494-501). Springer

Latuszynska, M., & Borawska, A. (2015). Economic experiments in virtual world. *Transformations in Business & Economics, 14(3C),* 495-506.

Li, N., Gao, C., Li, Y., & Liao, Q. (2023). Large language model-empowered agents for simulating macroeconomic activities. arXiv:2310.10436.

Lin, P. H., Brown, A. L., Imai, T., Wang, J. T., Wang, S. W., & Camerer, C. F. (2020). Evidence of general economic principles of bargaining and trade from 2,000 classroom experiments. *Nature Human Behaviour, 4,* 917-927. https://doi.org/10.1038/s41562-020-0916-8

Lorè, N., & Heydari, B. (2023). Strategic Behavior of Large Language Models: Game Structure vs. Contextual Framing. SSRN 4569717. Retrieved from https://ssrn.com/abstract=4569717

Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. In Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (pp. 1-22). https://doi.org/10.1145/3586183.3606763.

Phelps, S., & Russell, Y. I. (2023). Investigating emergent goal-like behaviour in large language models using experimental economics. arXiv:2305.07970.

Tsuchihashi, T. (2023). How much do you bid? Answers from ChatGPT in first-price and second-price auctions. *Journal of Digital Life, 3*. https://doi.org/10.51015/jdl.2023.3.11.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems, 35,* 24824-24837.

https://proceedings.neurips.cc/paper_files/paper/2022/file/9d5609613524ecf4f15af0f7b31abca4-Paper-Conference.pdf

joint conference
group and negotiation DECISION support systems technology
Porto 2024

24th International Conference on Group Decision and Negotiation &
10th International Conference on Decision Support System Technology
Human-Centric Decision and Negotiation Support for Societal Transitions

Reflection Machines and Adaptive Algorithmic Decision Support in Clinical Diagnostic Applications

Sabine T. Koeszegi[1*], Atefa Ahmadi[2], Josef Kastner[1] & Harald Kittler[2]
[1] TU Wien, Vienna, Austria
[2] MED Uni Wien, Vienna, Austria
[*] sabine.koeszegi@tuwien.ac.at

Abstract

The integration of Artificial Intelligence (AI) in clinical settings has raised high expectations to improve the quality of clinical diagnosis for various diseases. However, the effectiveness of AI integration depends on several factors, including user self-assessment, expectations of AI, and technology acceptance, and therefore requires careful design.
To overcome identified problems with existing algorithmic decision support (ADS), we propose the introduction of adaptive AI support and "reflection machines" that facilitate interactive dialogues between users and ADS. The proposed specific design allows for the preservation of human expertise, as reliance on AI alone precludes the acquisition of meaningful experience. By acting as a closed-loop expert system that only intervenes in work processes when noticeable deviations from its own predictions occur, this approach ensures that ADS complements rather than replaces human expertise in healthcare.
The primary aim of this study is to empirically assess the impact of adaptive algorithmic decision support (AADS) coupled with reflection machines on the quality of diagnostic and management decisions in the detection of skin cancer, one of the most common cancers worldwide with high mortality rates when detected late. In controlled laboratory experiments, participants will be assisted by different ADS designs to test the effectiveness of proposed adaptive AI and reflectance machines. We will compare the performance of subjects with varying levels of medical and dermatological expertise in analysing 1000 retrospectively collected images of pigmented skin lesions.

Keywords: decision support; artificial intelligence; skin cancer; dermatology; reflection machine

1. Introduction

In this project, we aim to investigate in detail how ADS and AI-based prediction models using dermoscopy images for skin cancer detection will affect the quality of diagnosis and management decisions.

Skin cancer, including both melanoma and non-melanoma skin cancers, is one of the most common malignant tumours in the world, accounting for 5% of all new cases of cancer each year (Cancer Stat Facts: Melanoma of the Skin, 2023). Melanoma is associated with one of the poorest prognoses of all cancers when metastases are present (Davis et al, 2019), but if detected early, it usually has a favourable outcome.

Skin cancer is usually diagnosed by experienced, board-certified dermatologists who are skilled in dermoscopy, a non-invasive diagnostic technique that requires specific training and expertise. The definitive diagnosis of skin cancer is usually made by biopsy or excision of the entire lesion, followed by detailed histopathological analysis. However, the limited number of experts and resources in this field necessitates the use of AI as a supporting tool to facilitate the diagnosis of skin cancer (Haenssle et al., 2018). Dermatology is a particularly attractive area for the implementation of AI systems, given its reliance on visual diagnosis.

AI has demonstrated expert-level performance in skin cancer detection: Codella et al. found that their deep learning algorithms outperformed dermatologists with 76% sensitivity and 62% specificity, compared to

70.5% and 59% for dermatologists (Codella et al., 2017). A comprehensive study involving over 500 participants (Tschandl et al., 2019) demonstrated that artificial intelligence (AI) systems outperformed both experienced experts and novices in the diagnosis of skin cancer. The study highlighted the ability of AI to accurately identify different forms of skin cancer, highlighting its potential as a valuable tool in dermatological diagnostics (Tschandl et al., 2019). Overall, human-AI collaboration has been shown to provide better outcomes than either human alone or AI-only decision making in cancer diagnosis (Wang et al., 2016), (Tschandl et al., 2019), as both offer unique capabilities (Krüger et al., 2017).

There are concerns, however, that these AI systems run the risk of performing poorly in realistic clinical scenarios or when used outside their intended scope. In addition, researchers have also expressed reservations about the widespread problem of inadequate reporting in deep learning research in high-risk applications (see, for example, Burger et al., 2022).

In collaborative decision making (AI-assisted decision making), the assessment of decision quality becomes crucial, goes beyond mere precision and accuracy criteria, and requires to incorporate social-psychological aspects into AI design. For instance, the principle of oversight emphasises that humans should play a corrective role, especially in critical decisions such as those in health care, and that reliance on human judgement is both necessary and sufficient to prevent poor decisions. Therefore, principled decisions about the use of ADSs and their implementation contexts should only be made after ensuring that the relevant factors have been properly considered. Relying solely on the ability and willingness of experts to take responsibility for ADS-based decisions is therefore insufficient.

Therefore, the full realisation of the positive potential of ADS in work processes requires a detailed socio-technical systems analysis and design (Zafari, S., Köszegi, S.T. and Filzmoser, M., 2021). In this study, we follow this approach and propose an "AI in the loop" design, rather than the human-in-the-loop model of collaborative decision making, where the AI system takes the lead and assigns supervisory tasks to humans. In the following sections, we briefly summarize related work (section 2), propose a specific design for adaptive ADS (section 3), describe the planned experiment design (section 4), and briefly summarize expected findings (section 5) and conclusions (section 6).

2. Related Work

Previous research has shown that user expectations, biases, and the timing of decision support influence the effectiveness of ADS.

Human decision makers develop expectations about the capabilities and features of ADS that are influenced by personal exposure and second-hand experience (Burton et al., 2019). Zhang et al. (2021) find that expectations extend beyond instrumental capabilities as they also include human-like behaviour and shared understanding. False expectations can lead to automation bias or algorithm aversion (Burton et al., 2019). Automation bias, where people trust even incorrect ADS output, leads to complacency and dependence on ADS (Parasuraman & Manzey, 2010; Bahner et al., 2008). Conversely, the Dunning-Kruger effect causes individuals to overestimate their competence compared to algorithmic support, leading to under-reliance on AI systems (He et al., 2023). Lay people who lack domain knowledge often trust algorithms over human expertise, while experts are less likely to rely on algorithmic predictions (Logg et al., 2019; Wouters et al., 2019). Algorithmic education has been proposed to build trust by creating reasonable expectations of AI capabilities, emphasising exposure to errors in automation during training (Bahner et al., 2008). The severity of decisions influences trust in algorithms, with greater consequences leading to reluctance (Filiz et al., 2023).

Furthermore, the timing of decision support affects user acceptance, with users feeling diminished if they are perceived as mere recipients of machine decisions (Lange et al., 2021). Skills may deteriorate with infrequent use of ADS, leading to deskilling processes (Bainbridge, 1983). The introduction of ADS may change the skills required by users (Smith et al., 2010). ADS may affect human knowledge, with Fügener et al. (2021) suggesting that collaborative decision making with AI leads to the loss of unique human knowledge, hindering the 'wisdom of crowds'.

Other studies have addressed long-term negative effects of ADS on self-evaluation, self-efficacy and human knowledge hypothesizing that they may offset these benefits, resulting in overall poorer socio-technical decision performance compared to human decisions alone. Users' perceptions of the potential role of ADS are intertwined with their self-evaluation of their competence, as shown, for example, in an experimental study by Jacobsen, Johansen and Bysted (Jacobsen et al., 2020). In a collaborative recycling sorting task, non-experts were given item classifications and confidence ratings by an AI system. At the same time, participants with ADS performed better overall; people who were assisted perceived themselves as less effective than they were (underconfident). In contrast, the opposite was true for people with AI support (overconfident). A qualitative analysis of participants' post-experimental perceptions in the same study (Papachristos et al., 2021) identified four distinct roles as projections of what users expected from interacting with the ADS: People who self-assessed relatively high competence in waste sorting (which is a non-trivial task) expected the system to confirm their decision and would ignore a system's opposite suggestion unless they were very uncertain. Here, ADSs were assigned a mirror role (inspired by the fairy tale of Snow White) or an assistant role. In both cases, the decision was made by the human. Participants who rated themselves as modest to low trusted the AI system to make a better judgement.

Delegating tasks to AI systems can have a long-term negative impact on human skills and know-how. Meaningful experience is no longer gained, as essential skills and abilities are only regained when needed and trained (see Bainbridge 1986). In addition, in the long run - without immediate feedback on the quality of a decision - a flawed self-assessment can result, either in a bias towards automation or in complacency. Again, in the long run, these dynamics will negate the potential benefits of AI decision support (ADS).

Transparency is another crucial element in good decision support: with explanations of how and why decisions are made, fostering understanding, trust and accountability (GDPR, 2016; Papagni et al., 2022, 2021a, 2021b; Felzmann et al., 2019; de Graaf et al., 2018; Miller, 2019). In addition, greater system transparency allows users to better understand the performance of the system and the processes that lead the system to make a particular decision or prediction (Felzmann et al., 2019). Closing the knowledge gap through system transparency increases technology acceptance (Malle et al., 2007). The plausibility of explanations, rather than accuracy, is key to promoting user understanding and accountability (Papagni et al., 2022, 2021a, 2021b).

Finally, the transfer of decision-making to algorithmic decision systems (ADS) also raises ethical and organisational challenges (Simmons, 2018). Apart from EU regulations such as Article 22 of the GDPR, which prohibits fully automated decisions for natural persons and emphasises the right to explanation (GDPR, 2016), ADS that replace humans in critical decisions face legitimacy issues (Simmons, 2018). Furthermore, autonomous AI systems challenge traditional accountability practices (Lei & Rau, 2021). Accountability dysfunctions arise in ADS-assisted decisions, with experts relying on machine outputs without understanding the process, leading to blame for false accusations (Smith et al., 2010, Kim & Hinds, 2006; Furlough et al., 2021). Therefore, different degrees of human supervision and control, represented by 'human-in-the-loop', 'human-on-the-loop' and 'human-in-command', have been proposed for different decision contexts. Particularly high uncertainty and ethical concerns necessitate human involvement, challenging the appropriateness of the latter approach “human in command” (Coeckelbergh, 2019; Zafari & Koeszegi, 2018; Ivanov, 2023).

3. Reflection Machines and Adaptive Algorithmic Decision Support

Therefore, it is crucial to adapt the system support to the user characteristics in order to optimise the decision quality of a collaborative decision process within the socio-technical system. We propose a system design that learns user preferences and expectations in advance and adapts its recommendations to specific users.

Furthermore, we propose design ADS in form of "reflection machines" (Haselager et al., 2023) as they provide a meaningful overview of human problems and control through a specific type of decision support. Rather than making suggestions to decision makers, a reflection machine challenges their reasoning and decisions. For example, it may highlight facts, raise critical issues, or present counter-arguments that are inconsistent with the proposed decision, thereby improving the problem-solving process and decision quality.

This proposal reverses the idea of "human in the loop" to an AI in the Loop (AILO) principle, using ADS as a monitoring system that only intervenes in work processes when it detects significant deviations from its own predictions. In this way, users will feel in control of their decisions and more comfortable working with autonomous AI systems. The proposed project aims to address this issue.

This form of support also enhances the individual's problem-solving skills and counteracts uncritical reliance on ADS. Experimental studies, particularly in the medical field, have already shown positive experiences with this approach (Haselager et al., 2023).

4. Methods and study design

In this study we aim to include 70 study participants with various level of expertise and experience in the diagnosis of skin cancer and in dermatoscopy. We will use a dataset of 1000 retrospectively collected dermatoscopic images of pigmented skin lesions. Pre-study surveys will be conducted on participants' demographics, competence levels, profession, age, and sex In each test round participants are required to answer ten questions with the assistance of an AI system with and without user adaptations. The sequence in which the AI systems are used is randomized. Participants are tasked with two primary objectives: firstly, to correctly diagnose the pigmented skin lesions, and secondly, to determine the appropriate treatment management. For each image, there are seven different diagnosis options and four different therapy options. The images are not repeated in the tests and the study participants will answer new questions each time. The images will be randomly selected with block randomization ensuring equal difficulty and composition of the images for each mode of AI-support. The participants will be exposed to each AI support mode only once.

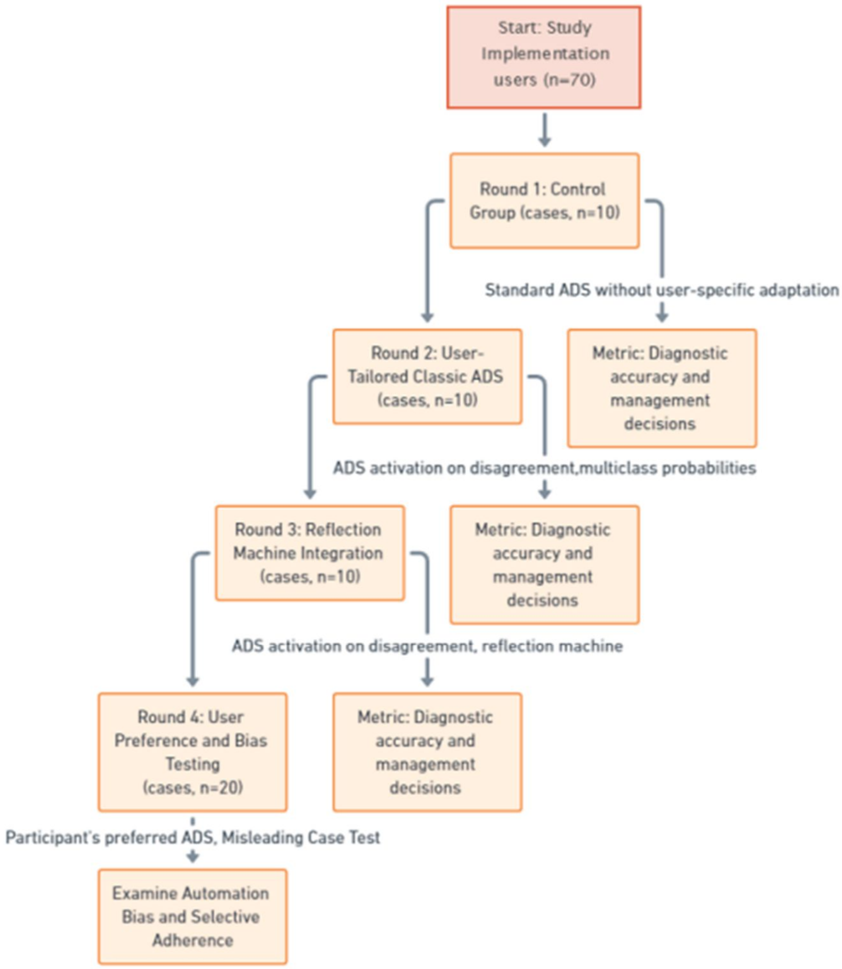


Figure 1: Study Flowchart

In our study, we will implement four rounds (Figure 1) of testing to assess different levels of Automated Decision Support (ADS) and its impact on user diagnosis accuracy.

1. Control Group: The first round employs standard ADS without user-specific adaptation. This will establish a baseline for comparison.

2. User-Tailored Classic ADS: In the second round, we introduce classic ADS with multiclass probabilities, now customized to each user. Here, ADS intervention occurs only if there is a discrepancy between the user's diagnosis and the ADS suggestion.

3. Reflection Machine Integration: The third round retains the condition of ADS activation upon disagreement with the user's diagnosis. However, instead of classic multiclass probabilities, we utilize a 'reflection machine'. This interface presents the user with three images similar to the top diagnosis suggested by the system, prompting the user to reconsider their decision.

4. User Preference and Bias Testing: In the final round, participants choose their preferred ADS support, which is then provided at their discretion. All users will evaluate the same randomly selected 20 images. Additionally, this round includes a specifically designed misleading case to Aims of the study 8 examine potential issues of "automation bias" and "selective adherence" in both modes of ADS support.

A post-study survey will evaluate the accuracy of diagnostic decisions, confidence levels, user experience and system interaction.

5. Expected Results

The objectives of this study are to test the impact of different ADS designs (user adaptation and reflection machines) on the quality of skin cancer diagnosis and management decisions. We expect answers to the following research questions:

1. Is there a significant difference in diagnostic accuracy among the different modes of AI-support with and without user adaptation?
2. Will management decision improve with adaptive AI?
3. Is there a difference in user experience and satisfaction between AI systems with and without user adaptations?

6. Conclusions

The use of AI in the diagnosis of skin cancer is expected to play a major role in the future. This makes it all the more important to develop systems that, overall, increase the decision quality in diagnosis and management decisions. With the proposed adaptation of standard AI support in skin cancer detection and the suggested experiment design, we not only contribute to a better understanding of algorithmic decision support but also make design propositions with high practical relevance.

List of Figures

References

Burton, J. W., Stein, M., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making*, *33*(2), 220–239. https://doi.org/10.1002/bdm.2155

Cancer Stat Facts: Melanoma of the Skin. (2023, November 12). National Cancer Institute. https://seer.cancer.gov/statfacts/html/melan.html

Codella, N. C. F., Nguyen, Q.-B., Pankanti, S., Gutman, D. A., Helba, B., Halpern, A. C., & Smith, J. R. (2017). Deep learning ensembles for melanoma recognition in dermoscopy images. *IBM Journal of Research and Development*, *61*(4/5), 5:1-5:15. https://doi.org/10.1147/JRD.2017.2708299

Davis, L. E., Shalin, S. C., & Tackett, A. J. (2019). Current state of melanoma diagnosis and treatment. *Cancer Biology*

& Therapy, *20*(11), 1366–1379. https://doi.org/10.1080/15384047.2019.1640032

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them. *Management Science*, *64*(3), 1155–1170. https://doi.org/10.1287/mnsc.2016.2643

Dragicevic, N., Ullrich, A., Tsui, E., & Gronau, N. (2020). A conceptual model of knowledge dynamics in the industry 4.0 smart grid scenario. *Knowledge Management Research & Practice*, *18*(2), 199–213. https://doi.org/10.1080/14778238.2019.1633893

Felzmann, H., Fosch-Villaronga, E., Lutz, C., & Tamo-Larrieux, A. (2019). Robots and Transparency: The Multiple Dimensions of Transparency in the Context of Robot Technologies. *IEEE Robotics & Automation Magazine*, *26*(2), 71–78. https://doi.org/10.1109/MRA.2019.2904644

Haenssle, H. A., Fink, C., Schneiderbauer, R., Toberer, F., Buhl, T., Blum, A., Kalloo, A., Hassen, A. B. H., Thomas, L., Enk, A., Uhlmann, L., Reader study level-I and level-II Groups, Alt, C., Arenbergerova, M., Bakos, R., Baltzer, A., Bertlich, I., Blum, A., Bokor-Billmann, T., … Zalaudek, I. (2018). Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Annals of Oncology: Official Journal of the European Society for Medical Oncology*, *29*(8), 1836–1842. https://doi.org/10.1093/annonc/mdy166

Haselager, P., Schraffenberger, H., Thill, S., Fischer, S., Lanillos, P., van de Groes, S., & van Hooff, M. (2023). Reflection Machines: Supporting Effective Human Oversight Over Medical Decision Support Systems. *Cambridge Quarterly of Healthcare Ethics*, 1–10. https://doi.org/10.1017/S0963180122000718

Jacobsen, R. M., Johansen, P. S., Bysted, L. B. L., & Skov, M. B. (2020). Waste Wizard: Exploring Waste Sorting using AI in Public Spaces. *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, 1–11. https://doi.org/10.1145/3419249.3420180

Krüger, M., Wiebel, C. B., & Wersing, H. (2017). From Tools Towards Cooperative Assistants. *Proceedings of the 5th International Conference on Human Agent Interaction*, 287–294. https://doi.org/10.1145/3125739.3125753

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, *267*, 1–38. https://doi.org/10.1016/j.artint.2018.07.007

Orlikowski, W. J. (2007). Sociomaterial Practices: Exploring Technology at Work. *Organization Studies*, *28*(9), 1435–1448. https://doi.org/10.1177/0170840607081138

Papachristos, E., Skov Johansen, P., Møberg Jacobsen, R., Bjørn Leer Bysted, L., & Skov, M. B. (2021). How do People Perceive the Role of AI in Human-AI Collaboration to Solve Everyday Tasks? *CHI Greece 2021: 1st International Conference of the ACM Greek SIGCHI Chapter*, 1–6. https://doi.org/10.1145/3489410.3489420

Papagni, G., & Koeszegi, S. (2020). Understandable and trustworthy explainable robots: A sensemaking perspective. *Paladyn, Journal of Behavioral Robotics*, *12*(1), 13–30. https://doi.org/10.1515/pjbr-2021-0002

Tschandl, P., Codella, N., Akay, B. N., Argenziano, G., Braun, R. P., Cabo, H., Gutman, D., Halpern, A., Helba, B., Hofmann-Wellenhof, R., Lallas, A., Lapins, J., Longo, C., Malvehy, J., Marchetti, M. A., Marghoob, A., Menzies, S., Oakley, A., Paoli, J., … Kittler, H. (2019). Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: An open, web-based, international, diagnostic study. *The Lancet. Oncology*, *20*(7), 938–947. https://doi.org/10.1016/S1470-2045(19)30333-X

Tschandl, P., Rinner, C., Apalla, Z., Argenziano, G., Codella, N., Halpern, A., Janda, M., Lallas, A., Longo, C., Malvehy, J., Paoli, J., Puig, S., Rosendahl, C., Soyer, H. P., Zalaudek, I., & Kittler, H. (2020). Human-computer collaboration for skin cancer recognition. *Nature Medicine*, *26*(8), 1229–1234. https://doi.org/10.1038/s41591-020-0942-0

Wang, D., Khosla, A., Gargeya, R., Irshad, H., & Beck, A. H. (2016). *Deep Learning for Identifying Metastatic Breast Cancer*. https://doi.org/10.48550/ARXIV.1606.05718

Zafari, S., & Koeszegi, S. T. (2021). Attitudes Toward Attributed Agency: Role of Perceived Control. *International Journal of Social Robotics*, *13*(8), 2071–2080. https://doi.org/10.1007/s12369-020-00672-7

Zafari, S., Köszegi, S.T. and Filzmoser, M., S. (2021). Human Adaption in the Collaboration with Artificial Agents. In *.), Konnektivität Über die Bedeutung von Zusammenarbeit in der virtuellen Welt* (S. 97–106). Waxmann Verlag GmbH. http://hdl.handle.net/20.500.12708/30581

Zammuto, R. F., Griffith, T. L., Majchrzak, A., Dougherty, D. J., & Faraj, S. (2007). Information Technology and the Changing Fabric of Organization. *Organization Science*, *18*(5), 749–762. https://doi.org/10.1287/orsc.1070.0307

Group Decision and Negotiation
- Risk evaluation and negotiation strategies -

Conference papers and extended abstracts

A study on risk allocation of public-private partnership to brownfield remediation projects in China

Yan Qin [1], Qing Xu [1], Ginger Y. Ke [2*] and Qingye Han [1*]

[1] College of Management Science and Real Estate, Chongqing University, Chongqing 400044, China
[2] Faculty of Business Administration, Memorial University of Newfoundland, St. John's, NL, Canada A1B 3X5
[*] gke@mun.ca [*] hanqy11@cqu.edu.cn

Abstract

With the renovation and relocation of highly polluting urban enterprises, brownfield sites have proliferated in recent years. Due to the public ownership of land and difficulty in delineating pollution responsibility, brownfield remediation in China relies heavily on local governments. Shortage of funds is a major obstacle. Although government-enterprise cooperation is considered an effective solution, financing brownfield remediation with social capital remains exploratory. Focusing on risk allocation in government-enterprise cooperation for brownfield remediation, this study identifies risk factors, determines evaluation indexes using the entropy weighting method, and constructs a risk allocation model with the TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) method. A case study is conducted to test the feasibility and applicability.

Keywords: brownfield; public-private partnership; risk allocation

1. Introduction

The global process of industrialization and deindustrialization has led to the emergence of numerous brownfields. Many developed countries have prioritized brownfield remediation and redevelopment, introducing incentive policies, such as tax reduction and exemption and special funds, to realize the potential economic, social, and environmental benefits (Sessa et al., 2022). However, China's awareness of brownfield sites is relatively late, and its policy system, especially regarding financing, is not yet well-developed (Han et al., 2019). Unlike private ownership in most countries, land in China is owned by the state or rural collectives. Additionally, many polluting companies have gone bankrupt or been reorganized under China's retreat to three policies. As a result, the Chinese government bears the responsibility for brownfield remediation (BR) projects, even amid fiscal deficits. The public-private partnership (PPP) is considered a potentially effective solution (Zhang et al., 2022), from which the local governments can alleviate financial pressure by seeking social capital.

However, because of the high uncertainty of BR projects and PPP projects, both the government and enterprises face high risks. Risk management is crucial to ensure the smooth operation and functioning of BR projects under the PPP mode. Generally, risk management of PPP projects includes risk identification, assessment, sharing, and control (Yuan et al., 2015). Studies have shown that reasonable risk sharing is a key factor in the success of PPPs and an effective measure for both the government and the enterprise to resolve potential conflicts (Jiang et al., 2023).

Existing research on brownfield remediation and public-private partnerships (PPP) focuses more on negotiation issues (Glumac et al., 2015), success factors (Li et al., 2016), and obstacles (Zhang et al., 2022). Most studies on the risks of BR projects are restricted in risk identification and assessment (Han et al., 2019), with limited research on risk allocation, while PPP risk allocation is mainly concentrated in areas such as water management projects (Shrestha et al., 2018), urban rail transit projects (Huang et al., 2022), and construction waste recycling projects (Wang et al., 2022). From the methodology perspective, on the other hand, the TOPSIS method is suitable for decision-making processes involving multiple evaluation criteria, which precisely describes the risk allocation problem for BR projects. Therefore, TOPSIS can properly assist decision-makers in weighing the importance of each criterion to determine the most suitable project.

2. Determination of risk evaluation indicators for BR projects under the PPP mode

2.1. Identification of initial risk factors for BR projects under PPP mode

By reviewing the relevant literature, the 12 most representative papers were screened, which had significant advantages in citation rate and influence or were published in high-level professional journals. From these 12 papers, a list of 48 initial risk factors is identified, as shown in Table 1.

Table 1 Initial risk list for BR projects under the PPP mode

No.	Risk factors	No.	Risk factors
1	Corruption	25	Insufficient experience with the government-enterprise partnership model
2	Stigma	26	Inadequate financial audits
3	Opposition from the public or other branches of government	27	Inappropriate allocation of risk responsibility or authority
4	Incorrect cost estimation	28	Environmental and ecological risks
5	Differences in implementation of relevant legal provisions	29	Delayed completion
6	Planning risks	30	Private investor capacity deficiencies
7	Funding availability	31	Insufficient attractiveness of the project to the investor
8	Stability risks	32	Changes in market demand
9	Excessive contractual changes	33	Substandard rehabilitation
10	Perceived risk and land epidemics	34	Force majeure
11	Delayed supply	35	Health risks
12	Impact on reputation and image	36	Delay in resolving contractual disputes, litigation or arbitration disputes
13	Technical Risks	37	Impact on the value of surrounding fixed assets
14	Lack of standard templates for government-enterprise cooperation agreements	38	Poor contract management skills
15	Cost overruns	39	Loan repayment risks
16	Failure to meet expected benefits	40	Changes in design programs
17	Lack of competitive bidding process	41	Regulatory risk
18	High financing costs	42	Project processing and approvals
19	Organizational and coordination risks	43	Liability risk
20	Lack of research and data	44	Immature relevant legal system
21	Change of private investors	45	Changes in remediation standards
22	High maintenance risk	46	Stakeholder reliability
23	Differences in the way government and business work	47	Other city projects
24	Sanctions and penalties	48	Litigation risks

2.2. Re-screening based on expert interview

Based on the 48 preliminary risk factors identified in Section 2.1, combined with the expert interview method, a questionnaire is used to further validate and determine the final risk factors. Invite 20 research scholars related to the fields of construction, land, engineering, and PPP field to conduct expert validation. The experts score the probability of occurrence and the degree of influence of the initial risk factors from 1 to 5. The final score of each risk factor is obtained by multiplying the two scores. Consequently, risk factors are categorized into three classes: high risk (18-25 points), medium risk (11-17 points), and low risk (1-10 points).

Risk factors with a score of 10 or more are selected as risk factors for the BR PPP projects studied in this paper. Based on the mean scores, 15 risk factors, numbered 1, 2, 4, 7, 11, 15, 16, 27, 28, 30, 32, 40, 42, 44, 46, are selected.

3. Construction of risk allocation model for BR projects under PPP mode

3.1. Determination of risk allocation indicators for BR projects under PPP mode

BR projects under PPP mode involve cooperation between the public and private sectors and require both parties to follow certain principles to reasonably and effectively allocate project risks to maximize the interests of both parties. Therefore, a scientific and reasonable risk allocation index system is needed to measure and guide the process of risk allocation. Through the literature review, four positive level risk allocation indicators are identified, namely risk perception ability, risk management ability, risk-taking cost, and risk-taking willingness.

3.2. Determination of risk allocation indicator weights for BR projects under the PPP mode

Different indicators have various impacts on the risk allocation of the project. Therefore, it is necessary to assign weights to the indicators that match their degree of contribution. The entropy weighting method relies only on the discrete nature of the data itself and is hence considered an objective method of assignment. The calculation steps are as follows.

(1) Standardize the judgment matrix.

$$b_{ij} = \frac{z_{ij}-z_{min}}{z_{max}-z_{min}} \qquad (3.1)$$

where z_{max}, z_{min} represent the most and least satisfied with the value of the program (usually, a smaller value indicates being more satisfied); b_{ij} $(i = 1, 2, \ldots, m; j = 1, 2, \ldots, n)$ is the standardized value of the j^{th} indicator of the i^{th} sample.

(2) Calculate the information entropy of each indicator according to the definition of information entropy in information theory.

$$E_j = -K \sum_{i=1}^{m} P_{ij}\, In(P_{ij}) \qquad (3.2)$$

where $K = 1/In(m)$, $P_{ij} = \frac{z_{ij}}{\sum_{i=1}^{m} z_{ij}}$, $0 \le E_j \le 1$.

(3) The weight of each indicator is calculated as

$$W_j = \frac{1-E_j}{K-\sum E_j}(i = 1, 2 \ldots, k) \qquad (3.3)$$

3.3. Construction of risk allocation model for BR project under PPP mode

The TOPSIS model is a comprehensive distance assessment method proposed by Hwang and Yoon (1981). To improve the reliability of the evaluation results, this study invites experts to evaluate the degree of risk-taking by governments and enterprises using a five-point scoring method, where the five levels are very small risk, low risk, medium risk, high risk, and very high risk. Each linguistic variable is expressed to the corresponding score.

In this study, it is clear that BR projects under PPP have two risk-allocating stakeholders, the government and social capital. Each stakeholder has 15 risk allocation indicators. So the original data matrix is:

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix} \qquad (3.4)$$

where $x_{ij}(i = 1, 2, \ldots, n; j = 1, 2, \ldots, m)$ is the j^{th} risk allocation indicator that represents the i^{th} risk allocation stakeholder. The data are normalized to construct the value matrix:

$$R = (r_{ij})_{n\times m} \qquad (3.5)$$

where r_{ij} is the standardized value of the j^{th} risk allocation indicator of the i^{th} risk allocation stakeholder and $r_{ij} \in [0, 1]$.

Calculate the normalized matrix $V_{ij} = w_j \times r_{ij}$, where w_j is the weight. Here, the weights are derived by the entropy weighting method. According to Matrix V_{ij}, the optimal V_j^* and worst V_j^- on each indicator can be determined.

Then, the Euclidean distances between each indicator value and respectively the corresponding optimal and worst solutions are calculated:

$$S_i^* = \sqrt{\sum_{j=1}^{n} (V_{ij} - V_j^*)^2} \qquad (3.6)$$

$$S_i^- = \sqrt{\sum_{j=1}^{n} (V_{ij} - V_j^-)^2} \qquad (3.7)$$

The closeness is thus:

$$C_i = \frac{S_i^-}{S_i^* + S_i^-} \qquad (3.8)$$

Sort the resulting values of C_i, where a larger value indicates a better decision.

4. Case study

In this paper, the effectiveness and applicability of the proposed risk allocation model are discussed in depth and verified using Project J (a pesticide factory) in the Jingkou Town of Chongqing, China, as a case study.

4.1. Determination of the weights of risk allocation indicators for the J BR project under PPP mode

Herein, five government and enterprise experts previously engaged in the PPP industry are invited to assess the importance of the four positive evaluation indicators of risk allocation (15-25 points: high risk; 5-14 points: medium risk; 1-4 points: low risk). Then the weights of each indicator are calculated by the entropy weight method (Table 2). As can be seen from the table, risk management ability and risk perception ability occupy the largest and smallest weights.

Table 2 Entropy weight coefficient and ranking of risk indicators

Index	Information entropy value *e*	Information utility value *d*	Weighting factor ranking *w*	Rank
Risk-taking cost	0.7936	0.2064	18.64%	3
Risk-taking willingness	0.6728	0.3272	29.54%	2
Risk management capability	0.5850	0.4150	37.46%	1
Risk perception ability	0.8409	0.1591	14.36%	4

4.2. Calculation of risk allocation ratio for Project J

Ten experts from universities, consulting companies, land reserve organizations, and real estate development companies are further invited to evaluate the risk allocation indexes of the 15 key risk factors in Section 2 with 1-5 points. The scores with obvious differences are excluded, and the average values are taken as the initial evaluation values.

Thus, a matrix of initial value ratings for each risk factor can be obtained:

$$X_1 = \begin{bmatrix} 3.6 & 3.6 & 2.4 & 3.4 \\ 2.5 & 1.9 & 2.1 & 1.6 \end{bmatrix} X_2 = \begin{bmatrix} 3.3 & 3.2 & 3.5 & 3.7 \\ 3.4 & 4.6 & 2.9 & 3.9 \end{bmatrix} X_3 = \begin{bmatrix} 3.7 & 3.5 & 4.1 & 1.2 \\ 1.8 & 2.1 & 1.9 & 2.2 \end{bmatrix}$$
$$X_4 = \begin{bmatrix} 1.7 & 1.9 & 1.8 & 2.5 \\ 3.5 & 4.1 & 3.3 & 3.5 \end{bmatrix} X_5 = \begin{bmatrix} 3.6 & 2.5 & 2.9 & 1.5 \\ 3.5 & 3.8 & 3.8 & 2.9 \end{bmatrix} X_6 = \begin{bmatrix} 3.6 & 3.7 & 3.9 & 2.1 \\ 2.6 & 3.3 & 3.7 & 3.8 \end{bmatrix}$$
$$X_7 = \begin{bmatrix} 3.6 & 3.8 & 2.8 & 2.9 \\ 2.7 & 2.5 & 3.2 & 3.9 \end{bmatrix} X_8 = \begin{bmatrix} 2.1 & 1.6 & 3.6 & 2.5 \\ 3.6 & 4.3 & 2.3 & 4.6 \end{bmatrix} X_9 = \begin{bmatrix} 1.9 & 2.3 & 2.1 & 1.3 \\ 3.2 & 3.5 & 4.1 & 4.3 \end{bmatrix}$$
$$X_{10} = \begin{bmatrix} 3.4 & 3.2 & 3.1 & 3.4 \\ 3.5 & 4.1 & 4.6 & 4.3 \end{bmatrix} X_{11} = \begin{bmatrix} 3.5 & 3.6 & 3.9 & 3.8 \\ 3.3 & 3.9 & 3.6 & 4.1 \end{bmatrix} X_{12} = \begin{bmatrix} 4.2 & 3.9 & 4.1 & 3.9 \\ 2.1 & 2.4 & 1.9 & 2.4 \end{bmatrix}$$
$$X_{13} = \begin{bmatrix} 2.6 & 2.1 & 1.6 & 2.5 \\ 3.7 & 3.7 & 3.2 & 2.6 \end{bmatrix} X_{14} = \begin{bmatrix} 1.9 & 2.5 & 2.1 & 2.7 \\ 2.4 & 3.1 & 3.2 & 3.5 \end{bmatrix} X_{15} = \begin{bmatrix} 3.4 & 2.3 & 2.7 & 3.5 \\ 2.6 & 1.6 & 2.5 & 1.4 \end{bmatrix}$$

where X_i refers to the i^{th} risk factor, the first row of the matrix represents the public sector, the second row shows the private sector, and the columns correspond to the four risk indicators Y_1、Y_2、Y_3、Y_4.

Next, by applying the TOPSIS method, we can compute the closeness values for all risk factors, as given in Table 3.

Table 3 Results of risk factor calculations

No.	Risk factors	Partner	Positive ideal solution distance	Negative ideal solution distance	Relative proximity	Rank
1	Corruption	Public	0	1.414	1	1
		Private	1.414	0	0	2
2	Funding availability	Public	1.276	0.611	0.324	2
		Private	0.611	1.276	0.676	1
3	Immature relevant legal system	Public	0.769	1.187	0.607	1
		Private	1.187	0.769	0.393	2
4	Cost overruns	Public	1.414	0	0	2
		Private	0	1.414	1	1
5	Stigma	Public	1.309	0.536	0.291	2
		Private	0.536	1.309	0.709	1
6	Failure to meet expected benefits	Public	0.769	1.187	0.607	1
		Private	1.187	0.769	0.393	2
7	Stakeholder reliability	Public	0.982	1.018	0.509	1
		Private	1.018	0.982	0.491	2
8	Incorrect cost estimation	Public	1.276	0.611	0.324	2
		Private	0.611	1.276	0.676	1
9	Private investor capacity deficiencies	Public	1.414	0	0	2
		Private	0	1.414	1	1
10	Environmental and ecological risks	Public	1.414	0	0	2
		Private	0	1.414	1	1
11	Changes in market demand	Public	1.158	0.812	0.412	2
		Private	0.812	1.158	0.588	1
12	Project processing and approvals	Public	0	1.414	1	1
		Private	1.414	0	0	2
13	Changes in design programs	Public	1.414	0	0	2
		Private	0	1.414	1	1
14	Delayed supply	Public	1.414	0	0	2
		Private	0	1.414	1	1
15	Improper allocation of risk responsibility or authority	Public	0	1.414	1	1
		Private	1.414	0	0	2

4.3. Results of risk allocation for brownfield remediation of Project J under PPP perspective

Based on Table 3 (the column of Relative Proximity), three types of risk allocation were identified:

(1) Risks should be allocated to the government,

(2) Risks should be allocated to social capital,

(3) Risks should be shared between the government and social capital.

As shown, 3 risk factors, “corruption”, “project processing and approvals”, and “improper allocation of risk responsibility or authority”, should be assigned to the government. 5 risk factors, “cost overruns”, “private investor capacity deficiencies”, “environmental and ecological risks”, “changes in design programs”, and “delayed supply” should be assigned to social capital. The remaining 7 factors should be shared between the government and the social capital, as shown in Table 4.

Table 4 Risk allocation ratio

Risk factors	Public sector share	Private sector share
Funding availability	0.324	0.676
Immature relevant legal system	0.607	0.393
Stigma	0.297	0.709
Failure to meet expected benefits	0.607	0.393
Stakeholder reliability	0.509	0.491
Incorrect cost estimation	0.324	0.676
Changes in market demand	0.412	0.588

5. Conclusion

Proper risk allocation between the public and private sectors is crucial for the success of PPP projects. However, achieving effective risk allocation can be challenging due to lacking a common risk management mechanism. This study identifies 15 key risks through a questionnaire and risk assessment matrix. Utilizing the entropy weighting method, four risk allocation indicators and their weights are determined. The TOPSIS method is then used to calculate allocation ratios for the 15 risks, with three assigned to the public sector, five to the private sector, and seven shared between both sectors. The findings of this study can lead public and private sectors to pay attention to 15 key risks, especially those that are borne entirely by their own parties. However, owing to the lack of public BR PPP project information, risk factor identification should be updated. Besides, future studies may contemplate the utilization of alternative multi-criteria decision-making techniques for cross-validation of findings.

Acknowledgment

This research was funded by National Natural Science Foundation of China (Award Number: 72104040), Chongqing Social Science Planning Fund (Award Number: 2020QNGL27), Central University Basic Research Fund of China (Award Number: 2021CDJSKJC21), Postdoctoral Research Foundation of China (Award Number: 2020M683253) and Natural Science Foundation Project of Chongqing, Chongqing Science and Technology Commission (Award Number: cste2020jcyjbshX0038). The third author also appreciates the support from Discovery Grants from the Natural Sciences and Engineering Research Council of Canada (RGPIN-2015-04013, RGPIN-2022-03514).

References

Glumac, B., Han, Q., Schaefer, W.W., & Krabben, E.V. (2015). Negotiation issues in forming public–private partnerships for brownfield redevelopment: Applying a game theoretical experiment. *Land Use Policy, 47,* 66-77.

Han, Q., Zhu, Y., Ke, G. Y., & Lin, H. (2019). A Two-Stage Decision Framework for Resolving Brownfield Conflicts. *International Journal of Environmental Research and Public Health, 16*(6), 1039.

Han, Q., Zhu, Y., Ke, G. Y., & Hipel, K. W. (2019). Public private partnership in brownfield remediation projects in China: Identification and structure analysis of risks. *Land Use Policy, 84,* 87-104.

Huang, Y., Xu, W., & Li, C. (2022). Information integration framework for a public–private partnership system of an urban railway transit project (Part A: System architecture). *Journal of Industrial Information Integration, 25,* 100244.

Jiang, F., Lyu, Y., Zhang, Y., & Guo, Y. (2023). Research on the Differences between Risk-Factor Attention and Risk Losses in PPP Projects. *Journal of the Construction Division and Management, 149*(9).

Li, X., Yang, H., Li, W., & Chen, Z. (2016). Public-private partnership in residential brownfield redevelopment: case studies of Pittsburgh. *Procedia Engineering, 145,* 1534-1540.

Sessa, M. R., Russo, A., & Sica, F. (2022). Opinion paper on green deal for the urban regeneration of industrial brownfield land in Europe. *Land Use Policy, 119,* 106198.

Shrestha, A., Chan, T., Aibinu, A. A., Chen, C., & Martek, I. (2018). Risk Allocation Inefficiencies in Chinese PPP Water Projects. *Journal of the Construction Division and Management, 144*(4).

Wang, Z., Zhou, Y., Jin, X., Zhao, N., & Sun, J. (2022). Risk allocation and benefit distribution of PPP projects for construction waste recycling: a case study of China. *Engineering, Construction and Architectural Management, 30*(9), 3927–3956.

Yuan, J., Chan, A. P. C., Xiong, W., Skibniewski, M. J., & Li, Q. (2015). Perception of Residual Value Risk in Public Private Partnership Projects: Critical Review. *Journal of Management in Engineering, 31*(3).

Zhang, H., Liu, G., Han, Q., & Chen, G. (2022). Mapping the Barriers of Utilizing Public Private Partnership into Brownfield Remediation Projects in the Public Land Ownership. *Land, 12*(1), 73.

Citizen collaboration in urban renewal: a fuzzy-based decision framework

Sha Lai [1], Qingye Han [1*] and Ginger Y. Ke [2*]

[1] Chongqing University, Chongqing, China
[2] Memorial University of Newfoundland, St. John's, Canada
*hanqy11@cqu.edu.cn
*gke@mun.ca

Abstract

Citizen collaboration has proven to be an effective way to tackle the challenges faced by urban renewal due to its capability of sharing costs, ensuring progress, and adjusting decisions. However, citizens' awareness of collaboration is relatively weak, and hence their participation is less than expected. This study proposes a decision framework that can be used to assist in achieving citizen collaboration in urban renewal. To be specific, indicators representing the explanatory conditions and outcomes of citizen collaboration are first derived from the existing literature and expert opinions. Then, through field research and interviews, key data on 16 urban renewal cases are collected and analyzed to summarize paths and strategies toward high levels of citizen collaboration by using Fuzzy Set Qualitative Comparative Analysis (fsQCA). The analytical results show that (i) there are three paths to high levels of citizen collaboration, namely interest-driven, culture-driven, and resource-driven paths, and (ii) government credibility is the core explanatory condition for all three paths. Corresponding strategies for enhancing government credibility are presented as well.

Keywords: Citizen collaboration; Urban renewal; fsQCA; Decision-making

1. Introduction

Urban renewal, reusing the existing resources (such as housing, infrastructure, and environment) and space, is an important issue faced by most cities with high rates of urbanization because it can deal with urban decay and bring new prosperity again (Deng et al., 2020; Wu et al., 2020). It also faces some challenges, such as the neglect of some groups' needs and the limited resources from the government, which can be obstacles to sustainable development (Huang et al., 2020). Around the world, urban renewal has been widely a public concern. Particularly for the context of urban renewal in China, it is much more complicated than that in developed countries. First, urban renewal in China has a strong attribute of public service. Secondly, the ownership and use rights of urban construction land are separated in China. Last but not least, a settlement in China usually consists of many high-rise dwellings, which adds more difficulties to the situation.

Community, as the fundamental component of urban space, seems to be the most suitable scale for analyzing citizens' behavior (Cheng et al., 2022; Pérez et al., 2018). Especially, after a long period of demolition and reconstruction, urban renewal in China has shifted to the renovation of existing buildings in response to sustainability. China has made "the renovation of old urban residential communities" an annual priority for the government. However, the collaboration of citizens, the most important stakeholders in the renewal process, has often been overlooked because of their weak sense of participation and less effective means of participation (Liu et al., 2020; Zhai & Ng, 2013).

To support precise decision-making for achieving citizen participation, this study first proposes a framework that can be used to summarize paths to citizen collaboration in urban renewal by analyzing the causal relationship between explanatory conditions and citizen collaboration using Fuzzy Set Qualitative

Comparative Analysis (fsQCA). As a type of QCA, fsQCA is more suitable for analyzing data expressed as continuous values. It can both accommodate data from a large number of cases and obtain sufficient configurations (combinations of multiple conditions) that achieve the target outcome (Fiss, 2011). These strengths distinguish fsQCA from qualitative analysis methods and linear regression methods commonly used in existing research, which can only yield correlations between explanatory conditions and resident cooperation. Then, empirical analyses of 16 urban renewal cases at the community level are conducted to obtain specific paths to achieve citizen cooperation and strategies to improve the core explanatory conditions. The results of the study can help the government guide citizens to participate in urban renewal, thus increasing citizens' motivation and satisfaction, and making better decisions in the process of urban renewal.

2. Conceptual framework

The framework integrates an indicator measurement system and an algorithm for analyzing causal relationships. By applying the indicator system, the explanatory conditions and levels of citizen collaboration in the cases can be assessed. Drawing on the results, equivalent and diverse paths to achieving citizen collaboration can be generated via the algorithm.

Multiple methods are adopted to develop the proposed framework. In more detail, the literature review and Delphi methods are applied to identify indicators that represent explanatory conditions and outcomes of citizen collaboration. Then, combined with field research, exact indicator measurements can be obtained. Finally, the algorithm is established with fsQCA.

2.1. Identifying and measuring the condition indicators

Condition indicators (X_1, X_2, X_3, X_4, X_5), i.e., explanatory conditions corresponding to the level of citizen collaboration, are identified through literature review and expert opinions. Specifically, 31 condition indicators affecting co-production in urban renewal are first distilled through the literature review. Then, five condition indicators were summarized and condensed following the steps of the Delphi method. The resulting indicators are shown in Table 1.

Table 1. Condition indicators

Condition indicators	Symbols	Sub-indicators	Symbols
External environment	X_1	Policy Environment	X_{1_1}
		Value of integrated resources around the neighborhood	X_{1_2}
		Economic value of the building	X_{1_3}
Internal environment	X_2	Comprehensive characteristics of citizens	X_{2_1}
		Internal relationship network	X_{2_2}
		Number of highly motivated citizens	X_{2_3}
		Neighborhood cultural environment	X_{2_4}
External resource support	X_3	External financial input	X_{3_1}
		The cooperation from external roles	X_{3-2}
		Richness of external roles	X_{3_3}
The interest-driven force from stakeholders	X_4	The fit between the type of urban renewal and citizens' interests	X_{4_1}
Government's credibility	X_5	Truth-building between government and citizens	X_{5_1}
		Raw trust between the government and citizens	X_{5_2}

Based on the identified condition indicators, this study uses the Multiple Indicator Scoring and Summation method to facilitate the measurement, which is nondimensionalized by the utility value within the range of [0, 1]. Finally, the score of the indicator is obtained by aggregating the corresponding sub-indicators. Note that we herein remove sub-indicators that are not easy to obtain data or cannot be understood by the interviewees and replace them with similar sub-indicators. The measurement of the condition indicators can be determined as:

$$Value(X_i) = Score(X_{i-1}) + Score(X_{i-2}) + \ldots + Score(X_{i-m}) \,, \tag{1}$$

where m denotes the number of sub-indicators for X_i, and X_{i-m} is the sub-indicator for X_i.

2.2. Defining the outcome indicator

The outcome indicator (Y) refers to the level of citizen collaboration, which corresponds to the sum of behaviors regarding collaboration from citizens. This paper uses citizen participation and input to measure this indicator. In terms of a full life cycle, inputs include those in the process and after completion, and participation includes citizen interaction and cooperation throughout the entire process. Accordingly, the measurements can be computed with Eqs. (2)-(6). The secondary or tertiary indicators are defined in Table 2.

$$Value(Y) = Score(Y_1)*25\% + Score(Y_2)*25\% + Score(Y_3)*25\% + Score(Y_4)*25\% \tag{2}$$

$$Value(Y_1) = Score(Y_{1\_1})*50\% + Score(Y_{1\_2})*50\% \tag{3}$$

$$Value(Y_2) = Score(Y_{2\_1})*50\% + Score(Y_{2\_2})*50\% \tag{4}$$

$$Value(Y_3) = Score(Y_{3\_1})*50\% + Score(Y_{3\_2})*50\% \tag{5}$$

$$Value(Y_4) = Score(Y_{4\_1})*50\% + Score(Y_{4\_2})*50\% \tag{6}$$

Table 2. Outcome indicator

Outcome indicator	Symbols	Sub-indicators	Tertiary indicators
The level of co-production	Y	Y_1 : inputs during the process	Y_{1_1} : direct investment and abandonment of public benefit s by citizens
			Y_{1_2} : labor input by citizens during the process
		Y_2 : inputs after completion	Y_{2_1} : citizens' payment for public services during the mai ntenance phase
			Y_{2_2} : labor input by citizens for maintenance
		Y_3 : citizen's interaction	Y_{3_1} : suggestions made by citizens in response to the rene wal of public areas
			Y_{3_2} : suggestions made by citizens in response to persona l interests
		Y_4 : citizen's cooperation	Y_{4_1} : cooperation of citizens in the survey process
			Y_{4_2} : attendance of citizens at public meetings

2.3. Calibration of indicators

After collecting data from the case set based on the indicators and measurements, we transform the data into fuzzy sets, which is a key step known as data calibration in fsQCA. Since participants in the urban renewal market evaluate the strengths and weaknesses of the community environment based on relative levels, it is appropriate for this study to use the calibration method, which is based on the relative position of the cases.

Based on the suggested values given by the Tosmana software (Cronqvist, 2019) and the author's judgment of the actual situation, the three thresholds were set to 95% (full in membership: N_1), 0.5 (crossover point: N_2), and 0.05 (full-out membership: N_3) of the cases (Andrews et al., 2016). Note that the membership function value indicates the degree of cases belonging to the target set.

2.4. Generating Paths to Co-production by Analysing Configuration

To empirically complete the generation of paths to citizen collaboration, fsQCA proceeds in three steps. The first step is using these calibrated set measures to construct a truth table, each row of which corresponds to a specific configuration of explanatory conditions. The cases are sorted into the truth table based on the set membership of explanatory conditions. In the second step, the rows of the truth table are reduced by two constraints: (1) the consistency threshold, i.e., the minimum level of consistency for a combination, and (2) the frequency threshold, i.e., the minimum number of cases required for a solution. “Consistency” in this context, calculated by Eq. (7), describes the degree to which the cases correspond to the set-theoretic relationships expressed in the combination. To improve the effectiveness of the solution, the value of 0.8 is selected as the minimum acceptable consistency, which is higher than the minimum recommended threshold of 0.75 (Ragin & Fiss, 2008). On the other hand, the frequency threshold ensures that a minimum number of empirical observations is obtained given the sample size (Schneider & Wagemann, 2012). The frequency threshold is set to 1 for a rather small case set. In the third step, the truth table is logically analyzed using the Boolean algebra-based algorithm to obtain the solutions. Based on this, three types of solutions can be obtained, namely the complex, intermediate, and parsimonious solutions. Each type of solution has its characteristics (as shown in Table 3). Based on these characteristics, the intermediate solution is chosen as the primary source for summarizing the paths to citizen collaboration, and parsimonious solutions as assistance in finding the core explanatory conditions in the paths.

$$\mathrm{Consistency}(C_j) = \frac{\sum_{k=1\ldots n} \left(Y_k^* \cap \min_{i=1,2\ldots5}(X_{ik}^*)\right)}{\sum_{k=1\ldots n} \min_{i=1,2\ldots5}(X_{ik}^*)}, \tag{7}$$

where N denotes the number of cases in the sample set, C_j is the jth conditional configuration, Y_k^* is the set membership of the outcome indicator in case k and $\min_{i=1,2\ldots5}(X_{ik}^*)$ computes the set membership of the conditional configuration C_j in case k.

Table 3. Three types of solutions

Solution Type	Characteristics
Complex solution	Not include analysis of configurations that are not covered in the case set.
Intermediate solution	Include analysis of easily identifiable parts of configurations that are not covered in the case set (e.g. incorporating necessary conditions into the analysis).
Parsimonious solution	Include analysis of all configurations that are not covered in the case set.

3. Case study

Located in Southwest China, as one of the municipalities directly under the central government of China, Chongqing is a mega-city with an urban population of more than 10 million. With a population of 10,477,600 in its central urban area and an urbanization rate of 93.3%, the city has a strong demand for renewal. By the end of 2022, a total of 3,993 old urban communities have been renovated in Chongqing. There are many completed urban renewal projects with significant variations, allowing for a balance between the core requirements of "similarity" and "heterogeneity". This study develops the case set in a database of completed urban renewal cases in six central urban areas. The final selection identifies 16 cases that represent the city of Chongqing and are generalizable across the country to constitute the urban renewal case set used for empirical purposes.

The results facilitate three sufficient paths (as shown in Table 4). P_1 represents the culture-driven path based on place identity, preferable for communities with long histories and high cultural effects, which contain citizens who have strong place identities. P_2 reflects the profit-driven path based on place dependence, focusing on communities with vulnerable external environments, including external policies, geographical location, or economic value of housing. P_3 depicts the resource-driven path based on the support of external actors, suitable for communities with great potential for resource redevelopment, which produces benefits for residents, profits for enterprises, and reduced burdens for the local government.

Table 4. The result of the case study

The result of fsQCA	Symbol of the path	Characteristics of the path	Communities suitable for the path
Configuration 1	P_1	The culture-driven path based on place identity	Communities with long histories and high cultural effects
Configuration 2	P_2	The profit-driven path based on place dependence	Communities with vulnerable external environments
Configuration 3	P_3	The resource-driven path based on the support of external actors	Communities with great potential for resource redevelopment

From the empirical results, it is clear that government credibility plays a key. This is consistent with the view of Chanley et al. (Chanley et al., 2000)that political trust has an important impact on policy attitudes. Through in-depth analyses of the 16 typical urban renewal cases in real-life situations, this study summarises the lessons that can help enhance government credibility in urban renewal at the community level from two aspects, 1) improving citizens' access to information and 2) building interpersonal trust.

4. Conclusion

This study aims to support decision-making in urban renewal through citizen collaboration. Based on the identified explanatory conditions of citizen collaboration in urban renewal at the community level, including the external and internal environment, external resource support, interest-driven forces of stakeholders, and credibility of the government, this study utilizes the fsQCA methodology to analyze the condition configurations of 16 typical Chinese urban renewal cases and finds that communities in different contexts and management styles have different paths to achieve citizen collaboration.

First, communities with disadvantages in the external environment and where residents are physically dependent due to economic constraints are suitable for the interest-driven path. A high level of citizen collaboration can be fostered by creating high credibility for the government and high interest-drivenness for the residents. Secondly, communities with a long history and culture, a large external effect, and residents with spiritual dependence are suitable for a culture-driven path, corresponding to the high level of citizen collaboration that can be realized by shaping the high credibility of the government and the high-level support from external resources. Finally, communities with high potential for resource development are suitable for the resource-driven path, which can realize the model of "residents receive benefits, enterprises gain profits, and the government reduces burdens", and should correspond to the realization of high levels of citizens' collaboration facilitated by shaping the government's high credibility, and external enterprises that are highly supportive and have interest-driven power.

In summary, this study concludes that citizen collaboration is an effective way to deal with the challenges facing urban renewal. We need to better understand its process and influencing mechanisms to provide better guidance for decision-making. By proposing an analytical framework to explore the complex causal relationship between explanatory conditions and citizen collaboration in urban renewal at the community level, this study deserves the attention of policymakers and administrative leaders to provide a decision-making basis for future innovative practices in urban renewal at the community level.

Acknowledgement

This research was funded by National Natural Science Foundation of China (Award Number: 72104040), Chongqing Social Science Planning Fund (Award Number: 2020QNGL27), Central University Basic Research Fund of China (Award Number: 2021CDJSKJC21), Postdoctoral Research Foundation of China (Award Number: 2020M683253) and Natural Science Foundation Project of Chongqing, Chongqing Science and Technology Commission (Award Number: cste2020jcyjbshX0038). The third author also appreciates the support from Discovery Grants from the Natural Sciences and Engineering Research Council of Canada (RGPIN-2015-04013, RGPIN-2022-03514).

References

Andrews, R., Beynon, M. J., & McDermott, A. M. (2016). Organizational capability in the public sector: A configurational approach. *Journal of Public Administration Research and Theory*, *26*(2), 239-258. https://doi.org/10.1093/jopart/muv005

Chanley, V. A., Rudolph, T. J., & Rahn, W. M. (2000). The origins and consequences of public trust in government: A time series analysis. *Public opinion quarterly*, *64*(3), 239-256.

Cheng, J., Mao, C., Huang, Z., Hong, J., & Liu, G. (2022). Implementation strategies for sustainable renewal at the neighborhood level with the goal of reducing carbon emission. *Sustainable Cities and Society*, *85*, 104047. https://doi.org/10.1016/j.scs.2022.104047

Cronqvist, L. (2019). Tosmana [Version 1.61]. *University of Trier. Internet: https://www. tosmana. net*.

Deng, Y., Qi, W., Fu, B. J., & Wang, K. (2020). Geographical transformations of urban sprawl: Exploring the spatial heterogeneity across cities in China 1992-2015. *Cities*, *105*, Article 102415. https://doi.org/10.1016/j.cities.2019.102415

Fiss, P. C. (2011). Building better causal theories: A fuzzy set approach to typologies in organization research. *Academy of management journal*, *54*(2), 393-420.

Huang, L., Zheng, W., Hong, J., Liu, Y., & Liu, G. (2020). Paths and strategies for sustainable urban renewal at the neighbourhood level: A framework for decision-making. *Sustainable Cities and Society*, *55*, 102074. https://doi.org/10.1016/j.scs.2020.102074

Liu, G., Wei, L., Gu, J., Zhou, T., & Liu, Y. (2020). Benefit distribution in urban renewal from the perspectives of efficiency and fairness: A game theoretical model and the government's role in China. *Cities*, *96*, 102422. https://doi.org/10.1016/j.cities.2019.102422

Pérez, M. G. R., Laprise, M., & Rey, E. (2018). Fostering sustainable urban renewal at the neighborhood scale with a spatial decision support system. *Sustainable Cities and Society*, *38*, 440-451. https://doi.org/10.1016/j.scs.2017.12.038

Ragin, C. C., & Fiss, P. C. (2008). Net effects analysis versus configurational analysis: An empirical demonstration. *Redesigning social inquiry: Fuzzy sets and beyond*, *240*, 190-212.

Schneider, C. Q., & Wagemann, C. (2012). *Set-theoretic methods for the social sciences: A guide to qualitative comparative analysis*. Cambridge University Press. http://ci.nii.ac.jp/ncid/BB10589525

Wu, Y. Z., Galdini, R., Hui, E. C. M., & Long, H. L. (2020). Urban regeneration and re-use: China and Europe. *Cities*, *106*, Article 102863. https://doi.org/10.1016/j.cities.2020.102863

Zhai, B. Q., & Ng, M. K. (2013). Urban regeneration and social capital in China: A case study of the Drum Tower Muslim District in Xi'an. *Cities*, *35*, 14-25. https://doi.org/10.1016/j.cities.2013.05.003

24th International Conference on Group Decision and Negotiation &
10th International Conference on Decision Support System Technology
Human-Centric Decision and Negotiation Support for Societal Transitions

Research on the Evolutionary Game of Credit Risk of New Energy Vehicle Supply Chain Finance

Yeqing Guan*, Ying Cao

College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, P.R. China
* yuanovoo@163.com

Abstract

In recent years, China's new energy vehicles have broad development prospects. However, due to the high investment in research and development of new energy vehicles and the long return cycle, they are considered high-risk projects in the risk assessment of financial institutions, causing financing difficulties. Supply chain finance has been believed to be a feasible way to alleviate financing difficulties. However, due to the large number of participants in supply chain finance and the continuous innovation of its business models, credit risk events occur frequently. Based on the above background, in order to solve the credit risk problem of new energy vehicle enterprises under the supply chain finance model, this paper constructs a tripartite evolutionary game model among financing enterprises, core enterprises and financial institutions. First, this paper analyses the source of credit risk of new energy vehicle supply chain finance, including the operating risks of small and medium-sized financing enterprises and the principal-agent risk caused by the collusion between financing enterprises and core enterprises, and determine the basic assumptions and relevant parameter definitions of the evolutionary game model. Then based on the model assumptions and parameters, replicated dynamic equations are constructed and solved, 14 equilibrium points are obtained, and the local stability of 8 pure strategy equilibrium points is determined. When certain conditions are met, the tripartite evolutionary stability strategies are obtained: (default, passive supervision, lax supervision), (default, passive supervision, strict supervision), (default, active supervision, strict supervision) and (keeping promise, active supervision, lax supervision) respectively. Finally, the model is assigned numerical values based on the actual situation, and MATLAB is used to conduct numerical simulation to verify the effectiveness of the evolutionary stability analysis, and explore the impact of relevant parameter changes on the risk evolution mechanism of the game system, and put forward some suggestions to prevent credit risks. The analysis found that (1) Increasing the promise-keeping profits of financing enterprises, additional profits from the good operation of the supply chain, the revenue of financial institutions when defaults of financing enterprises are discovered, and the profits from active supervision of core enterprises are all conducive to controlling credit risks. (2) Reducing the cost of strict supervision of financial institutions and increasing the cost of lax supervision can better deal with the occurrence of credit risks. (3) Appropriately increasing penalties for defaults of financing enterprises and passive supervision of core enterprises will have a normative effect on financing enterprises and core enterprises, and can also prevent credit risks.

Keywords: evolutionary game; supply chain finance; credit risk; new energy vehicles

1. Introduction

In recent years, the country has focused on adjusting the structure of the energy industry and promoting the application of new energy technologies. China's new energy vehicle industry chain has begun to take shape, and new energy vehicle products are flourishing. Due to the high investment in research and development of new energy vehicles and the long return cycle, many enterprises lack fixed asset guarantees and are easily considered high-risk projects in the risk assessment of financial institutions, causing financing difficulties (Xiong & Chen, 2016). In order to fully meet the financing needs of the new energy vehicle industry, supply chain finance is considered a feasible way (Lu & Dong, 2016). With the continuous innovation of supply chain finance business models, the speculative behavior of some financing enterprises has also led to frequent credit risk events in supply chain finance.

Supply chain finance participants mainly include financing enterprises, core enterprises and financial institutions. In the new energy vehicle industry chain, financing enterprises are generally small and medium-sized enterprises (SMEs) in the upstream, midstream and downstream, and core enterprises are generally larger vehicle manufacturing enterprises with large fixed assets (Yang & Xie, 2016). Because of the "financing difficulty" problem of SMEs, they need to rely on the credibility and contractual capacity of the core enterprises to obtain financing. Financial institutions regard the core enterprises and financing enterprises in the supply chain as a whole. The core enterprises provide guarantees for financing enterprises, and the financial institutions provide financial services for financing enterprises. When the supply chain finance business is completed, the core enterprises get the interest paid by SMEs and the subsidies provided by the government departments (Zheng & Zhang, 2020).

SMEs are the main subjects served by supply chain finance. Since SMEs themselves have few information elements and the lack of information resources is very common, it is difficult to use existing credit risk analysis models for assessment. Core enterprises are large in scale and strong in strength, but the risks they bear are limited, and many core enterprises participate in joint counterfeiting. Over-reliance on the credit of core enterprises also has certain risks (Sun et al., 2022). As fund providers, financial institutions need to assess and supervise the credit risks of financing enterprises and core enterprises.

Therefore, this paper uses the evolutionary game theory to focus on the supply chain finance participants of the new energy vehicle, and constructs a three-party game model composed of financing enterprises, core enterprises, and financial institutions, and then reveals the mechanism of the decision-making behavior of the participants on the evolution of credit risk. Different from the existing industry credit risk assessment mainly concentrated on some traditional industries, this paper integrates the risk characteristics of supply chain finance into the credit risk assessment of the new energy vehicles, making the research on supply chain finance of the new energy vehicles more comprehensive. In the establishment of the game model, most researches divide the strategy of the core enterprises into "guarantee" and "no guarantee". If the core enterprises do not provide guarantee, it will violate the core requirements of supply chain finance. Therefore, this paper directly considers the operation process of supply chain finance business when the core enterprises "guarantee" for the financing enterprises, and classifies the strategy of the core enterprises into "active supervision" and "passive supervision". At the same time, considering the collusion behavior of the core enterprise and the financing enterprises, it is more in line with the actual situation of the occurrence of supply chain finance credit risk.

2. Literature Review

2.1. Risks of new energy vehicles supply chain

Some scholars have studied the risks of new energy vehicles from the supply chain perspective. Based on various ways of the supply chain financing of new energy vehicles, Yang & Xie (2016) comprehensively considered the risks brought by each way, and sought an efficient path to solve the financial problem for small and medium-sized enterprises in the supply chain. Yan et al. (2020) conducted risk assessment on China's new energy vehicle supply chain from three dimensions: market risk, operational risk and environmental risk, and the results show that the risk of China's new energy vehicle supply chain was at a relatively high level. Li (2022) analysed the risk sources at various stages of the supply chain of new energy vehicles, such as planning, procurement, production, logistics distribution and returns, and constructed a set of influencing factors. The research shows that inventory control, distribution service responsiveness, transportation network design, insufficient capital or technical level, inventory control risk of order-driven returns, and resource depletion are the key influencing factors of new energy vehicle supply chain risk. Yang et al. (2023) found that the key nodes in the supply chain network of new energy vehicles that have a strong influence on risk communication not only contain core enterprises such as batteries in the traditional sense, but also include parts enterprises with the attribute of hidden champions in the industry.

2.2. Supply chain finance

From the perspective of supply chain finance credit risk, relevant research evaluates credit risk based on different industries, different models and different perspectives, and conducts empirical research with the help of specific models or quantitative methods. Tian & He (2016) studied three supply chain financing models, added core enterprises and other factors under the traditional credit risk evaluation system, and comprehensively assessed the size of the credit risk of financing enterprises. Based on the current realistic trend of supply chain finance development, Fan et al. (2017) effectively prevented and controlled the credit risk of supply chain according to the characteristics of supply chain finance, content and methods of risk management. Based on the credit risk theory of supply chain finance and the industry characteristics of real estate enterprises, Xu & Li (2018) identifies the default risk of real estate enterprises in order to provide new ideas for bank credit financing. Zhang (2020) established a credit risk assessment system from the perspective of sustainable supply chain finance to predict the performance of financing enterprises for financial institutions.

Since the decision-making choices of supply chain finance participants will be affected by the decision-making choices of other participants, and in turn affect the decision-making choices of other participants, and this interaction between participants is dynamic, many scholars choose evolutionary games as an analytical tool to more scientifically interpret the relationship among participants and reveal the dynamic evolution mechanism among them. In order to promote the development of supply chain finance in the cooperation between commercial banks and third-party logistics, Xu & Cong (2022) used evolutionary game to explore the evolutionary stability strategy of cooperation and its influencing factors. Lou et al. (2022) studied the impact of blockchain on the supply chain finance system through the two-way perspective between banks and enterprises, using three-party game and dynamic evolutionary game. Among the current supply chain finance research on specific industries, there are many research contents on agricultural supply chain finance. Li (2023) used the evolutionary game to analyse the influencing factors of the willingness of agricultural supply chain financial participants (farmers, core enterprises, and financial institutions) based on finite rationality. Based on the perspective of agricultural supply chain finance, Li & Qu (2023) constructed a tripartite evolutionary game model of "leading enterprises in agricultural industrialization-new agricultural operating entities-financial institutions" to analyse the strategic choices and influencing factors of financing participants.

3. Problem Description and Basic Assumptions

3.1. Problem description

The sources of credit risk for new energy vehicles under the supply chain finance model mainly include two parts. First, the operating risk of financing enterprises. Since most financing enterprises are SMEs in the new energy vehicle industry chain, they are often short of funds. The initial investment in construction is large but the enterprises' fixed assets are small. During the loan repayment process, defaults may occur due to irregular operations. The second is the principal-agent risk. In order to obtain more profits, the financing enterprises and the core enterprises may reach illegal interest alliance during the operation process, and this collusion behavior will lead to the emergence of credit risk problems.

3.2. Model assumptions and description of parameters

Assumption 1: Financing enterprises, core enterprises and financial institutions in new energy vehicle supply chain finance are the three players in the evolutionary game. The three players are all finite rational, and there is information asymmetry in the game process. The players will adjust their own strategy based on the situation of other players.

Assumption 2: In the evolutionary game process, the behavioral strategy set of financing enterprise is $S_1\{K_1$ keeping promise, K_2 default$\}$. The financing enterprise repays the loan on time as agreed, that is, it chooses the "keeping promise" strategy; after the maturity of the loan, the financing enterprise deliberately fails to repay, that is, it chooses the "default" strategy. The behavioral strategy set of core enterprise is $S_2\{M_1$

active supervision, M_2 passive supervision}. As the assisting supervisor of financial institution, core enterprise actively supervises the business processes of financing enterprise. When the defaults are discovered, financing enterprise will be subject to penalty, that is, it chooses the "active supervision" strategy; when the financing enterprise defaults, the core enterprise enters into a collusive relationship with the financing enterprise in order to profit from it, that is, it chooses the "passive supervision" strategy. The behavioral strategy set of financial institution is S_3{N_1 strict supervision, N_2 lax supervision}. Financial institution strictly supervises the business processes of financing enterprise and punishes financing enterprise when it finds defaults, that is, it chooses the "strict supervision" strategy. Due to the high cost of strict supervision, financial institution adopts lax supervision behaviors in order to save costs during the supervision process, that is, it chooses the "lax supervision" strategy.

Assumption 3: In the initial stage of the game, the probability of the financing enterprise choosing the "keeping promise" strategy is x, and the probability of choosing the "default" strategy is $1-x$; and the probability of the core enterprise choosing the "active supervision" strategy is y, the probability of choosing the " passive supervision" strategy is $1-y$; the probability of financial institution choosing the "strict supervision" strategy is z, the probability of choosing the "lax supervision" strategy is $1-z$. Among them, $0 \le x \le 1$, $0 \le y \le 1$, $0 \le z \le 1$, x,y,z change with the passage of time.

Assumption 4: After obtaining the loan, the financing enterprise operates well and repays the loan on time when due， so as to obtain the profits I_1 of keeping the promise. Financing enterprise can obtain the default profits I_2 by adopting some means to defraud and evade loans, and $I_1 < I_2$. However, once the default behavior is discovered by the core enterprise or financial institution, it will face default penalties F_1 and credit losses R_1.

Assumption 5: Core enterprise assists financial institution in actively supervising financing enterprise, which is conducive to subsequent cooperation with financial institution and can obtain profits P_1. When financing enterprise defaults, in order to evade the supervision of the core enterprise, it will induce the core enterprise to conduct passive supervision with certain profits. The core enterprise and the financing enterprise will enter into a collusive relationship, and core enterprise can obtain profits P_2, and $P_1 < P_2$. However, when this illegal interest alliance is discovered by financial institution, the core enterprise will face penalties F_2 and credit losses R_2.

Assumption 6: When financial institution strictly supervises, the cost is C_1. As the financial institution chooses lax supervision because the supervising cost is too high or the credit status of financing enterprise is good, the supervising cost is C_2, and $C_1 > C_2$. When financing enterprise keeps promise, financial institution can collect loans in a timely manner and obtain revenue I_4. When financing enterprise defaults, if the financial institution strictly supervises it, it can detect the financing enterprise's default in time and ask for the loan, and gain revenue I_5; if the financial institution relaxes supervision, but the core enterprise actively supervises, default can still be discovered in time and financial institution can gain revenue I_5, but financial institution will face penalties F_3 due to lax supervision; if financial institution relaxes supervision and core enterprise passively supervises, the financing enterprise's default is not discovered at this time, and financial institution will face losses L_1.

Assumption 7: If the financing enterprise chooses to "keeping promise" and core enterprise chooses "active supervision", then the supply chain operates well and enterprise in the chain can obtain additional profits I_3.

3.3. Game evolution strategy choice and payoff matrix

According to the behavioral strategies of financing enterprise, core enterprise and financial institution, eight game combinations can be obtained, namely (K_1 keeping promise, M_1 active supervision, N_1 strict supervision), (K_1 keeping promise, M_1 active supervision, N_2 lax supervision), (K_1 keeping promise, M_2 passive supervision, N_1 strict supervision), (K_1 keeping promise, M_2 passive supervision, N_2 lax supervision), (K_2 default, M_1 active supervision, N_1 strict supervision), (K_2 default, M_1 active supervision, N_2 lax supervision), (K_2 default, M_2 passive supervision, N_1 strict supervision), (K_2 default, M_2 passive supervision, N_2 lax supervision). Based on the strategic analysis and parameter definition of each game player, the payoff matrix of the evolutionary game model is obtained, as shown in Table 1 below.

Table 1. Behavioral Strategy Combination and Payoff Matrix of Evolutionary Game

Strategy combination	Financing enterprise	Core corporate	Financial institution
(K_1, M_1, N_1)	$I_1 + I_3$	$P_1 + I_3$	$I_4 - C_1$
(K_1, M_1, N_2)	$I_1 + I_3$	$P_1 + I_3$	$I_4 - C_2$
(K_1, M_2, N_1)	I_1	0	$I_4 - C_1$
(K_1, M_2, N_2)	I_1	0	$I_4 - C_2$
(K_2, M_1, N_1)	$I_2 - F_1 - R_1$	P_1	$I_5 - C_1$
(K_2, M_1, N_2)	$I_2 - F_1 - R_1$	P_1	$I_5 - C_2 - F_3$
(K_2, M_2, N_1)	$I_2 - F_1 - R_1$	$P_2 - F_2 - R_2$	$I_5 - C_1$
(K_2, M_2, N_2)	I_2	P_2	$-C_2 - L_1$

4. Evolutionary Game Analysis

4.1. Construction and solution of replicated dynamic equation

Evolutionary game uses replicated dynamic equations to describe the changing rules of how participants choose strategies in the process of evolution. Its basic form is $\frac{dx}{dt} = x(f(s_i, x) - f(x, x))$, where "$x$" represents the proportion of a strategy "s_i" in the population at time "t", "$f(s_i, x)$" represents the expected revenue when participant chooses "s_i" and "$f(x, x)$" represents the average expected revenue.

Assume that the expected revenue of financing enterprise choosing the "keeping promise" and "default" strategies are, respectively U_x, U_{1-x} and the average expected revenue is $\overline{U}_x$, then:

$$U_x = yz(I_1 + I_3) + y(1 - z)(I_1 + I_3) + (1 - y)zI_1 + (1 - y)(1 - z)I_1 \tag{1}$$

$$\begin{aligned} U_{1-x} = yz(I_2 - F_1 - R_1) + y(1 - z)(I_2 - F_1 - R_1) + (1 - y)z(I_2 - F_1 - R_1) \\ + (1 - y)(1 - z)I_2 \end{aligned} \tag{2}$$

$$\overline{U}_x = xU_x + (1 - x)U_{1-x} \tag{3}$$

Based on the equation (1), (2) and (3), the replicated dynamic equation of financing enterprise is as follows:

$$\begin{aligned} F(x) = \frac{dx}{dt} = x(U_x - \overline{U}_x) = x(1 - x)(U_x - U_{1-x}) \\ = x(1 - x)(I_1 - I_2 + yI_3 + yF_1 + zF_1 + yR_1 + zR_1 - yzF_1 - yzR_1) \end{aligned} \tag{4}$$

Assume that the expected revenue of core enterprise choosing the "active supervision" and "passive supervision" strategies are, respectively U_y, U_{1-y} and the average expected revenue is $\overline{U}_y$, then:

$$U_y = xz(P_1 + I_3) + x(1 - z)(P_1 + I_3) + (1 - x)zP_1 + (1 - x)(1 - z)P_1 \tag{5}$$

$$U_{1-y} = (1 - x)z(P_2 - F_2 - R_2) + (1 - x)(1 - z)P_2 \tag{6}$$

$$\overline{U}_y = yU_y + (1 - y)U_{1-y} \tag{7}$$

Based on the equation (5), (6) and (7), the replicated dynamic equation of the core enterprise is as follows:

$$\begin{aligned} F(y) = \frac{dy}{dt} = y(U_y - \overline{U}_y) = y(1 - y)(U_y - U_{1-y}) \\ = y(1 - y)(P_1 - P_2 + xI_3 + zF_2 + xP_2 + zR_2 - xzF_2 - xzR_2) \end{aligned} \tag{8}$$

Assume that the expected revenue of financial institution choosing the "strict supervision" and "lax supervision" strategies are, respectively U_z, U_{1-z} and the average expected revenue is $\overline{U}_z$, then:

$$U_z = xy(I_4 - C_1) + x(1-y)(I_4 - C_1) + (1-x)y(I_5 - C_1) + (1-x)(1-y)(I_5 - C_1) \tag{9}$$

$$U_{1-z} = xy(I_4 - C_2) + x(1-y)(I_4 - C_2) + (1-x)y(I_5 - C_2 - F_3) + (1-x)(1-y)(-C_2 - L_1) \tag{10}$$

$$\overline{U}_z = zU_z + (1-z)U_{1-z} \tag{11}$$

Based on the equation (9), (10) and (11), the replicated dynamic equation of the financial institution is as follows:

$$\begin{aligned} F(z) &= \frac{dz}{dt} = z(U_z - \overline{U}_z) = z(1-z)(U_z - U_{1-z}) \\ &= z(1-z)(-C_1 + C_2 + I_5 + L_1 - xI_5 - yI_5 + yF_3 - xL_1 - yL_1 + xyI_5 - xyF_3 + xyL_1) \end{aligned} \tag{12}$$

Let $F(x) = 0$, $F(y) = 0$, $F(z) = 0$, we can get the equilibrium point and Jacobi matrix of the replicated dynamic system. The 14 equilibrium points are respectively $E_1(0,0,0)$, $E_2(0,1,0)$, $E_3(0,0,1)$, $E_4(0,1,1)$, $E_5(1,0,0)$, $E_6(1,1,0)$, $E_7(1,0,1)$, $E_8(1,1,1)$, $E_9((C_2 - C_1 + I_5 + L_1)/(I_5 + L_1), 0, (I_2 - I_1)/(F_1 + R_1))$, $E_{10}(0, (C_2 - C_1 + I_5 + L_1)/(I_5 - F_3 + L_1), (P_2 - P_1)/(F_2 + R_2))$, $E_{11}((-F_2 - P_1 + P_2 - R_2)/(I_3 - F_2 + P_2 - R_2), (-I_1 + I_2 - F_1 - R_1)/I_3, 1)$, $E_{12}((P_2 - P_1)/(I_3 + P_2), (I_2 - I_1)/(I_3 + F_1 + R_1), 0)$, $E_{13}(x_1^*, y_1^*, z_1^*)$, $E_{14}(x_2^*, y_2^*, z_2^*)$. (x_1^*, y_1^*, z_1^*) and (x_2^*, y_2^*, z_2^*) are the solution of equations set (13), which is not listed here due to its complexity.

$$\begin{cases} I_1 - I_2 + yI_3 + yF_1 + zF_1 + yR_1 + zR_1 - yzF_1 - yzR_1 = 0 \\ P_1 - P_2 + xI_3 + zF_2 + xP_2 + zR_2 - xzF_2 - xzR_2 = 0 \\ -C_1 + C_2 + I_5 + L_1 - xI_5 - yI_5 + yF_3 - xL_1 - yL_1 + xyI_5 - xyF_3 + xyL_1 = 0 \end{cases} \tag{13}$$

In asymmetric games, if the equilibrium of the evolutionary game is an evolutionary stability strategy, it must be a strict Nash equilibrium, and a strict Nash equilibrium is a pure strategy equilibrium. In other words, the mixed strategy equilibrium in asymmetric games must not be evolutionary stability strategy (Selten R., 1980; Wang et al., 2022). Since $E_9 \sim E_{14}$ are not strict Nash equilibrium solutions, only $E_1 \sim E_8$ will be analysed below.

4.2. Evolutionary equilibrium point analysis

According to Friedman's research, the local stability of the equilibrium point $E_1 \sim E_8$ is analysed based on Jacobi matrix eigenvalues and theory of Lyapunov asymptotic stability. The form of the Jacobi matrix obtained by replicated dynamic equations is:

$$J = \begin{bmatrix} \frac{\partial F(x)}{\partial x} & \frac{\partial F(x)}{\partial y} & \frac{\partial F(x)}{\partial z} \\ \frac{\partial F(y)}{\partial x} & \frac{\partial F(y)}{\partial y} & \frac{\partial F(y)}{\partial z} \\ \frac{\partial F(z)}{\partial x} & \frac{\partial F(z)}{\partial y} & \frac{\partial F(z)}{\partial z} \end{bmatrix}$$

Taking the equilibrium point $E_1(0,0,0)$ as an example, the Jacobi matrix J_1 of the E_1 is obtained by substituting (0,0,0) into the formula:

$$J_1 = \begin{bmatrix} I_1 - I_2 & 0 & 0 \\ 0 & P_1 - P_2 & 0 \\ 0 & 0 & C_2 - C_1 + I_5 + L_1 \end{bmatrix}$$

The eigenvalues of J_1 that can be obtained are $\lambda_1 = I_1 - I_2$, $\lambda_2 = P_1 - P_2$, $\lambda_3 = C_2 - C_1 + I_5 + L_1$. In the same way, the remaining seven equilibrium points of the replicated dynamic system are substituted into the Jacobi matrix formula, and the Jacobi matrix eigenvalues corresponding to the equilibrium points are obtained as shown in Table 2 below.

Table 2. Equilibrium Point and its Eigenvalues

Equilibrium point	Eigenvalue λ_1	Eigenvalue λ_2	Eigenvalue λ_3
$E_1(0,0,0)$	$I_1 - I_2$	$P_1 - P_2$	$C_2 - C_1 + I_5 + L_1$
$E_2(0,1,0)$	$P_2 - P_1$	$C_2 - C_1 + F_3$	$I_1 - I_2 + I_3 + F_1 + R_1$
$E_3(0,0,1)$	$C_1 - C_2 - I_5 - L_1$	$I_1 - I_2 + F_1 + R_1$	$F_2 + P_1 - P_2 + R_2$
$E_4(0,1,1)$	$C_1 - C_2 - F_3$	$P_2 - P_1 - F_2 - R_2$	$I_1 - I_2 + I_3 + F_1 + R_1$
$E_5(1,0,0)$	$P_1 + I_3$	$C_2 - C_1$	$I_2 - I_1$
$E_6(1,1,0)$	$C_2 - C_1$	$-I_3 - P_1$	$I_2 - I_1 - I_3 - F_1 - R_1$
$E_7(1,0,1)$	$P_1 + I_3$	$C_1 - C_2$	$I_2 - I_1 - F_1 - R_1$
$E_8(1,1,1)$	$C_1 - C_2$	$-I_3 - P_1$	$I_2 - I_1 - I_3 - F_1 - R_1$

According to the theory of Lyapunov asymptotic stability: if all eigenvalues of the Jacobi matrix are negative, then the equilibrium point is evolutionary stability strategy (ESS); if one eigenvalue of the Jacobi matrix is non-negative, the equilibrium point is an unstable point. The positive and negative of the eigenvalues are determined, and the stability of each equilibrium point is analysed. If the positive or negative of the eigenvalue is uncertain, it is represented by the symbol "×", and the local stability of the equilibrium point is shown in Table 3 below.

Table 3. Local Stability Analysis of Equilibrium Point

Equilibrium point	Eigenvalue symbol	Stability conditions	Evolutionary results
$E_1(0,0,0)$	(-, -, ×)	①	ESS
$E_2(0,1,0)$	(+,×,×)	\	Unstable point
$E_3(0,0,1)$	(×,×,×)	②	ESS
$E_4(0,1,1)$	(×,×,×)	③	ESS
$E_5(1,0,0)$	(+,-, +)	\	Unstable point
$E_6(1,1,0)$	(-,-, ×)	④	ESS
$E_7(1,0,1)$	(+,+,×)	\	Unstable point
$E_8(1,1,1)$	(+,-,×)	\	Unstable point

It can be seen from the table that the equilibrium points E_2, E_5, E_7, E_8 all have non-negative eigenvalues, so they are unstable points. The equilibrium point E_1, E_3, E_4, E_6 are ESS when the parameters meet specific conditions. Next, the stability of the equilibrium point E_1, E_3, E_4, E_6 is analysed.

When condition ①$C_1 - I_5 > C_2 + L_1$ is met, the equilibrium point $E_1(0,0,0)$is the ESS. At this time, financing enterprise chooses "default", core enterprise chooses " passive supervision", and financial institution chooses "lax supervision". That is, when the cost of strict supervision of financial institution minus the financial institution's revenue if the financing enterprise's default is discovered is greater than the sum of the financial institution's cost of lax supervision and the financial institution's losses if the financing enterprise's default is not discovered, the price of strict supervision of financial institution is higher than that of lax supervision. Therefore, financial institution ultimately tends to choose "lax supervision".

When condition ②$C_1 - I_5 < C_2 + L_1$, $I_1 < I_2 - F_1 - R_1$, $P_1 < P_2 - F_2 - R_2$ is met, the equilibrium point $E_3(0,0,1)$ is the ESS. At this time, financing enterprise chooses "default", core enterprise chooses "passive supervision", and financial institution chooses "strict supervision". That is, when the cost of strict supervision of financial institution minus the financial institution's revenue if the financing enterprise's default is discovered is less than the sum of the financial institution's cost of lax supervision and the financial institution's losses if the financing enterprise's default is not discovered, the price of lax supervision paid by the financial institution is higher than that of strict supervision. Therefore, financial institution ultimately tends to choose "strict supervision". When the financing enterprise's profits from keeping promise is less than the financing enterprise's profits from default minus default penalties and reputation losses, the financing enterprise is more profitable if it chooses to default, so the financing enterprise will eventually choose "default". When core enterprise's profits from active supervision are less than core enterprise's profits from passive supervision deducting penalties for passive supervision and reputation losses from passive supervision, core enterprise is more profitable if it chooses to passively supervise. Therefore, core enterprise eventually tends to be "passive supervision".

When conditions③$C_1 < C_2 + F_3$, $P_1 > P_2 - F_2 - R_2$, $I_2 - F_1 - R_1 > I_1 + I_3$ is met, the equilibrium point $E_4(0,1,1)$ is the ESS. At this time, financing enterprise chooses "default", core enterprise chooses "active supervision", and financial institution chooses "strict supervision". That is, when the cost of strict supervision of financial institution is less than the sum of the cost and penalties of lax supervision of financial institution, the price of lax supervision of financial institution is higher than the cost of strict supervision, so financial institution ultimately tends to choose "strict supervision". When the profits of core enterprise from active supervision are greater than the profits of core enterprise from passive supervision deducting the penalties for passive supervision and reputation losses from passive supervision, core enterprise is more profitable if it chooses to actively supervise. Therefore, core enterprise eventually tends to be "active supervision". When the financing enterprise's default profits minus default penalties and reputation losses is greater than the financing enterprise's profits from keeping promise and the additional profits from the good operation of the supply chain, the financing enterprise is more profitable if it chooses to default, so the financing enterprise will eventually choose to "default".

When condition ④$I_2 - F_1 - R_1 < I_1 + I_3$ is met, the equilibrium point $E_6(1,1,0)$ is the ESS. At this time, financing enterprise chooses to "keeping promise", core enterprise chooses "active supervision", and financial institution chooses "lax supervision". That is, when the financing enterprise's default profits minus the default penalty and reputation losses is less than the financing enterprise's profits from keeping promise and the additional profits from the good operation of the supply chain, the financing enterprise is more profitable if it chooses to keep promise, so the financing enterprise will eventually choose to "keep the promise".

Through the above analysis, it can be found that changes in parameters have a significant impact on system stability.

5. Numerical Simulation Analysis

In order to verify the effectiveness of evolutionary stability analysis and explore the impact of parameter changes on the risk evolution mechanism of the game system, this paper assigns numerical values to the model based on the actual situation and uses MATLAB for numerical simulation.

5.1. Numerical simulation of stability analysis

Assume that the initial values of the parameters are $I_1 = 30$, $I_2 = 150$, $I_3 = 5$, $I_4 = 120$, $I_5 = 40$, $C_1 = 100$, $C_2 = 8$, $F_1 = 100$, $F_2 = 45$, $F_3 = 50$, $L_1 = 50$, $P_1 = 10$, $P_2 = 40$, $R_1 = 5$, $R_2 = 5$, array 1 satisfies $C_1 - I_5 > C_2 + L_1$. The simulation results of replicated dynamic equation set evolving 50 times over time are shown in Figure 1 (a).

Assume that the initial values of the parameters are $I_1 = 30$, $I_2 = 100$, $I_3 = 5$, $I_4 = 60$, $I_5 = 40$, $C_1 = 30$, $C_2 = 8$, $F_1 = 60$, $F_2 = 20$, $F_3 = 50$, $L_1 = 50$, $P_1 = 10$, $P_2 = 40$, $R_1 = 5$, $R_2 = 5$, array 2 satisfies $C_1 - I_5 < C_2 + L_1$, $I_1 < I_2 - F_1 - R_1$ and $P_1 < P_2 - F_2 - R_2$. The simulation results of replicated dynamic equation set evolving 50 times over time are shown in Figure 1 (b).

Assume that the initial values of the parameters are $I_1 = 30$, $I_2 = 150$, $I_3 = 5$, $I_4 = 120$, $I_5 = 40$, $C_1 = 100$, $C_2 = 8$, $F_1 = 100$, $F_2 = 45$, $F_3 = 50$, $L_1 = 50$, $P_1 = 10$, $P_2 = 40$, $R_1 = 5$, $R_2 = 5$, array 3 satisfies $C_1 < C_2 + F_3$, $P_1 > P_2 - F_2 - R_2$ and $I_2 - F_1 - R_1 > I_1 + I_3$. The simulation results of replicated dynamic equation set evolving 50 times over time are shown in Figure 1 (c).

Assume that the initial values of the parameters are $I_1 = 30$, $I_2 = 100$, $I_3 = 5$, $I_4 = 60$, $I_5 = 40$, $C_1 = 30$, $C_2 = 8$, $F_1 = 150$, $F_2 = 45$, $F_3 = 50$, $L_1 = 50$, $P_1 = 10$, $P_2 = 40$, $R_1 = 5$, $R_2 = 5$, array 4 satisfies $I_2 - F_1 - R_1 < I_1 + I_3$. The simulation results of replicated dynamic equation set evolving 50 times over time are shown in Figure 1 (d).

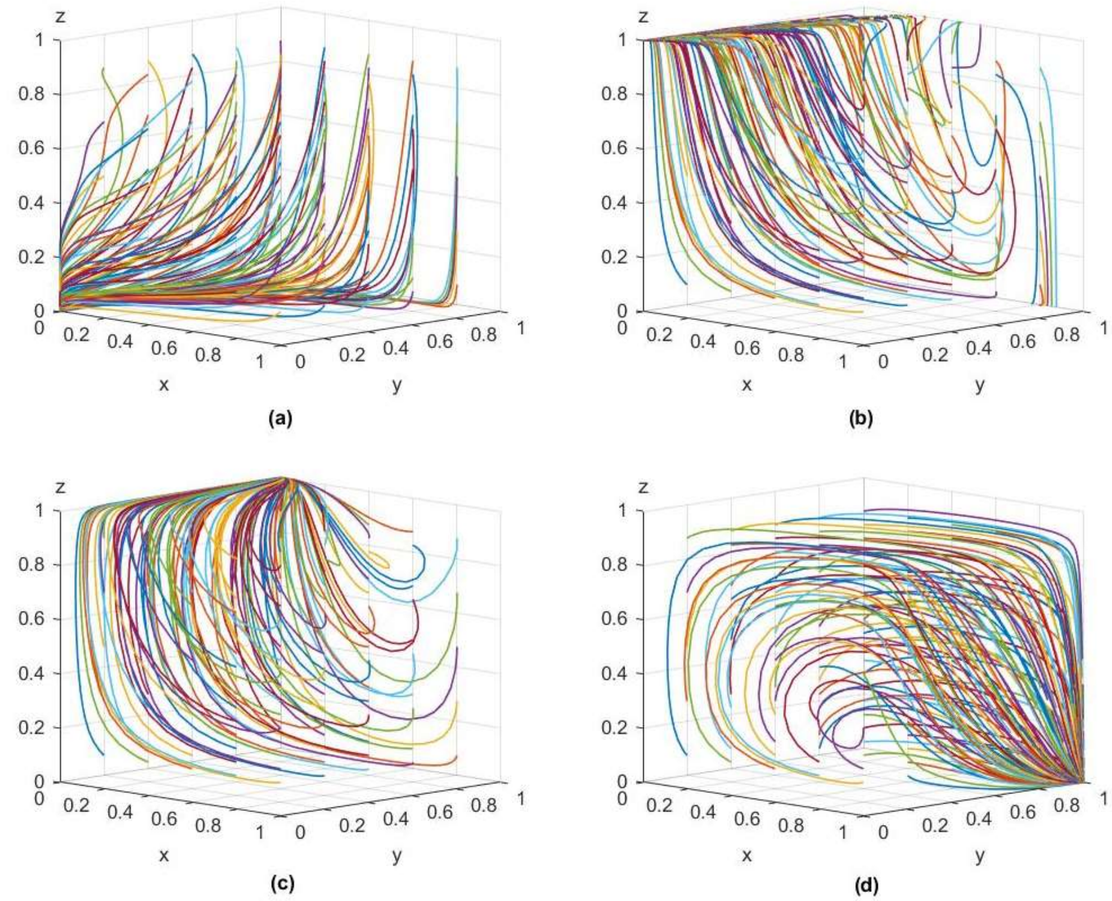


Figure 1: Evolution Results of Arrays 1 to 4

By simulating four sets of values, it can be seen that array 1 converges to (0,0,0), array 2 converges to (0,0,1), array 3 converges to (0,1,1), and array 4 converges to (1,1,0), consistent with the previous stability analysis results, that is, when certain conditions are met, the evolutionary stability strategies are (default, passive supervision, lax supervision), (default, passive supervision, strict supervision), (default, active supervision, strict supervision) and (keeping promise, active supervision, lax supervision) respectively.

5.2. Impact of initial probability on the strategy evolution of players

The initial probability of financing enterprise to "keeping promise", the initial probability of core enterprise to "active supervision" and the initial probability of financial institution to "strict supervision" are all set to 0.2, 0.5 and 0.7, and array 4 is used as the initial value of the parameter to simulate and analyse the evolution process of the behavior strategy of the three players. The evolution results are shown in Figure 2.

The analysis of Figure 2 shows that in this supply chain finance model, no matter the initial probability is 0.2, 0.5 or 0.7, financing enterprise will choose the "keeping promise" strategy and have high enthusiasm, and core enterprise will choose "active supervision". When the initial probability is 0.2, financial institution will tend to "strict supervision" for a period of time. This is due to losses caused by financial institution to prevent defaults of financing enterprise and penalties brought about by lax supervision. However, as time evolves, financing enterprise tends to "keeping promise", which in turn affects financial institution, and financial institution eventually tends to "lax supervision" in order to save costs. In addition, it can be seen that as the initial probability ratio increases, the game system evolves to a stable strategy faster.

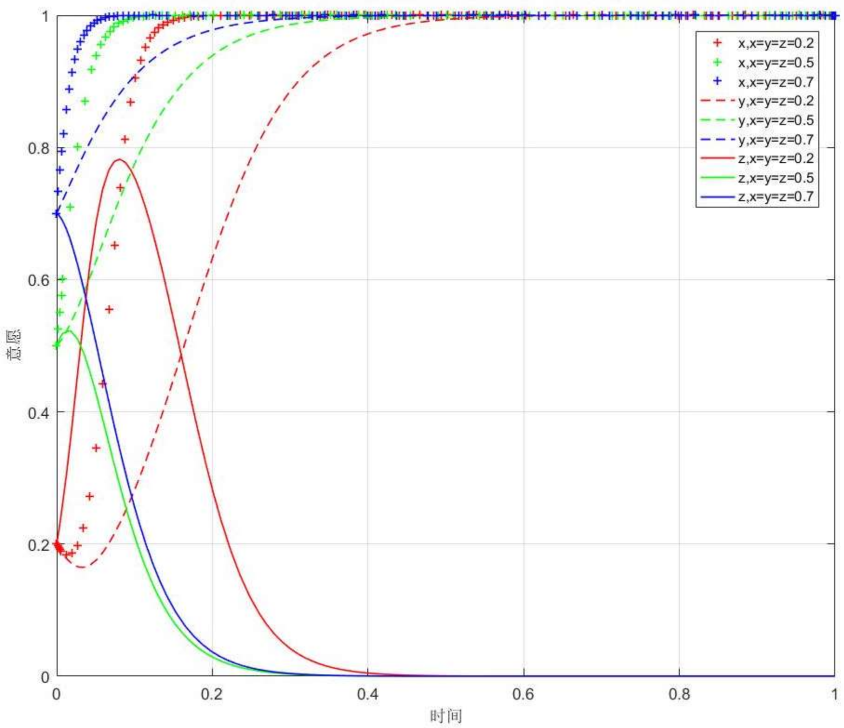


Figure 2: Evolution Strategies with Different Initial Probability

5.3. Impact of different parameters on the strategy evolution of players

Through the stability analysis of the system, it can be found that changes in the value of key parameters will have an impact on the choice of the main strategy. Therefore, the parameters are divided into three categories for analysis, including revenue parameters I_1, I_3, I_5, P_1, cost parameters C_1, C_2, and penalty parameters F_1, F_2, F_3. Still use array 4 as the initial value of the parameters, change one of the parameters to study the impact of parameter changes on the credit risk evolution mechanism of the game system.

- Revenue parameters

I_1 is assigned to 0, 30 and 60 respectively, and the simulation results of the replicated dynamic equation set evolving 50 times with the probability of (0.2, 0.2, 0.2) are shown in Figure 3; I_3 is assigned to 0, 5 and 10 respectively, and the simulation results are shown in Figure 4; I_5 is assigned to 0, 40 and 80 respectively, and the simulation results are shown in Figure 5; P_1 is assigned to 5, 10 and 15 respectively, and the simulation results are shown in Figure 6.

Analysing Figure 3 shows that as financing enterprise increase its promise-keeping profits I_1, the evolution speed of financing enterprise's "keeping promise" strategy accelerates, the probability of "strict supervision" of financial institution decreases, and the probability of "active supervision" of core enterprise decreases. Therefore, an appropriate increase in the promise-keeping profits of financing enterprises will help financing enterprises reduce default risks, thereby preventing credit risks, while saving supervision costs for financial institutions and core enterprises. Analysing Figure 4, 5, and 6, we can see that with the increase in the additional profits I_3 from the good operation of the supply chain, the financial institution's revenue I_5 if the default of the financing enterprise is discovered, and the core enterprise's profits P_1 by active supervision, the probability of "active supervision" of core enterprise increases, and the "strict supervision" of financial institution increases. Therefore, increasing the additional profits from the good operation of the supply chain, the revenue of financial institutions when defaults of financing enterprises are discovered, and the profits from active supervision of core enterprises will help mobilize the supervision enthusiasm of financial institutions and core enterprises, thereby reducing credit risks.

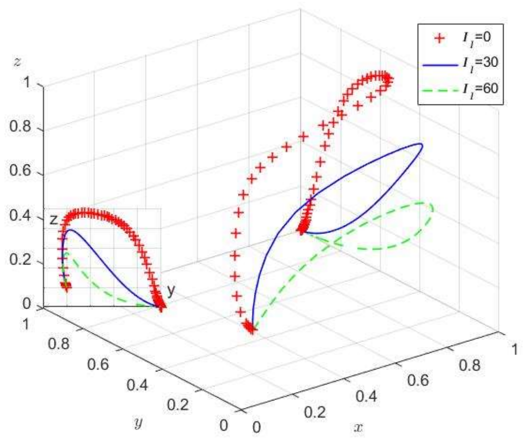


Figure 3: Evolution paths when I_1 is changed

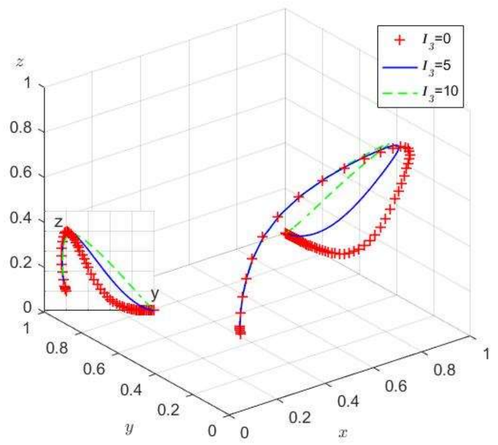


Figure 4: Evolution paths when I_3 is changed

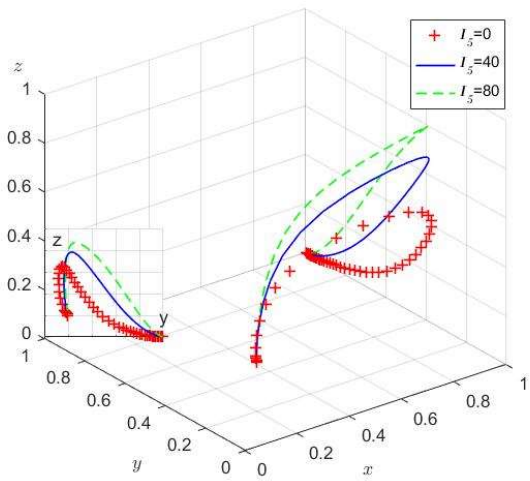


Figure 5: Evolution paths when I_5 is changed

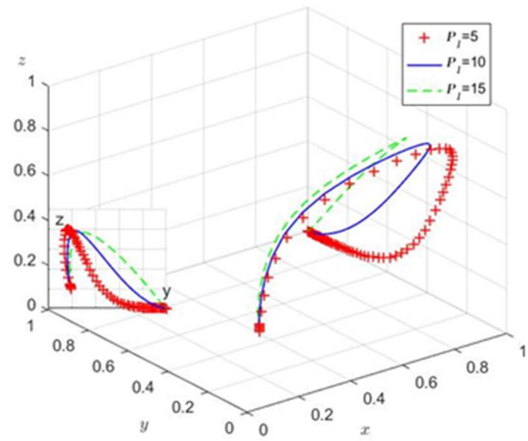


Figure 6: Evolution paths when P_1 is changed

- Cost parameters

C_1 is assigned to 10, 30 and 50 respectively, and the simulation results of the replicated dynamic equation set evolving 50 times with the probability of (0.2, 0.2, 0.2) are shown in Figure 7; C_2 is assigned to 2, 8 and 14 respectively, and the simulation results are shown in Figure 8.

Analysing Figure 7 shows that as the cost C_1 of strict supervision of financial institution increases, financial institution can save more costs by lax supervision. Therefore, the evolution speed of financial institution's "lax supervision" strategy accelerates. At this time, the probability of "strict supervision" of financial institution decreases, and the probability of financing enterprise "keeping promises" also decreases, which is not conducive to the prevention of credit risks. Analysing Figure 8 shows that as the cost C_2 of lax supervision of financial institution increases, the probability of "strict supervision" of financial institution increases, and the probability of "active supervision" of core enterprise increases. Therefore, we can reduce the cost of strict supervision of financial institutions and increase the cost of lax supervision to better deal with the occurrence of credit risks. On the one hand, it helps financial institutions to supervise the credit risks of financing enterprises, and on the other hand, it can also promote the financing enterprises to keep their promises and the core enterprises to actively supervise.

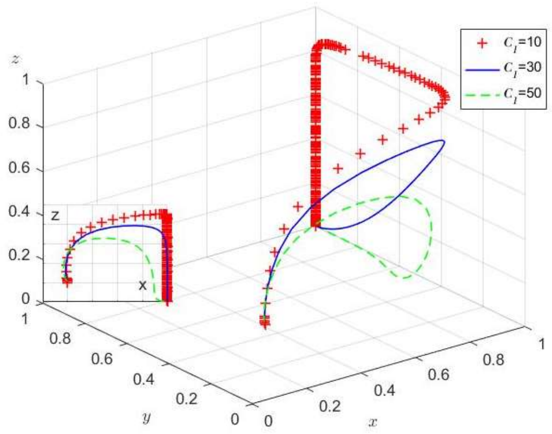


Figure 7: Evolution paths when C_1 is changed

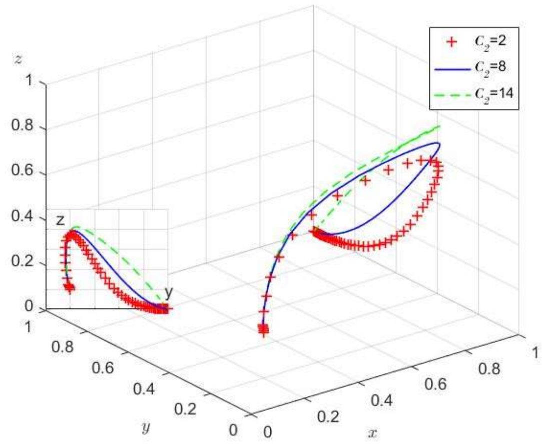


Figure 8: Evolution paths when C_2 is changed

- Penalty parameters

F_1 is assigned to 100, 150 and 200 respectively, and the simulation results of the replicated dynamic equation set evolving 50 times with the probability of (0.2, 0.2, 0.2) are shown in Figure 9; F_2 is assigned to 0, 45 and 90 respectively, and the simulation results are shown in Figure 10; F_3 is assigned to 0, 50 and 100 respectively, and the simulation results are shown in Figure 11.

Analysing Figure 9 shows that as the default penalties F_1 of financing enterprise increase, the evolution speed of financing enterprise's "keeping promise" strategy accelerates. At this time, the probability of "active supervision" of core enterprise decreases, and the probability of "strict supervision" of financial institution also decreases. It can be seen that increasing default penalties for financing enterprises can restrain the occurrence of defaults. At this time, financial institutions can also appropriately relax supervision and reduce economic costs. Analysing Figure 10, we can see that before the evolution of the core enterprise's "keeping promise" strategy stabilizes at 1, with the increase of passive supervision penalties F_2 of core enterprise, the probability of "strict supervision" of financial institution increase. When the core enterprise's "keeping promise" strategy stabilizes at 1, the probability of "strict supervision" of financial institution decreases, and the increase of F_2 will increase the probability of "active supervision" of core enterprise. It can be seen that increasing the penalties for passive supervision of core enterprises will help prevent the occurrence of credit risk problems. Analysing Figure 11 shows that as financial institution's lax regulatory penalties F_3 increases, the probability of "strict supervision" of financial institution increases, and the probability of "active supervision" of core enterprise also increases, but the overall impact is not significant. Therefore, the occurrence of credit risks can be prevented by appropriately increasing the penalties for defaults of financing enterprises and passive supervision of core enterprises. On the one hand, this can promote financing enterprises to keep their promises and core enterprises to actively supervise, and on the other hand, it can also reduce the cost of supervision by financial institutions.

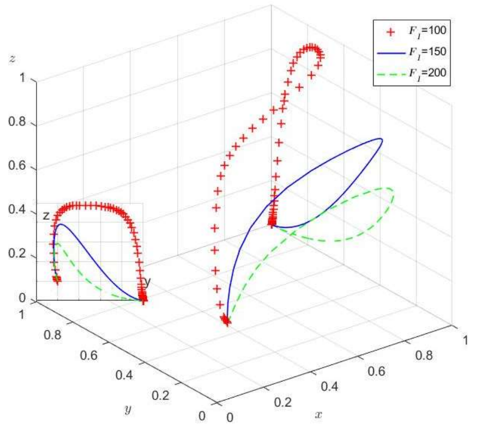


Figure 9: Evolution paths when F_1 is changed

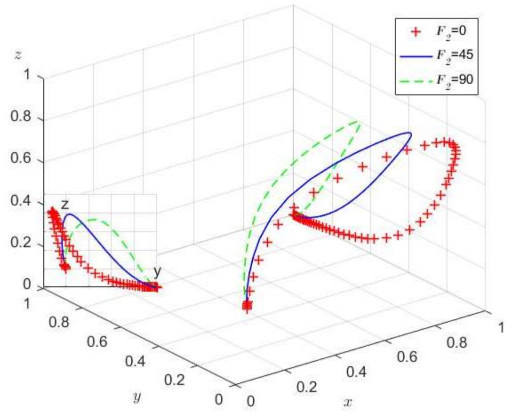


Figure 10: Evolution paths when F_2 is changed

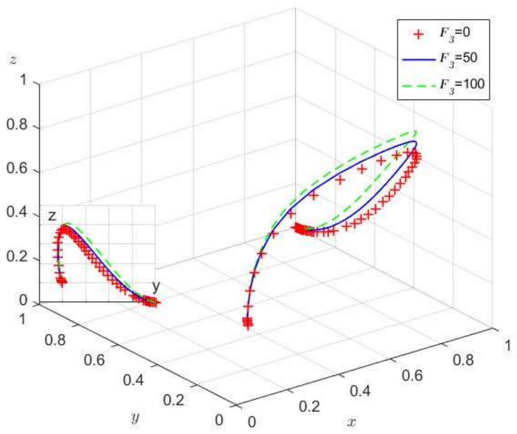


Figure 11: Evolution paths when F_3 is changed

6. Conclusion

Aiming at the credit risk problem existing in new energy vehicle supply chain finance, this paper constructs a tripartite evolutionary game model and its replicated dynamic equation of financing enterprises, core enterprises and financial institutions, solves 14 equilibrium points and analyses the local stability of 8 pure strategy equilibrium points. Finally, the evolution process of financing enterprises, core enterprises and financial institutions is numerically simulated to verify the validity of stability analysis, analyse the impact of parameter changes on the evolution results, and put forward suggestions to prevent credit risks.

The analysis found that (1) Increasing the promise-keeping profits of financing enterprises, additional profits from the good operation of the supply chain, the revenue of financial institutions when defaults of financing enterprises are discovered, and the profits from active supervision of core enterprises are all conducive to controlling credit risks. In addition, increasing the promise-keeping profits of financing enterprises is not only beneficial to control the default risk of financing enterprises, but can also save the supervision costs of financial institutions and core enterprises. (2) Reducing the cost of strict supervision of financial institutions and increasing the cost of lax supervision can better deal with the occurrence of credit risks. On the one hand, it helps financial institutions to supervise the credit risks of financing enterprises, and on the other hand, it can also promote the financing enterprises to keep their promises and the core enterprises to actively supervise. (3) Appropriately increasing penalties for defaults of financing enterprises and passive supervision of core enterprises will have a normative effect on financing enterprises and core enterprises, and can also prevent credit risks.

Based on the research in this paper, the following suggestions are given: (1) Strengthen the business audit and risk assessment of financing enterprises and core enterprises to prevent the occurrence of credit risk at the source. Investigate the transactions between financing enterprises and core enterprises to prevent the occurrence of collusion. (2) Combine blockchain, Internet of Things, big data, artificial intelligence and other financial technologies with the means of risk supervision of financial institutions, and dynamically and timely track the operating conditions of financing enterprises. It can not only improve the efficiency and accuracy of financial institutions' supervision, but also reduce the supervision cost. (3) Financial institutions shall improve credit reward and punishment mechanisms. Provide certain credit concessions to enterprises with good credit to encourage enterprises to keep promises. At the same time, they should appropriately increase the penalties for defaulting enterprises, publicise the information of credit-defaulting enterprises and restrict their financing activities.

Finally, the limitations of this paper and future research directions are as follows: (1) The game model constructed in this paper mainly considers financing enterprises, core enterprises and financing enterprises. In the subsequent research, third-party logistics enterprises can be added to the game process, and the game subjects can be further refined. (2) In the simulation of the game model of the participants in the supply chain finance of new energy vehicles, only the influence of the change of a single factor on the evolution path of credit risk is studied in this paper, and the comprehensive influence of the change of multiple factors on the evolution result of credit risk can be discussed in the subsequent research.

References

[1] Fan Fangzhi, Su Guoqiang, Wang Xiaoyan (2017). Study on assessment and management of SMEs' credit risk under the model of supply chain finance. *Journal of Central University of Finance & Economics*,2017(12):34-43.

[2] Jian Ni, Jun Zhao, Lap Keung Chu (2021). Supply contracting and process innovation in a dynamic supply chain with information asymmetry. *European Journal of OperationalResearch*,288(2):552-562.

[3] Li Da (2023). Agricultural supply chain financial subject strategy evolutionary game coordination analysis. *Science Technology and Industry*, 23(13): 148-156.

[4] Li Meirao, Qu Lili (2023). Research on the evolutionary game of the financing constraint relief mechanism of new agricultural operating entities — a discussion based on the perspective of agricultural supply chain finance. *Financial Theory & Practice*, 2023(09):59-73.

[5] Li Xiaoxiao (2022). Identification of key factors influencing new energy vehicle supply chain risks. *Logistics Sci-Tech*, 45(15):119-123.

[6] Lou Yong, Chang Yuxing, Hao Fengxia (2022). The impact of blockchain on supply chain finance — based on tripartite game theory and dynamic evolutionary game theory. *Chinese Journal of Management Science*, 30(12): 352-360.

[7] Lu Huifang, Dong Guoshu (2016). Discussion on supply chain financing model of new energy vehicles in China. *Shanghai Finance*, (08):81-84.

[8] Paul K.M., Linda B., Jon H. (2021). Blockchain and its impacts on agri-food supply chain network management. *Sustainability*, 13(4):2168.

[9] Selten R. (1980). A note on evolutionarily stable strategies in asymmetric animal conflicts. *Journal of Theoretical Biology*,84(1):93-101.

[10] Sun Rui, He Dayi, Su Huilin (2022). Application of blockchain technology to preventing supply chain finance based on evolutionary game. *Chinese Journal of Management Science*, 1-18

[11] Tian Meiyu, He Wenyu (2016). Small and medium-sized enterprise credit risk assessment in the SCF financing mode — empirical research in auto industry as an example. *Journal of Industrial Technology and Economy*, 35(06):154-160.

[12] Wang Xianjia, Gu Cuiling, He Qilong (2022). Evolutionary game analysis on credit market of supply chain finance. *Operations Research and Management Science*, 31(01):30-37.

[13] Xiong Yongqing, Chen Manlin (2016). Study of the new energy vehicles demand market's cultivated policy orientation: supply side or demand side. *China Population, Resources and Environment*,26(05):129-137.

[14] Xu Lingling, Cong Xuewei (2022). Research on the evolutionary game of supply chain finance in cooperation between commercial banks and third-party logistics. *Financial Theory & Practice*, (04):10-18.

[15] Xu Yongge, Li Ran (2018). Credit risk of real estate enterprises under the perspective of supply chain finance. *Finance and Accounting Monthly*,2018(08):164-169.

[16] Yan Q, Zhang M, Li W, et al. (2020). Risk assessment of new energy vehicle supply chain based on variable weight theory and cloud model: a case study in China. Sustainability, 12(8).

[17] Yang Guang, Xie Jiaping (2016). Research on new energy vehicle supply chain finance model. *Modernization of Management*,36(02):21-23.

[18] Yang Xiaobo, Gao Haiwei, Liu Tianyue, et al (2023). Key risk node identification methods in new energy vehicle supply chain. *Computer Science*, 50(S1):846-852.

[19] Zhang Jiaxin (2020). Credit risk evaluation model for small and medium-sized enterprises under supply chain finance model. *China Market*,2020(16):188-189+193.

[20] Zhang Yu (2018). The financial risk management and inquiry of China's banking supply chain — based on the game theory. *Journal of Technical Economics &Management*, (1):72-76.

[21] Zheng Yu, Zhang Kaixi (2020). Research on the risk management of the supply chain finance — from the perspective of SME financing. *Journal of Financial Development Research*,2020(10):45-51.

Evaluation of NEV market acceptance based on descriptive information and comparative information

Rui Liu[1], Haiyan Xu[1*], Dequn Zhou[1], Limei Hu[1]

[1] College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing, China
[*] corresponding.author@mail.com

Abstract

With advances in information and communication technologies (ICT), customers increasingly share their comments about new electric vehicles (NEVs) on various professional websites. By mining these consumer comments, decision makers (DMs) can analyze consumer preferences and evaluate the market acceptability of NEVs to optimize the layout of NEV products. Although the traditional multi-attribute decision-making (MADM) method can provide relevant technical support, it is still challenged by two aspects in the context of big data: (1) Although studies of existing sentiment analysis (SA) on NEVs offer an insight into sentiment polarity, it does not consider the sentiment analysis results of different attributes of new energy vehicles containing comparative relations and (2) Various types of uncertain data need to be effectively processed in the evaluation process. To address these inadequacies, we provide an integrated framework for the problem of NEV market acceptance evaluation using the evidential reasoning -PageRank (ER-PR) approach. Firstly, SA is performed on the descriptive and comparative information of NEVs, and the sentiment values of different sentiment intensity levels are converted into the belief structure of the NEVs' market acceptability. On this basis, the integrated description information and comparison information of the directed weighted graph of NEVs are constructed. Then, based on the evaluation framework of ER-PR, the sorting value solution algorithm of the market acceptance of NEVs is given. Finally, real data extracted from Autohome websites are used as experimental data to illustrate the implementation of the proposed method and to demonstrate the effectiveness of our approach.

Keywords: descriptive information; comparative information; sentiment analysis; graph model; ER-PR method

1. Introduction

With the growth of transportation tools (such as fuel vehicles), the problem of pollutant emissions is becoming increasingly serious and it severely threatens the health and harmonious development of people and environment (He et al., 2017). Developing and cultivating the new electric vehicle (NEV) industry is an important strategic measure to cope with climate change and conserve fossil energy (He & Wang, 2023). Many

countries have formulated various action plans to substantially expand the NEV market (Sun et al., 2019). For example, China has actively introduced a series of fiscal and tax policies to increase consumer purchase rates (Wang et al., 2021). NEV sales in China cultivated 5.365 million units in 2022, an increase of 81.6% year-on-year (CPCA, 2023). European countries have increased subsidies for NEVs, thus boosting their sales volume to over 1.1 million in 2022, up 28% from 2021 (IEA, 2023). In 2022, the United States proposed the Inflation Reduction Act, which explicitly supports the development of electric vehicles in the next decade. Despite a general contraction in the US auto market that year, sales of NEVs rose by two-thirds to reach 807,180 units (ANL, 2023). The promising market prospects will attract many new manufacturers to start producing NEVs. At the same time, foreign-funded enterprises are pouring into the Chinese market. For example, the production capacity of the gigafactory in Shanghai reached over 750,000 units of pure NEV Model Y in 2022 (Tesla, 2023). The NEVs market has huge growth potential, and market demand is an important driving factor for the development of China's NEV industry. Therefore, in order to achieve the industry target of a 25% market penetration of NEV by 2025, it must be achieved through market-oriented driving.

The NEV industry in China has developed rapidly in recent years, but it is still in the growth stage, and there are still many problems and challenges in the development. Firstly, although the NEV industry in China has experienced a high level of prosperity in the past few years, it still relies on government support. On the one hand, the purchase price of consumers is a subsidy of tens of thousands of yuan given by the government, which makes up for the price disadvantage of NEV compared with fuel vehicles. On the other hand, enterprises on the industrial chain settle in the park and enjoy a large amount of government land and tax subsidies. But in the future, subsidies will be reduced and cancelled, and consumers may not buy NEV because of the reduction in price subsidies. Secondly, there are bottlenecks in the development of current industrial technology, and there is still a gap in core technology, such as the high energy density and safety of power batteries, which have not been balanced and solved. The inadequate infrastructure, imperfect market supervision system, single marketing channels, and low consumer acceptance have also had a negative impact on increasing the market penetration rate of new energy vehicles. Therefore, as the NEV market competition intensifies, the development of NEV has shifted from policy driven to market driven, quick occupation of the market and improving consumers' willingness to buy products are important topics for decision makers (DMs). Market acceptability refers to the degree to which a product is accepted by consumers in a completely competitive market (Du et al., 2018), and it reflects the matching degree of various product indicators from the market with the diversified needs of consumers. At present, many DMs try to accurately grasp consumer preferences and personal demands to purchase NEVs through the assessment of market acceptability, and then optimize the product to achieve sales growth.

The development and popularity of mobile Internet has reached an unprecedented scale. Billions of users can obtain information, exchange information, express their opinions and share their experiences, and express various emotions and sentiments, such as criticism and praise, on the Internet (Kumar & Alok, 2020). Domestic professional automotive platforms such as AutoZone have accumulated many real user reviews. Some studies focus on mining customer's sentiment on a certain attribute from text reviews, whereas other studies directly mine comparative sentences and relationships from text reviews (Chen & Li, 2023). We divide the above two categories of heterogeneous information into descriptive information and comparative information. Descriptive information is the evaluation of a single entity, which effectively reflects the quality of a single entity. Comparative information is focused on comparisons between pairwise products. New energy vehicles of different brands that appear together in the comparison sentences can be regarded as competitive products. Enterprises can identify competitors from the comparison sentences of related products, find the shortcomings of their own products, and improve the competitiveness of products. At the same time, these comparative sentences can be used to analyze the advantages and disadvantages of different new energy vehicles, and objectively provide reference opinions for consumers.

These extensive review texts, being consumers' after-sales feedback, contain both positive and negative sentiments of various consumer groups, which strongly influence future consumer purchase behavior (Carpenter et al., 2014). NEV manufacturers can obtain and analyze customers' preferences and insights, and can identify the strengths and weaknesses of existing products by mining consumers' sentimental attitudes of heterogeneous information in online reviews (Gong et al., 2020). But existing online reviews of users' sentiment tendencies composed of several complex descriptive and comparative sentences, which are difficult to be measured directly (Fan et al., 2020). Therefore, sentiment analysis (SA) results that elucidate product or service features for DMs from online reviews is an important topic.

Both SA of descriptive information and comparative information are complex system project, in which SA of descriptive information involves sentiment mining, sentiment polarity recognition, sentiment content representation, and aggregation of SA results for different attributes (Camilleri, 2020). Different from the SA of descriptive information, SA of comparative information needs to add comparison sentence recognition and comparison relation extraction. Based on the granularity of the processed text, SA can be divided into chapter/sentence-level SA and aspect-based SA. Unlike chapter/sentence-level SA, aspect-based SA needs to consider both the target information (attributes) and their corresponding sentiments. SA of online reviews of NEVs belongs to aspect-based SA because the reviews contain information about the various attributes of NEVs such as performance, appearance and interior. A natural question is how to mine and describe the sentiment tendencies on its attributes and how to aggregate the sentiment information on the attributes to form the overall SA results.

Common purposes of SA are to analyze the positive, negative, or neutral sentiment tendencies expressed or implied in a text. However, text-based SA often contains various ambiguities that must be resolved (Pathak et al., 2021). These ambiguous heterogeneous texts may lead to inaccurate classification results from sentiment classification models (Karthik, & Ganapathy, 2021). Therefore, the uncertainty of sentiment analysis results for different attributes must be considered when performing SA on NEVs. As a nonlinear information fusion method, Evidential reasoning (ER) can effectively deal with the uncertainty of attribute values. Consequently, the ER was introduced to describe the uncertainty in the SA process for NEVs. (Sun et al., 2019).

When evaluating the market acceptance of multiple NEVs, it is necessary to consider not only one-dimensional description information, but also two-dimensional comparison information. Then what mechanism or model to adopt to integrate different dimensional information is the difficulty of this paper. The advantage of the description information for individual NEVs is that it contains a detailed evaluation of individual products, and the disadvantage is that all NEVs are evaluated in isolation. The advantage of comparative information is that it provides a comparison of the relative advantages or disadvantages between NEVs, the disadvantage is that this abstract comparison is a relative relationship, and the evaluation of the NEVs themselves with which it is compared cannot be determined. Therefore, the combination of descriptive information and comparative information can more comprehensively evaluate the market acceptance of new energy vehicles. Relevant scholars proposed that graph models could be used to solve the above difficulties (Zhang et al., 2010; Li et al., 2011).

In summary, in order to integrate the SA results of heterogeneous information, an analysis framework based on ER-PR method is proposed to comprehensively evaluate the market acceptance of NEVs. Firstly, SA is performed on the descriptive and comparative information of NEVs, and the sentiment values of different sentiment intensity levels are converted into the belief structure of the NEVs' market acceptability. On this basis, the integrated description information and comparison information of the directed weighted graph of NEVs are constructed. According to the feature description information, the characteristics of NEVs are objectively weighted based on the maximum entropy optimization model, and ER analysis algorithm is used to determine the node weight of the directed weighted graph, and the directed edge and directed edge weight of the directed weighted graph are determined according to the comparative relationship between different NEVs. Then, based on the original principle of PR algorithm, the algorithm for solving the ranking value of the market acceptance of NEV is given.

Compared to previous research efforts, the major contributions and key features of our research are summarized as follows:

(1) Firstly, in the field of NEVs, previous studies usually study the descriptive information of individual NEVs, however, this paper studies the evaluation of market acceptance of NEVs based on descriptive and comparative information from the perspective of customers

(2) Secondly, this paper considers the results of multi-level intensity SA of descriptive information and comparative information, which can more delicately reflect the evaluation of NEVs in each sentiment intensity.

(3) Finally, using the directed weighted graph, an evaluation framework based on the ER-PR method is designed to evaluate the market acceptance of NEVs in uncertain environments, providing robust and reliable evaluation results.

The rest of the paper is organized as follows. Section 2 introduces the methodologies that were adopted. Section 3 provides the conclusion.

2. Methodology

2.1. The concept

In order to make full use of online review information to evaluate the market acceptance of NEVs, a graph model is used to integrate the descriptive and comparative information, and a framework for evaluating the market acceptance of NEVs based on ER-PR method is proposed. The flow chart is shown in Figure 1.

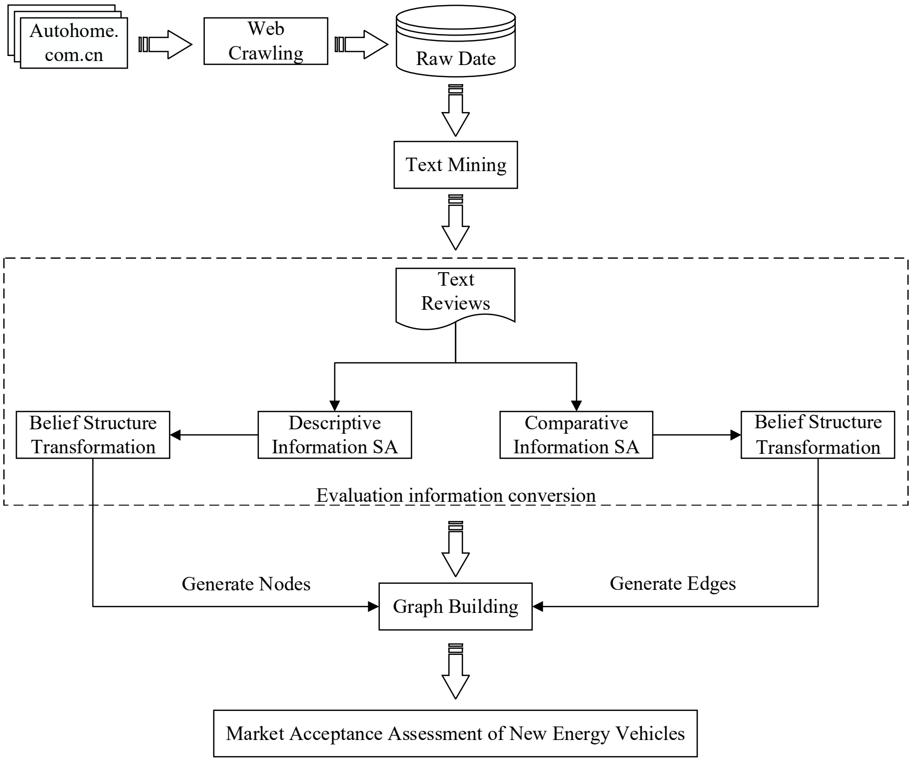


Figure 1: Flowchart of Integrating Heterogeneous Information for NEVs.

2.2. Sentiment analysis based on descriptive information and comparative information

(1) Sentiment analysis based on descriptive information

This paper analyzes the sentiment of descriptive information based on sentiment dictionary. In advance, a sentiment dictionary containing several groups of sentiment words is constructed according to different sentiment categories. Each dictionary can express different sentiment tendencies, and the sentiment value of the text to be analyzed is calculated by word matching and certain sentiment calculation rules. Based on the Chinese sentiment dictionary HowNet analysis, this paper calculates the sentiment value of n descriptive text comments of NEV I_j under the index system, and obtains the sentiment value matrix by normalization processing:

$$A^0 = \begin{pmatrix} a_{11}^0 & a_{12}^0 & \cdots & a_{1L}^0 \\ a_{21}^0 & a_{22}^0 & \cdots & a_{2L}^0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}^0 & a_{n2}^0 & \cdots & a_{nL}^0 \end{pmatrix} \tag{1}$$

(2) Sentiment analysis based on comparative information

The SA of comparative information includes the recognition of comparative sentences, the extraction of comparative relations and the determination of the sentiment tendency of comparative sentences. The main research includes the following aspects:

1) Identify the comparison sentences using the associative feature list

In this paper, an association feature dictionary is constructed, and the summarized comparative sentence features are stored into the association feature dictionary in a regular way for the recognition of Chinese comparison sentences.

2) Conditional random field model was used to extract comparison relationships

This paper uses the learning method of conditional random field and the prediction method to extract the features of the comparison sentence and train the model, extract the main components of the comparison sentence, and analyze the comparison subject, comparison object, comparison attribute and comparison result combined with the sentence structure characteristics of the comparison sentence.

3) The sentiment dictionary is used to calculate the propensity of the comparison entities

The tendentious analysis of the comparative sentence is mainly to judge the commentator's emotional tendentiousness towards the comparative subject and object in the comparative sentence. Based on the research work of the first two parts, this paper calculates the propensity of new energy vehicles in comparative sentences by constructing multiple sentiment dictionaries.

The specific system process is shown in Figure 2.

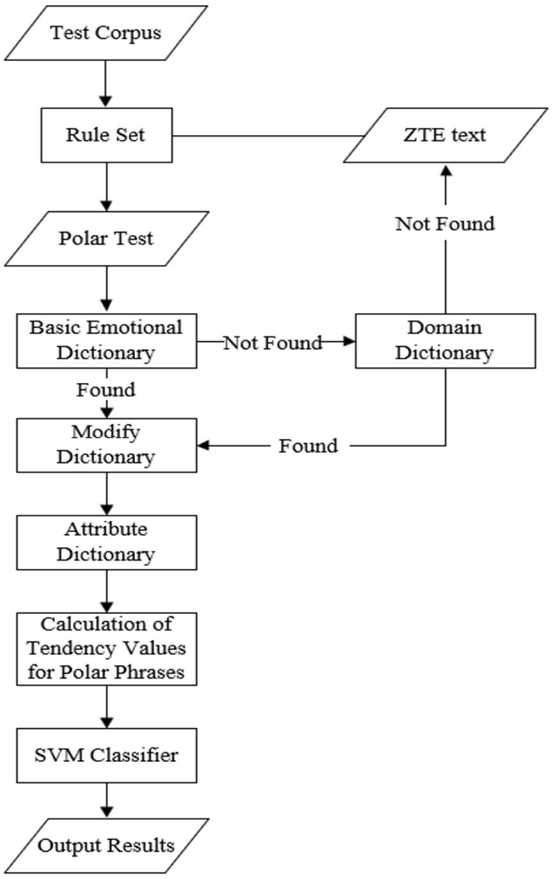


Figure 2: The Entity Propensity Analysis Process was Compared

2.3. Transformation of belief structures for affective intensity ratings based on ER method

Due to the limitations of some online review publishers on the understanding of the NEVs they review, which leads to the uncertainty of the SA results. Because ER can deal with unknown and uncertain information (Wang, et al., 1995). Therefore, in order to better deal with the multi-attribute online comment decision problem with uncertain affective tendency information, this paper uses the ER method to transform the belief structure of the affective intensity level.

Suppose that the problem of multiple attribute decision making involves M NEVs $\{I_1, I_2, \dots, I_j, \dots, I_M\}$and L attributes $\{C_1, C_2, \dots C_q, \dots, C_L\}$. $H_n(n = 1,2\dots, N)$ denotes the evaluation grades of attributes. In general, we suppose that H_{n+1} is preferred to H_n. If an NEV I_j is assessed on attribute C_q to a grade H_n with the degree of belief $\beta_n\left(I_j^q\right)$, then we use the belief structure $\left\{H_n, \beta_n\left(I_j^q\right)\right\}$ to describe the attribute information which requires $0 \leq \beta_n\left(I_j^q\right) \leq 1; \sum_{n=1}^{N} \beta_n\left(I_j^q\right) \leq 1$. If $\sum_{n=1}^{N} \beta_n\left(I_j^q\right) = 0$, it means that the ith index information of NEV I_j is missing. If $\sum_{n=1}^{N} \beta_n\left(I_j^q\right) = 1$, the evaluation of NEV I_j on attribute C_q is said to be complete. In this paper, sentiment level intensity dictionary is used to transform the text SA results of descriptive information, and the confidence structure of the index is obtained.

A comparative sentence between two brands of NEVs I_i and I_j is defined as $D = \{d_1, d_2, \dots, d_h\}$. Let ${D_{ij}}^{kq} = (I_i, I_j, {\alpha_{ij}}^{kq}(I_i|I_i, I_j), {\beta_{ij}}^{kq}(I_j|I_i, I_j))$, (i,j=1,2,...,M,q=1,2,...,L,) denote the emotional tendency index vector of NEV I_i and I_j on attribute C_q in comparison sentence $d_k \in D$. ${D_{ij}}^{kq} = (I_i, I_j, R^{kq}(I_i|I_i, I_j), R^{kq}(I_j|I_i, I_j))$ can have two situations. If the NEV I_i is better than I_j in attribute C_q, then ${D_{ij}}^{kq} = \left(I_i, I_j, {\alpha_{ij}}^{kq}, 0\right)$; If not, then ${D_{ji}}^{kq} = (I_i, I_j, 0, {\beta_{ij}}^{kq})$.

Let $L = \{L^1, L^2, \dots, L^r, \dots, L^v\}$ be the set of emotional intensity levels of NEVs, where L^r denotes the rth emotional intensity level, and a higher r denotes a higher emotional intensity level. The set of emotional intensity levels $L = \{L^1, L^2, \dots, L^r, \dots, L^v\}$ of NEVs alternatives is considered as the identification frame $H_n(n = 1,2\dots, N)$, and attribute $\{C_1, C_2, \dots C_q, \dots, C_L\}$ is considered as the evidence set. Firstly, according to the comment frequency t_{jq}^r of emotional intensity level L_q^r and the comment frequency t_{jq}^H of uncertain emotional intensity level L_q^H in NEV I_j's online comments about attribute C_q, the probability P_q^r and P_q^H of emotional intensity level L_q^r and uncertain emotional intensity level L_q^H of NEVs I_j about attribute C_q is calculated. And its probability P_q^r is expressed as the belief structure $\left\{H_n, \beta_n\left(I_j^q\right)\right\}$ of NEVs $I_j$about the attribute C_q, where v is equal to n.

2.4. Construction of directed weighted graph based on descriptive and comparative information of NEVs

Suppose that the NEV directed weighted graph is represented as $G = \left(V, E, w(I_j), w(I_i, I_j)\right)$ by 4-tuple, where V is the set of nodes, E is the set of directed edges, $w(I_j)$ is the node weight, and $w(I_i, I_j)$ is the directed edge weight. Figure 3. is a simple example to illustrate how to build the weighted digraph structure.

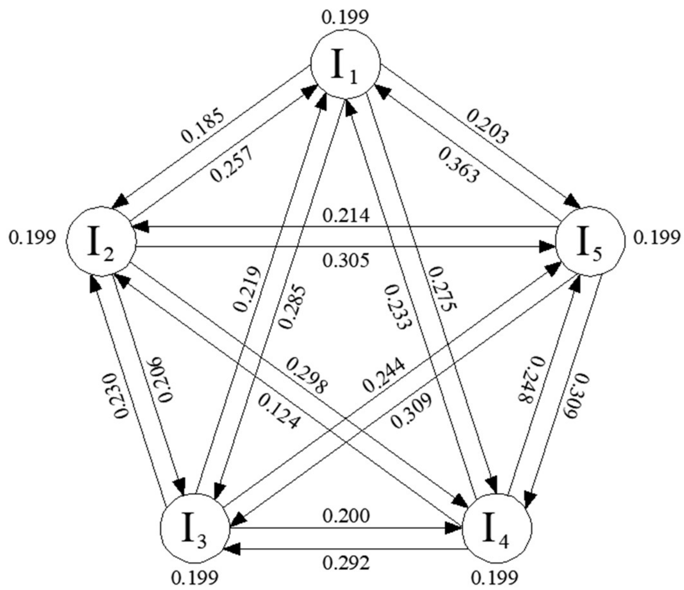


Figure 3: Example of Weighted Digraph for Five NEVs.

(1) Determination of node weights based on directed weighted graph of NEVs

The weight of nodes in the directed weighted graph represents the importance of nodes, which can be determined based on the decision matrix $A = \left[a_{jq}\right]_{M \times L}$ of the attribute information of NEVs, and the decision matrix $B = \left[b_{jq}\right]_{M \times L}$ is obtained by normalizing it. In order to determine the weight of each node in the directed weighted graph based on $B = \left[b_{jq}\right]_{M \times L}$, it is necessary to determine the weight w_q of L attributes of NEVs. Different attribute weights will produce different evaluation results. In view of this problem, this paper follows the principle of fairness to find a specific combination of four groups of weights to form a convex combination, so that the distance from the four groups of weights is as equal as possible, and the maximum entropy optimization model is constructed as follows:

$$max - \sum_{k=1}^{4} \frac{dist(\overline{w}, w^k)}{\sum_{l=1}^{4} dist(\overline{w}, w^l)} \cdot log_2 \frac{dist(\overline{w}, w^k)}{\sum_{l=1}^{4} dist(\overline{w}, w^l)} \tag{2}$$

$$s.t.\ dist(\overline{w}, w^k) = \sqrt{\sum_{i=1}^{L} \left(\overline{w}_i - w_i^k\right)^2}, k = 1, \cdots, 4; \tag{3}$$

$$\overline{w}_i = \sum_{k=1}^{4} w_i^k \theta_k, i = 1, \cdots, L; \tag{4}$$

$$0 \leq \theta_k \leq 1, k = 1, \cdots, 4; \tag{5}$$

$$\sum_{k=1}^{4} \theta_k = 1 \tag{6}$$

Based on the fair weight $\overline{w}$ and the normalization matrix $B = \left[b_{jq}\right]_{n \times L}$, the analytical algorithm in the ER framework is used to synthesize the emotional tendency of NEVs on each attribute, and then the utility function of each NEV is calculated. The utility function is used as the weight $w'(I_j)$ of each node in the directed weighted graph, as follows:

1) Calculate the belief structure of the NEVs' descriptive information.

For a given I_j, the basic probability assignment function of index q is measured by the following programming:

$$
\begin{aligned}
& m_{n,q} = m_q(H_n) = w_q \beta_n(I_j^i); n = 1,2,\ldots,N; q = 1,2,\ldots,L; j = 1,2,..,M \\
& m_{H,q} = m_q(H) = 1 - \sum_{n=1}^{N} m_{n,q} = 1 - w_q \sum_{n=1}^{N} \beta_n\left(I_j^q\right) \\
& \overline{m}_{H,q} = \overline{m}_q(H) = 1 - w_q; \widetilde{m}_{H,q} = \widetilde{m}_q(H) = w_q\left(1 - \sum_{n=1}^{N} \beta_n\left(I_j^q\right)\right) \\
& m_{H,q} = \overline{m}_{H,q} + \widetilde{m}_{H,q}; \sum_{q=1}^{L} w_q = 1
\end{aligned}
\tag{7}
$$

Here, $m_{n,q}$is the basic probability mass of I_j assessed as the evaluation grade H_n on the attribute C_q. The unassigned belief is denoted by $m_{H,q}$, consisting of $\overline{m}_{H,q}$and $\widetilde{m}_{H,q}$, where $\overline{m}_{H,q}$ is caused by the attribute weights and $\widetilde{m}_{H,q}$ is caused by the uncertainty of the measurement on the attribute C_q for I_j. For NEV I_j, if the qth indicator information is missing, then $m_{n,q} = 0$ and $m_{H,q} = 1$.

Then, the belief structure of the NEVs' descriptive information can be obtained by using the evidential-reasoning rule:

$$
\begin{aligned}
& \{H_n\}: m_n = \overline{k}\left[\prod_{q=1}^{L}\left(m_{n,q} + \overline{m}_{H,q} + \widetilde{m}_{H,q}\right) - \prod_{q=1}^{L}\left(\overline{m}_{H,q} + \widetilde{m}_{H,q}\right)\right] \\
& \{H\}: \widetilde{m}_H = \overline{k}\left[\prod_{q=1}^{L}\left(\overline{m}_{H,q} + \widetilde{m}_{H,q}\right) - \prod_{q=1}^{L}\left(\overline{m}_{H,q}\right)\right] \\
& \{H\}: \overline{m}_H = \overline{k}\left[\prod_{q=1}^{L}\left(\overline{m}_{H,q}\right)\right] \\
& \overline{k} = \left[\sum_{n=1}^{N}\prod_{q=1}^{L}\left(m_{n,q} + \overline{m}_{H,q} + \widetilde{m}_{H,q}\right) - (N-1)\prod_{q=1}^{L}\left(\overline{m}_{H,q} + \widetilde{m}_{H,q}\right)\right]^{-1} \\
& \{H_n\}: \beta_n = m_n/(1 - \overline{m}_H); n = 1,2,\ldots,N \\
& \{H\}: \beta_H = \widetilde{m}_H/(1 - \overline{m}_H)
\end{aligned}
\tag{8}
$$

where β_n and β_H represent the belief degrees of the NEVs' market acceptability assessed as grade H_n and assessment frame H, respectively.

2) Calculate the expected utility of the NEVs' descriptive information.

Suppose the utilities of the assessment grades are $u(H_1), u(H_2), \ldots, u(H_N)$, the maximum and minimum utilities of each NEV are calculated by the following formulas:

$$
\begin{aligned}
&u(I_j)\left(\beta_N(I_j)+\beta_H(I_j)\right)(H_N)\sum_{n=1}^{N-1}\beta_n(I_j)\,(H_n)_{max} \\
&u(I_j)\sum_{n=2}^{N}\beta_n(I_j)\,(H_n)\left(\beta_1(I_j)+\beta_H(I_j)\right)(H_1)_{min}
\end{aligned}
\tag{9}
$$

$$
u_{avg}\left(I_j\right)=\frac{u_{\max}\left(I_j\right)+u_{\min}\left(I_j\right)}{2} \tag{10}
$$

The average utility value $u_{avg}(I_j)$ is taken as the weight $w'(I_j)$ of each node in the directed weighted graph, that is $w'(I_j)=u_{avg}(I_j)$. The weights of each node in the directed weighted graph were normalized as follows:

$$
w(I_j)=w'(I_j)/\sum_{j=1}^{M}w'(I_j) \tag{11}
$$

(2) Determination of directed edge and directed edge weight based on comparative information of NEVs

Determine the comparative emotional tendency relationship $D_{ij}{}^{kq}=(I_i,I_j,R^{kq}(I_i|I_i,I_j),R^{kq}(I_j|I_i,I_j))$ between NEVs in every pair of nodes under attribute C_q. If the NEV I_i is better than I_j in the C_q attribute, then $D_{ij}{}^{kq}=(I_i,I_j,\alpha_{ij}{}^{kq},0)$, there is a directed edge from NEV I_j to I_i, and the weight of the directed edge is calculated as follows:

$$
w_q(I_i,I_j)=\frac{\alpha_{ij}{}^{kq}}{\sum_{j=1}^{M}\rho\left(R^{kq}(I_i|I_i,I_j)\right)} \tag{12}
$$

where, for $\forall j\in\{1,2,\cdots,M\}$, $D_{ij}{}^{kq}=(I_i,I_j,\alpha_{ij}{}^{kq},0)$, then $\rho=1$; For $\forall j\in\{1,2,\cdots,M\}$,$D_{ij}{}^{kq}=(I_i,I_j,0,\beta_{ij}{}^{kq})$, then ρ=0.

Then, using the ER algorithm, it is determined that for L attributes $\{C_1,C_2,\dots C_q,\dots,C_L\}$ (q=1,2,...,L) of the integrated weight of the directed edges. The calculation formula is as follows:

$$
w(I_i,I_j)=\frac{u_{avg}^{R}(I_i,I_j)}{\sum_{j=1}^{M}u_{avg}^{R}(I_i,I_j)} \tag{13}
$$

Among them, on L attributes, the ER is used to integrate the directed edge weights when NEV I_i is superior to I_j, denoted by $u_{avg}^{R}(I_i,I_j)$.

2.5. NEV market acceptance ranking method based on directed weighted graph

According to the directed weighted graph of NEVs constructed in the previous section, this paper classifies different brands of NEVs based on the PR algorithm. In PR algorithm, web pages are regarded as nodes, and the weight value of each node represents the quality of the corresponding page. The hyperlinks between web pages are regarded as directed edges, and the weights of directed edges represent the importance of page links. Let the probability that the customer continues to search for other NEVs be σ, then the probability that the customer stops searching is $1-\sigma$. The PR value $PCS(I_j)$ of the NEV $I_j$is as follows:

$$
PCS(I_j)=(1-\sigma)w(I_j)+\sigma\sum_{i=1}^{M}\beta PCS(I_i)\,w(I_i,I_j) \tag{14}
$$

Here, σ is the damping coefficient and M is the number of NEVs. The generation of NEVs market acceptance ranking can be transformed into the calculation of a matrix characteristic direction, which is expressed as follows:

$$PCS = (1 - \sigma) * w + \sigma * M * PCS \quad (15)$$

Where M is the transition matrix and has the form:

$$M = \begin{bmatrix} 0 & w(I_1, I_2) & \cdots & w(I_1, I_M) \\ w(I_2, I_1) & 0 & \cdots & w(I_2, I_M) \\ \vdots & \vdots & \vdots & \vdots \\ w(I_M, I_1) & w(I_M, I_2) & \cdots & w(I_M, I_M) \end{bmatrix} \quad (16)$$

Due to the convergence of PCS_k to a stationary probability distribution, formula (15) can be solved by iterative method. According to references (Li et al., 2017), the specific steps for solving formula (15) by iterative method are given.

1) PCS is assigned an initial value of PCS_0;

2) The initial value PCS_0 is substituted into equation (15) to obtain PCS_1;

3) PCS_1 is assigned to PCS_0 to obtain PCS_2;

4) Repeat step 3), that is, continuously assign the value of PCS_{k+1} to PCS_k until the iteration stops when the condition $|PCS_{k+1} - PCS_k| < \varepsilon$ is met, where ε is the condition for stopping the iteration, which can be given by the DM;

5) Finally, the $PCS(I_j)$ of NEVs is calculated by the iterative algorithm, and the market acceptance of NEVs is ranked based on this.

2.6. Comparison with other methods and management insights

(1) Comparison with other methods

In this subsection, we compare the experimental results obtained by the method in this article with two different methods.

The first method is to rank products based on text sentiment (Tian, et al., 2009). The advantage of this method is that text sentiment provides rich information, but its disadvantage is that it only describes the evaluation of a single product, which cannot fully evaluate the relative advantages or disadvantages between products and cannot rank all products based on the relationship between different products.

The second method is to rank products based on comparison sentences (Zhang, et al., 2013). The advantage of comparison sentences is to directly compare the relative advantages and disadvantages of products with similar products, and each node in the product comparison graph model has no weight, which will ignore the text sentiment of individual products.

This paper obtains the sales ranking of NEVs through third-party websites. The performance of the ranking method is measured by the correlation between the ranking results of NEVs obtained by our method and other methods and the sales ranking of NEVs obtained by third-party websites. The Spearman rank correlation coefficient can evaluate the correlation between two rankings, and its formula is as follows:

$$\phi\left(\vec{R_{x_i}} - \vec{R_{y_i}}\right) = 1 - \frac{6\sum_{i=1}^{n}\left(R_{x_i} - R_{y_i}\right)^2}{n(n^2-1)} \tag{17}$$

Where, $\vec{R_{x_i}}$ represents the market acceptance ranking of new energy vehicles obtained by the method in this paper and the other two methods, and $\vec{R_{y_i}}$ represents the sales ranking obtained on third-party websites. The closer the correlation coefficient ϕ is to 1, the more superior and reliable the results calculated by this method are.

Among all the sorting methods, this paper integrates the sentiment of a single NEV text and the sentiment of two NEVs comparing each other in sentences, and integrates rich information to comprehensively and scientifically evaluate the acceptance of the NEVs market. The correlation coefficient is the highest, so we have a related measure of the market acceptance of NEVs.

(2) Management insights

Based on the results of the assessment framework and the corresponding analysis, our study attained several potential management insights, which are summarized as follows.

1) The herding effect makes consumers pay attention to the experience and feedback of other consumers while purchasing NEVs. This phenomenon stems from the human inclination to seek social validation and minimize the risks associated with novel and complex purchases, such as NEVs. The ready availability of information on various NEV models allows consumers to make comprehensive comparisons and draw accurate conclusions. Consumers can not only use the textual emotions in a single NEV to judge its own value, but also quickly determine which NEVs are more favored by a wider audience through pairwise comparisons of NEVs, thus gaining valuable insights into the popular and reliable choices. Understanding the impact of the herding effect on consumer decision-making is vital for automakers and marketers in the NEV industry. Managers can capitalize on this phenomenon by fostering positive consumer experiences and encouraging satisfied customers to share their feedback.

2) In the rapidly evolving landscape of NEVs, understanding market competition is crucial for companies to thrive. Traditional approaches to evaluating competition may fall short in capturing nuanced sentiments and preferences results of consumers with multiple emotional intensities from both a single subject and a comparison between two main subjects simultaneously. This study proposes a novel method that leverages SA of online reviews descriptive information and comparative information to assess the competitive dynamics between NEVs. Sentiment analysis offers a powerful tool for extracting and quantifying consumer sentiments expressed in online reviews. By employing sentiment analysis techniques, NEV companies can measure the overall sentiment polarity and intensity associated with their products as well as those of their competitors. The dominance probability, derived from sentiment analysis, serves as an indicator of market competition between NEVs. When the directed edge weights of two NEVs are closely matched, it suggests a higher level of competitive threat. Such a scenario calls for sales companies to shift their focus and closely monitor their competitors' strategies, product enhancements, and customer experiences to maintain or improve their market position.

3. Conclusions

Despite the promising prospect of EVs, the number of EVs in the market is still insufficient and the willingness of private individuals to buy these is not strong enough. Under the limited market demand, one of the effective ways to promote purchasing is to optimize the product structure of EVs to provide products that are acceptable to consumers. Therefore, it is imperative to measure the market acceptability of EVs. This paper proposes an evaluation framework based on the ER-PR method to evaluate the market acceptance of NEVs in uncertain environments, providing robust and reliable evaluation results. This model uses graph model to integrate description information and two-dimensional information, which makes up for the defects caused by only using a single form of data, thereby improving the reliability of the developed model, thereby improving the reliability of the developed model. Finally, the rationality and practicability of the method was verified via an empirical example.

In theory, belief structure in evidence theory is used to transform the SA results, taking into account the uncertainty of the sentiment classification model, and a graph model is constructed to analyze the emotional results of descriptive information and comparative information based on the evaluation framework of the ER-PR method. These optimized methods can provide reliable decision results, which is of great significance for the development and enrichment of scheme ranking theory based on online comments. In practice, this study did not use traditional simulation data, but instead adopted a new data source - big data, which combines big data with traditional multi-attribute decision-making methods to evaluate NEV market acceptance, providing effective management insights for consumers, enterprises, and manufacturers.

While this study provides a comprehensive evaluation framework, there are limitations that can be addressed in future research. Here, we summarize several future research directions. Firstly, although the article collects information from recent consumer reviews to mine consumer preferences, these may change over time. Therefore, continuous understanding of the market acceptability of an electric vehicle is required for dynamic decisions in the future. Secondly, the evaluation method proposed in this paper assumes independent features, highlighting the need for data-driven approaches that consider feature relationships. Finally, in this paper, the whole evaluation process is carried out by one manager. Group consensus can improve the reliability of decision-making. Therefore, data-driven group decision making is also a future research direction.

References

Argonne National Laboratory (ANL). (2023, April). *Light duty electric drive vehicles monthly sales update.* https://www.anl.gov/es/light-duty-electric-drive-vehicles-monthly-sales-updates/.

Camilleri, A. R. (2020). The importance of online reviews depends on when they are presented. *Decision Support Systems*, *133*, 113307. http://dx.doi.org/10.1016/j.dss.2020.113307.

Carpenter, T., Golab, L., & Syed, S. J. (2014, June). Is the grass greener? Mining electric vehicle opinions. In Proceedings of the 5th international conference on Future energy systems (pp. 241-252). https://doi.org/10.1145/2602044.2602050.

Chen, J., & Li, X. (2023). Doctors ranking through heterogeneous information: The new score functions considering patients' emotional intensity. *Expert Systems with Applications*, *219*, 119620. http://dx.doi.org/10.1016/j.eswa.2023.119620.

China Passenger Car Association (CPCA). (2023, May). *2023 New energy vehicle industry report*. http://www.cpcaauto.com/news.php?types=bgzl&anid=126.

Du, J., Meng, X., Li, J., Wu, X., Song, Z., & Ouyang, M. (2018). Insights into the characteristics of technologies and industrialization for plug-in electric cars in China. *Energy*, *164*, 910-924. http://dx.doi.org/10.1016/j.energy.2018.09.060.

Fan, Z. P., Li, G. M., & Liu, Y. (2020). Processes and methods of information fusion for ranking products based on online reviews: An overview. *Information Fusion*, *60*, 87-97. http://dx.doi.org/10.1016/j.inffus.2020.02.007.

Gong, B., Liu, R., & Zhang, X. (2020). Market acceptability assessment of electric vehicles based on an improved stochastic multicriteria acceptability analysis-evidential reasoning approach. *Journal of Cleaner Production*, *269*, 121990. http://dx.doi.org/10.1016/j.jclepro.2020.121990.

He, L. Y., Yang, S., & Chang, D. (2017). Oil price uncertainty, transport fuel demand and public health. *International Journal of Environmental Research and Public Health*, *14*(3), 245. https://doi.org/10.3390/ijerph14030245.

He, S., & Wang, Y. (2023). Evaluating new energy vehicles by picture fuzzy sets based on sentiment analysis from online reviews. Artificial Intelligence Review, *56*(3), 2171-2192. https://doi.org/10.1007/s10462-022-10217-1.

International Energy Agency (IEA). (2023, April). *Global EV outlook 2023*. https://www.iea.org/reports/global-ev-outlook-2023.

Karthik, R. V., & Ganapathy, S. (2021). A fuzzy recommendation system for predicting the customers interests using sentiment analysis and ontology in e-commerce. *Applied Soft Computing*, *108*, 107396. http://dx.doi.org/10.1016/j.asoc.2021.107396.

Kumar, R. R., & Alok, K. (2020). Adoption of electric vehicle: A literature review and prospects for sustainability. *Journal of Cleaner Production*, *253*, 119911. http://dx.doi.org/10.1016/j.jclepro.2019.119911.

Li, S., Ming, Z. Y., Leng, Y., & Guo, J. (2017). Product ranking using hierarchical aspect structures . *Journal of Intelligent Information Systems*, 48(2), 453-474. https://doi.org/10.1007/s10844-016-0421-8.

Li, S., Zha, Z. J., Ming, Z., Wang, M., Chua, T. S., Guo, J., & Xu, W. (2011, July). Product comparison using comparative relations. *In Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval* (pp. 1151-1152). https://doi.org/10.1145/2009916.2010094.

Pathak, A. R., Pandey, M., & Rautaray, S. (2021). Topic-level sentiment analysis of social media data using deep learning. *Applied Soft Computing*, *108*, 107440. http://dx.doi.org/10.1016/j.asoc.2021.107440.

Peng, H., Cambria, E., & Hussain, A. (2017). A review of sentiment analysis research in Chinese language. *Cognitive Computation*, *9*, 423-435. http://dx.doi.org/10.1007/s12559-017-9470-8.

Sun, X., Liu, X., Wang, Y., & Yuan, F. (2019). The effects of public subsidies on emerging industry: An agent-based model of the electric vehicle industry. *Technological Forecasting and Social Change*, *140*, 281-295. http://dx.doi.org/10.1016/j.techfore.2018.12.013.

Tesla. (2023, January). *Tesla's 2022 Q4 financial statement*. https://ir.tesla.com/#quarterly-disclosure.

Tian, P., Liu, Y., Liu, M., & Zhu, S. (2009). Research of product ranking technology based on opinion mining. *Proceedings of Second International Conference on Intelligent Computation Technology and Automation* (pp. 239-243). https://doi.org/10.1109/ICICTA.2009.773.

Wang, J., Yang, J. B., & Sen, P. (1995). Safety analysis and synthesis using fuzzy sets and evidential reasoning. Reliability Engineering & System Safetty, 47(2), 103-118. https://doi.org/10.1016/0951-8320(94)00053-Q.

Wang, X. W., Cao, Y. M., & Zhang, N. (2021). The influences of incentive policy perceptions and consumer social attributes on battery electric vehicle purchase intentions. *Energy Policy*, *151*, 112163. http://dx.doi.org/10.1016/j.enpol.2021.112163.

Zhang, K., Narayanan, R., & Choudhary, A. (2010). Voice of the customers: mining online customer reviews for product feature-based ranking. In *3rd Workshop on Online Social Networks (WOSN 2010)*. https://www.usenix.org/legacy/event/wosn10/tech/full_papers/Zhang.pdf.

Zhang, Z., Guo, C., & Goes, P. (2013). Product comparison networks for competitive analysis of online word of mouth. *ACM Transactions on Management Information System*, 3(4), 1-22. https://doi.org/10.1145/2407740.2407744.

24th International Conference on Group Decision and Negotiation &
10th International Conference on Decision Support System Technology
Human-Centric Decision and Negotiation Support for Societal Transitions

A group-based failure mode and effect analysis approach with D-S evidence theory and social network for medical waste management

Limei Hu [1,2], Haiyan Xu [1*], Dequn Zhou [1], Rui Liu [1]

[1] College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu, China
[2] College of Management, Anhui Science and Technology University, Bengbu, Anhui, China
[*] xuhaiyan@nuaa.edu.cn

Abstract

As a recognized and powerful risk analysis tool, failure mode and effects analysis (FMEA) has proven its effectiveness in medical waste management, which has profound significance to human health and ecological environment. However, the traditional FMEA have several limitations in the uncertainty modeling and the weight determination of the risk indicators (RFs), and existing FMEA studies also pay little attention to the social network of team members (TMs), potentially leading to inappropriate results. Therefore, this paper constructs a group-based FMEA approach with Dempster-Shafer evidence theory (D-S evidence theory) and social network for medical waste management, which can enhance the reliability of FMEA. Firstly, in order to fully consider the existence of multiple emotional preferences and unknown preferences in evaluation information, the belief structure of D-S evidence theory is employed to process and transform evaluations of failure modes (FMs). Secondly, a dual network structure is constructed by combining the undirected network graph based on the opinion similarity and the directed network graph based on TMs' trust relationship to obtain the weights of TMs. The evaluation information from TMs is integrated based on the modified Dempster's combination rule. Thirdly, the weights of risk factors (RFs) from an objective view is calculated by using the maximizing deviation method based on the belief structure. Afterwards, an integrated evaluation can be calculated to prioritize FMs by fusing the RFs' evaluations. Finally, the effectiveness of the proposed method is verified through the case of medical waste risk management. The proposed group-based FMEA approach, which fully considered the uncertainty of evaluation information and the relationship network between TMs, can get more accurate and practical risk ranking results to help the healthcare industry formulate accurate risk prevention and control plans.

Keywords: Medical waste management; Failure mode and effects analysis; Evidence theory; Social network; the maximizing deviation method

1. Introduction

Medical waste is one of the toughest challenges facing the healthcare industry and the entire human race in decades (Perumal et al., 2021). Infectious medical waste, as a special kind of harmful medical waste, carries pathogenic microorganisms and has the risk of causing the spread of infectious diseases. It will pose a huge threat to ecology and human beings handled non-properly (Luo and Liao, 2022). Therefore, it is highly necessary for all countries to prevent catastrophic accumulation of infectious waste with sustainable medical waste management (Ghoushchi et al., 2022).

As an ex-ante risk and reliability management technique, FMEA aims to identify, analyze and prevent potential failure modes in products, processes and systems. In order to reduce healthcare risks and improve service quality, FMEA has been successfully and widely used in the healthcare industry (Ghoushchi et al., 2022). The RPN method, a classical FMEA method, is used to evaluate and rank FMs in infectious medical waste disposal (Ho and Liao, 2011). However, RPN method has some recognized defects (Zheng et al., 2022): (i) The traditional FMEA method determines the evaluations of failure modes with the crisp number without considering the uncertainty of risk assessment problems; (ii) Equal weight of different risk indicators are assigned in the RPN method, while the relative importance of risk indicators is seldom coincidental in practice; (iii) The risk ranking sequence numbers of different failure modes with an equal value of RPN are the same, although there are different risk indicators ratings.

In response to the evaluation information processing problem, more and more technologies use fuzzy theory to describe the uncertainty of risk assessment information with the increasing complexity of risk assessment in FMEA problems. For example, Carnero (Carnero, 2020) improves the FMEA model with intuitionistic fuzzy sets for healthcare waste segregation. Spherical fuzzy sets and the multi-objective optimization using the ratio analysis (MOORA) method is employed to evaluate and rank FMs in sustainable medical waste design (Ghoushchi et al., 2022). The intuitionistic fuzzy sets are used to evaluate FMs of the intelligent minimally invasive medical equipment system (Liang et al., 2023). As an extended version, Liu et al. (Liu et al., 2021) adopt q-rung orthopair fuzzy sets in FMEA of blood transfusion issues. A FMEA model is constructed with internal and external TMs for medical waste management system by using probabilistic linguistic term sets (PLTSs) and D-S evidence theory to process evaluations of FMs (Liu et al., 2023) .

The determination of risk priority of FMs can be regarded as a multi-criteria group decision-making (MCGDM) problem. In the aspect of RFs' weights calculation, relevant scholars have done a lot of researches. For example, the weights of RFs are calculated using fuzzy Analytic Hierarchy Process (AHP) and Shannon entropy principle (Das et al., 2020). Combining linguistic Z-numbers (LZNs) and fuzzy entropy measure, a weighted entropy measure is proposed to determine the weights of RFs (Hu, 2023). Based on the idea of maximizing deviation, a standard weight determination method considering the different risk attitudes of the FMEA expert is constructed (Li et al., 2023). The RFs' weights are determined by combining the different importance of TMs on RFs and the maximizing deviation model (Liu et al., 2024).

The implementation of FMEA is a typical group decision-making (GDM) problem, so the weight calculation of TM is also one of the contents of the research. For example, the weights of FMEA TMs are derived from the similarity of FMEA TMs' opinions (Hu, 2023). A comprehensive weight determination method is established which takes the subjective and objective aspects of TMs' evaluations and revision factors into account (Li et al., 2023). The progress of modern network and communication technology has created conditions for the practice of FMEA in a distributed environment, the social network of experts will affect the evaluation of FMEA, and a weight determination method to determine TMs' weights is proposed that comprehensively considers the conflict level and social importance of TMs (Liu et al., 2023). For infectious medical waste management systems, a large number of experts from different departments or institutions were gathered to form an FMEA team, and TMs' weights are determined by considering the TMs' relationships and risk assessment information (Liu et al., 2024).

In order to overcome the shortcomings of traditional RPN method and obtain the results of risk prioritization of FMs, many studies have adopted different multi-criteria decision-making (MCDM) techniques. For example, the fuzzy Technique for Order Performance by Similarity to Ideal Solution (TOPSIS) method is adopted to rank the risk priorities of FMs by comprehensively considering all of RFs (Li et al., 2020). A new method is constructed to extend the prospect theory with the TOPSIS method to obtain the risk priorities of FMs (Fang et al.,2020). The risk priorities of FMs can be also obtained by employing the Vise Kriterijumska Optimizacija Kompromisno Resenje (VIKOR) method to improve the FMEA with Z-number which can capture the information reliably (Das et al., 2020). A new FMEA risk assessment method combining rough set theory with ELimination and Choice Translating REality II (ELECTRE II) method to obtain the FMs' ranking result (Sarwar et al., 2021). A generalized Interactive and Multi-criteria Decision Making (TODIM) method based on LZNs is constructed to obtain the risk ranking result of FMs (Hu, 2023). And an integrated DEMATEL-TODIM decision framework is proposed to prioritize FMs (Liu et al., 2023).

Existing studies have made significant contributions to the FMEA problem. Great progress has been made on the basis of the traditional RPN method. However, there are some research gaps:

(1) Evaluation information processing -- Existing researches cannot adequately address the circumstance in which there are multiple emotional preferences and unknown preferences in evaluation information. This paper will further process and transform evaluation information based on the belief structure of evidence theory which can distinguish well between uncertainty and ignorance.

(2) RFs' weights calculation -- The existing literature shows the superiority of the maximization deviation idea. In this paper, the weights of RFs from an objective view will be calculated by using the maximizing deviation method based on the belief structure.

(3) TMs' weights calculation -- The existing research is more based on the opinion similarity to obtain the weights of TMs and even considering the trust relationship between TMs, it is more used to adjust TMs' opinions. In this paper, a dual network structure will be constructed by combining the undirected network graph based on the opinion similarity of TMs and the directed network graph based on the trust relationship of TMs to obtain the weights of TMs by comprehensive calculation.

(4) Risk priority ranking -- Existing methods are more integrated with MCDA method for calculation, considering the fusion of belief structure evaluation expression form, this paper directly adopts Dempster's combination rule which can distinguish between uncertainty and ignorance while managing conflict between evidence. Inspired by the social networks, a modified Dempster's combination rule will be proposed for evaluation information integration to avoid counterintuitive synthesis results in medical waste management.

Motivated by the research status and gaps of FMEA, a group-based FMEA approach with D-S evidence theory and social network for medical waste management is established in this paper. The rest of this paper is organized as follows. Section 2 introduces some basic knowledge about D-S evidence theory and social network required for this article. Section 3 gives the theoretical structure and calculation steps of the proposed group-based FMEA approach. Section 4 illustrates the feasibility and practicality of the proposed FMEA model by applying it to a medical waste management system. Section 5 concludes the whole study.

2. Preliminaries

In this section, the fundamental concepts and theories of the D-S evidence theory and social network are introduced.

2.1 D-S evidence theory

The evidence theory provides a powerful tool for decision making through the representation and integration of uncertain information (Dempster, 1967; Shafer, 1976). The belief structure which can be used to better represent the evaluation can be defined as follows.

(1) Basic probability assignment

Definition 1. Let Θ be a set of mutually exclusive and collectively exhaustive events, often referred to as a frame of discernment. A basic probability assignment (BPA) is a mapping $m: 2^{\Theta} \rightarrow [0,1]$, which satisfies the following conditions:

$$m(\phi) = 0 \text{ and } \sum_{A \subseteq 2^{\Theta}} m(A) = 1,$$

where $m(A)$ is called the basic probability assignment of A. It is called the focal element for all A when it satisfies the condition of $m(A) > 0$.

(2) Dempster's combination rule

Let m_1 and m_2 be two BPAs on the same frame of discernment Θ, the Dempster's rule of combination can be used to combine them denoted by $m = m_1 \oplus m_2$ and defined by:

$$m(A)=\begin{cases}\frac{1}{K}\sum_{B\cap C=A} m_1(B)m_2(C), & A\neq\varnothing;\\ 0, & A=\varnothing,\end{cases} \tag{1}$$

where $K=\sum_{B\cap C\neq\varnothing} m_1(B)m_2(C)$ and K is a normalization constant. It reflects the degree of conflict between evidences. The Dempster's rule of combination is only applicable to such two belief structures which satisfy the condition $K<1$. This rule satisfies commutative and associative properties, i.e., (i) $m_1\oplus m_2=m_2\oplus m_1$ and (ii) $(m_1\oplus m_2)\oplus m_3=m_1\oplus(m_2\oplus m_3)$.

Assume that the reliability of the mass function m provided by a source of evidence is α (0<α<1), the discounted evidence $m^{'}$ on Θ can be defined as follows:

$$\begin{cases} m^{'}(A)=\alpha m(A), \ \forall A\in\Theta, A\neq\Theta\\ m^{'}(\Theta)=1-\alpha+\alpha m(\Theta)\end{cases} \tag{2}$$

where α is called the discount coefficient.

(3) A distance of evidence

An evidence distance formula is proposed to measure the degree of similarity between the fused evidences (Jousselme et al., 2001) as follows:

Definition 2. Let m_1 and m_2 be two BPAs on the same frame of discernment Θ, containing N mutually exclusive and exhaustive hypotheses. The distance between m_1 and m_2 is:

$$d_{BPA}(m_1,m_2)=\sqrt{\frac{1}{2}(m_1-m_2)^T \underline{\underline{D}}(m_1-m_2)}\,. \tag{3}$$

where $\underline{\underline{D}}$ is an $2^N\times 2^N$ matrix whose elements are $D(A,B)=\frac{|A\cap B|}{|A\cup B|}$, $A,B\in P(\Theta)$, which represents the relationship among the subsets of Θ. The $D(A,B)$ must be to unity when A is closer to B, on the contrary, the $D(A,B)$ must be to zero.

2.2 Social network

(1) The undirected network graph

The complex network is a kind of graph (see Figure 1) which can be used to describe the undirected network graph constructed in this paper based on the opinion similarity of FMEA team. A graph is a two-tuple composed of a set of points $V=\{v_i\}$ and a set of unordered pairs of elements in $E=\{e_k\}$, denoted as $G=\{V,E\}$. The element v_i in V is called a node, and the element e_k in E is called an edge. We can use $e_{ij}=1$ to represent node i is associated with node j. When $e_{ij}=0$, it means node i is unrelated to node j. In a complex network, the edge value can only be 0 or 1. In this paper, the edge value is extended to an interval [0, 1]. The closer the edge value between two nodes is to 1, the higher the similarity between the two nodes is.

In a complex network, the number of all edges connected by each node is the degree of this node, which is an indicator used to quantitatively reflect the importance of nodes. As a generalization of the degree of nodes, the vertex strength VS_i of node v_i can be defined as follows.

$$VS_i = \sum_{j \in N_i} e_{ij} \tag{4}$$

where N_i is the set of adjacent points of node v_i ; e_{ij} is the value of the edge between node v_i and node v_j . It is worth mentioning that e_{ij} represents the degree of interaction between node v_i and node v_j . The physical meaning of e_{ij} can be defined according to the concrete conditions. For example, in transportation network, e_{ij} may be defined as the distance between two points.

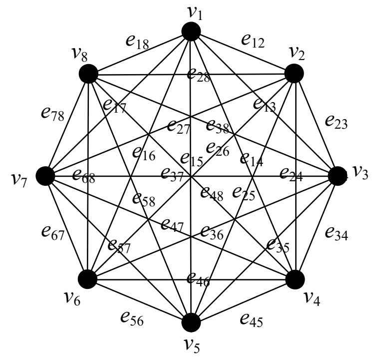


Figure 1: The undirected network graph.

(2) The directed network graph

The existing social network analysis technique has been used to simulate the social trust relationships among individuals, and has been proven to be effective in many GDM problems (Li et al., 2021; Liu et al, 2021). Let $DM = \{DM_1, DM_2, ..., DM_m\}$ be a group of individuals in the social network, the social network can be represented by a directed graph (see Figure 2), in which $DM_l \to DM_h$ signifies that individual DM_l directly trusts DM_h . Let $B = (b_{ij})_{m \times m}$ be an adjacency matrix to represent the trust relationships among individuals. There exists a direct trust relationship from DM_i to DM_j when $b_{ij} = 1$.

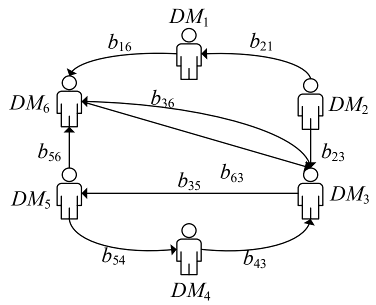


Figure 2: The directed network graph

Social trust network is utilized to describe the directed network graph constructed in this paper based on the trust relationship of FMEA team, which belongs to a type of social network. Individual DM_i can utilized a value between 0 and 1 to represent the trust level from DM_i to DM_j . A great trust value means higher trust degree of DM_i to DM_j . Here, the matrix $B = (b_{ij})_{m \times m}$ is still utilized to denote the trust relationships between individuals and the trust value $b_{ij} \in [0,1]$.

The relative node in-degree centrality index, associated with $DM_k \in DMs$:

$$C(DM_k) = \frac{1}{m-1} \sum_{h=1, h \neq k}^{m} b_{hk} \tag{5}$$

Clearly, a larger $C(DM_k)$ value indicates a higher weight of individual DM_k. Therefore, the weight of DM_k can be obtained using the following equation:

$$\lambda_k = \frac{C(DM_k)}{\sum_{h=1}^{m} C(DM_h)} \tag{6}$$

3. The proposed FMEA approach based on D-S evidence theory and social network

This section formalizes the FMEA issues studied in this paper and introduces the proposed FMEA approach. Figure 3 is the complete flowchart of the proposed FMEA approach.

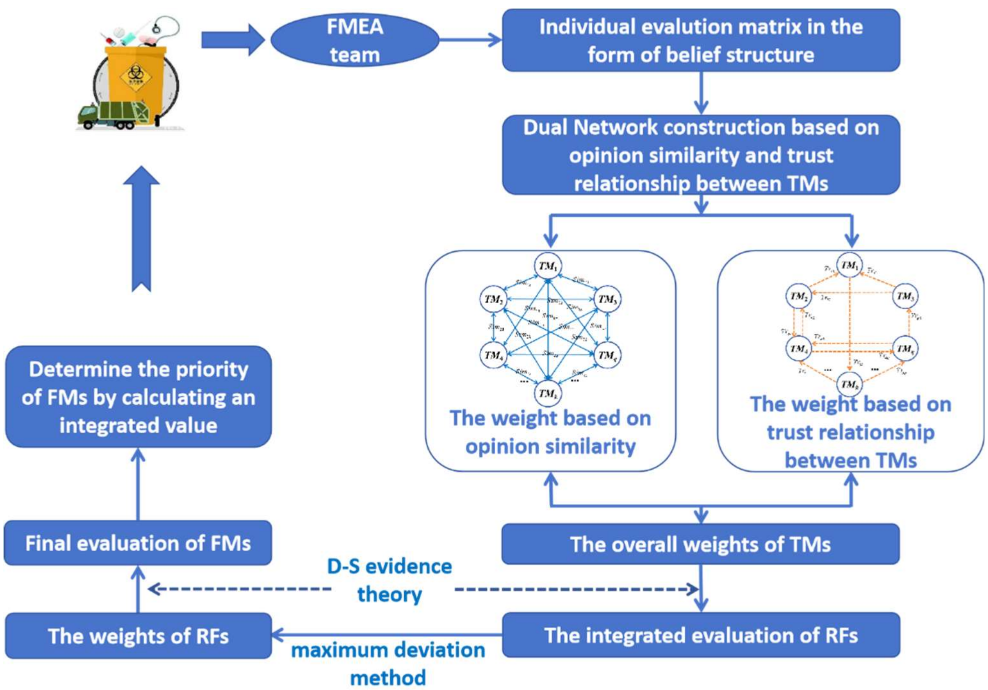


Figure 3: The Complete Flowchart of the Proposed FMEA Approach

3.1 Description of a typical FMEA problem

In a group FMEA decision situation, suppose an FMEA team process consists of q ($q \geq 2$) TMs denoted by $TM = \{TM_1, TM_2, \ldots, TM_q\}$. The weights of TMs can be represented as $w = \{w_1, w_2, \ldots, w_q\}$, where w_k satisfying $w_k > 0$ (k = 1, 2, …, q) and $\sum_{k=1}^{q} w_k = 1$. All TMs are invited to evaluate n potential FMs: $FM = \{FM_1, FM_2, \ldots, FM_n\}$ according to m RFs: $RF = \{RF_1, RF_2, \ldots, RF_m\}$ using the frame of discernment is $\Theta = \{s_\alpha \mid \alpha = 0, 1, \ldots, \tau\}$ (τ is even number).

3.2 The proposed FMEA approach

This section describes the building idea of the proposed group-based FMEA approach which can be broken down into four Steps: (1) processing evaluation information, (2) calculate the weights of TMs from the perspective of social network, (3) calculate the weights of *RFs* and (4) determine the priority of FMs.

Step 1: Processing evaluation information

It is known that the frame of discernment is $\Theta = \{s_\alpha \mid \alpha = 0, 1, \ldots, \tau\}$. The power set of Θ can be expressed as H_g (g=1, 2, …, $2^\tau-1$) in sequence. As a result, the evaluation of FM_i (i=1, 2, …, n) under RF_j (j=1, 2, …, m) provided by TM_k (k=1, 2, …, q) can be represented as $m_{ij}^k(H_g)$ (g=1, 2, …, 2^τ-1), shown as in Table 1.

Table 1 The Evaluation of FM_i under RF_j Provided by TM_k

TM	***RF***	***FM*₁**	**…**	***FMᵢ***	**…**	***FMₙ***
	RF_1	$m_{11}^1(H_g)$	…	$m_{i1}^1(H_g)$	…	$m_{n1}^1(H_g)$
	…	…	…	…	…	…
TM_1	RF_j	$m_{1j}^1(H_g)$	…	$m_{ij}^1(H_g)$	…	$m_{nj}^1(H_g)$
	…	…	…	…	…	…
	RF_m	$m_{1m}^1(H_g)$	…	$m_{im}^1(H_g)$	…	$m_{nm}^1(H_g)$
…	…	…	…	…	…	…
	RF_1	$m_{11}^k(H_g)$	…	$m_{i1}^k(H_g)$	…	$m_{n1}^k(H_g)$
	…	…	…	…	…	…
TM_k	RF_j	$m_{1j}^k(H_g)$	…	$m_{ij}^k(H_g)$	…	$m_{nj}^k(H_g)$
	…	…	…	…	…	…
	RF_m	$m_{1m}^k(H_g)$	…	$m_{im}^k(H_g)$	…	$m_{nm}^k(H_g)$
…	…	…	…	…	…	…
	RF_1	$m_{11}^q(H_g)$	…	$m_{i1}^q(H_g)$	…	$m_{n1}^q(H_g)$
	…	…	…	…	…	…
TM_q	RF_j	$m_{1j}^q(H_g)$	…	$m_{ij}^q(H_g)$	…	$m_{nj}^q(H_g)$
	…	…	…	…	…	…
	RF_m	$m_{1m}^q(H_g)$	…	$m_{im}^q(H_g)$	…	$m_{nm}^q(H_g)$

Step 2: Calculate the weights of TMs from the perspective of social network

Inspired by social networks, the most important thing to construct a network is to define the set of points and the value of edges. In this paper, q TMs are taken as the point set. Considering the opinion interaction and trust interaction among TMs, a dual network structure diagram is constructed based on opinion similarity and trust value, as shown in Figure 4. It can be seen from the graph that the weights of q TMs nodes are respectively solved and weighted by completely undirected networks based on opinion similarity and directed networks based on trust relationship. The specific calculation process is as follows:

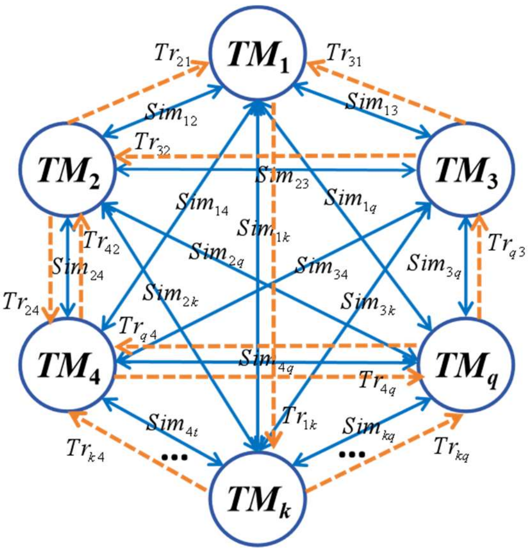


Figure 4: Dual network structure diagram based on opinion similarity and trust relationship

(1) The TMs' weights derived based on opinion similarity

Since the edges are generally expressed as the interactions among TMs, the value of the edges are defined as the similarity between TM_k and TM_t as follows:

$$Sim_{kt} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} Sim(m_{ij}^{k}, m_{ij}^{t})}{mn} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}(1 - d(m_{ij}^{k}, m_{ij}^{t}))}{mn}, k \neq t \tag{7}$$

where $d(m_{ij}^{k}, m_{ij}^{t})$ is the distance of evidence between m_{ij}^{k} and m_{ij}^{t} according to Eq. (3). The smaller the distance, the larger the similarity.

The vertex strength VS_k of node TM_k is defined as the sum of the related edge weight:

$$VS_k = \sum_{t=1, t\neq k}^{q} Sim_{kt}, (k = 1,2,...,q) \tag{8}$$

The weight of TM_k based on the consideration of the opinion similarity can be obtained after the following normalization process:

$$\mu_k = \frac{VS_k}{\sum_{t=1}^{q} VS_t}, (k = 1,2,...,q) \tag{9}$$

(2) The TMs' weights derived based on trust relationship

The matrix $Tr = (Tr_{kt})_{q\times q}$ is still utilized to denote the trust relationships between individuals and the trust value $Tr_{kt} \in [0,1]$.

The relative node in-degree centrality index, associated with $TM_k \in TMs$:

$$C(TM_k) = \frac{1}{q-1}\sum_{t=1, t\neq k}^{q} Tr_{kt} \tag{10}$$

Clearly, a larger $C(TM_k)$ value indicates a higher weight of individual TM_k. Therefore, the weight of TM_k based on the consideration of the trust relationship can be obtained using the following equation:

$$\lambda_k = \frac{C(TM_k)}{\sum_{t=1}^{q} C(TM_t)}, (k = 1,2,...,q) \tag{11}$$

(3) The overall weights

Based on the opinion similarity and trust relationship of TM_k on the network, we can get the overall weight of TM_k:

$$w_k = \alpha\mu_k + (1-\alpha)\lambda_k, (k = 1,2,...,q) \tag{12}$$

in which $\alpha \in [0,1]$ represents the proportion of the weight calculated based on the opinion similarity in the overall weights of TMs.

Evaluations can be further integrated according to the weights of TMs calculated based on social network. To avoid the problem that the Dempster's rule cannot directly handle the combination of highly conflicting evidences, the weighted average belief structure of H_g for TM_k can be obtained as follows:

$$\overline{m}_{ij}(H_g) = \sum_{k=1}^{q} w_k \cdot m_{ij}^k(H_g), \quad g = 1,2,...,2^5 - 1 \tag{13}$$

As a result, the weighted belief structure of TM_k can be expressed as follows:

$$\overline{m}_{ij}(H_g) = \overline{m}_{ij}^1(H_g) = \overline{m}_{ij}^2(H_g) = ... = \overline{m}_{ij}^q(H_g). \tag{14}$$

An integrated belief structure of FM_i with RF_j (shown in Table 2) can be obtained through fusing belief structures of TMs one by one with the use of Eq. (1) for q-1 times.

Table 2 The Integrated Evaluation of FM_i under RF_j

RF	***FM*₁**	**...**	***FMᵢ***	**...**	***FMₙ***
RF_1	$m_{11}^f(H_g)$	$\cdots$	$m_{i1}^f(H_g)$	$\cdots$	$m_{n1}^f(H_g)$
...	...	...	...	...	...
RF_j	$m_{1j}^f(H_g)$	$\cdots$	$m_{ij}^f(H_g)$	$\cdots$	$m_{nj}^f(H_g)$
...	...	...	...	...	...
RF_m	$m_{1m}^f(H_g)$	$\cdots$	$m_{im}^f(H_g)$	$\cdots$	$m_{nm}^f(H_g)$

Step 3: Calculate the weights of *RFs*

In FMEA, if the evaluations of FMs have great difference regarding a RF, then such RF should be assigned a big weight. A larger difference, a relatively important role plays in the priority procedure. The maximizing deviation model which is a popular method in determining criteria weights objectively (Duan et al., 2019) can be used to calculate the *RF*'s weight. Suppose the weight of RF_j is ω_j, where ω_j satisfying $\omega_j > 0$ ($j = 1, 2, \ldots, m$) and $\sum_{j=1}^{m} \omega_j = 1$. The deviation of FM_i from FM_x ($x = 1,2,...,n; x \neq i$) under RF_j can be computed by Eq. (15):

$$Dev_j = \sum_{i=1}^{n} \sum_{x=1, x\neq i}^{n} d(m_{ij}^f, m_{xj}^f) \tag{15}$$

where $d(m_{ij}^f, m_{xj}^f)$ is the distance between m_{ij}^f and m_{xj}^f.

According to the maximizing deviation method, the weights of *RF*s are computed by Eq. (16):

$$\omega_j = \frac{\sum_{i=1}^{n} \sum_{x=1, x\neq i}^{n} d(m_{ij}^f, m_{xj}^f)}{\sum_{j=1}^{m} \sum_{i=1}^{n} \sum_{x=1, x\neq i}^{n} d(m_{ij}^f, m_{xj}^f)}, \ j = 1,2,...,m \tag{16}$$

The weighted average belief structure of H_g for RF_j can be obtained as follows:

$$\overline{m}_i(H_g) = \sum_{j=1}^{m} \omega_j \cdot m_{ij}^f(H_g), \quad g = 1,2,...,2^5 - 1 \tag{17}$$

As a result, the weighted belief structure of RF_j can be expressed as follows:

$$\overline{m}_i(H_g) = \overline{m}_{i1}(H_g) = \overline{m}_{i2}(H_g) = ... = \overline{m}_{im}(H_g) \tag{18}$$

An integrated evaluation of FM_i (shown in Table 3) can be obtained through fusing belief structures of the RFs with Eq. (1) for n-1 times.

Table 3 The Final Evaluation of FM_i

FM	***FM*₁**	**...**	***FMᵢ***	**...**	***FMₙ***
$m^f(H_g)$	$m_1^f(H_g)$	$\cdots$	$m_i^f(H_g)$	$\cdots$	$m_n^f(H_g)$

Step 4: Determine the priority of FMs

With the use of the corresponding score of each level of the linguistic terms as shown in Table 4, an integrated value that represents the overall risk priority of the FM_i can be calculated as follows:

$$u_i = \sum_{g=1}^{2^{\tau}-1} \sum_{\theta \in H_g} \frac{u(\theta) m_i(H_g)}{|H_g|} \tag{19}$$

Then, the complete ranking of FMs can be obtained according to the values. The larger the u_i value, the higher the FM_i ranking, which needs to be paid great attention to.

Table 4. Utility Functions for Assessment Grade

Assessment grade	s_0	...	$s_{\tau/2}$	...	s_τ
$u(*)$	0	...	0.5	...	1.0

4. An illustrative example of infectious medical waste management

Infectious medical waste constitutes a distinct and hazardous category of medical waste that possesses the potential to spread infections through various pathways, thereby endangering patients, medical staff, and the general public. Consequently, it demands significant attention from healthcare institutions and other relevant parties. Employing a professional risk management approach becomes imperative in guiding the proper handling, transportation, and disposal of medical waste. This section employs the hazards and risks identified in a medical waste management system listed in (Makajic-Nikolic et al., 2016), as the FMs in the illustrative example (refer to Table 5). Subsequently, the proposed group-based FMEA model is employed by the TMs to evaluate and rank the FMs based on their perceived hazard levels.

By referring Table 5, 11 FMs are identified and evaluated under O, S, D three RFs. The FMEA team for this problem consists of 6 experts {$TM_1, TM_2, \ldots, TM_6$}. A 5-scale LTS S= {s_0 =very bad, s_1 = bad, s_2 = medium, s_3 = good, s_4 = very good} and belief structures are adopted for evaluation. The original individual evaluation matrices of all TMs can be found in Appendix A. The trust relationship of TMs can be visualized as Figure 5.

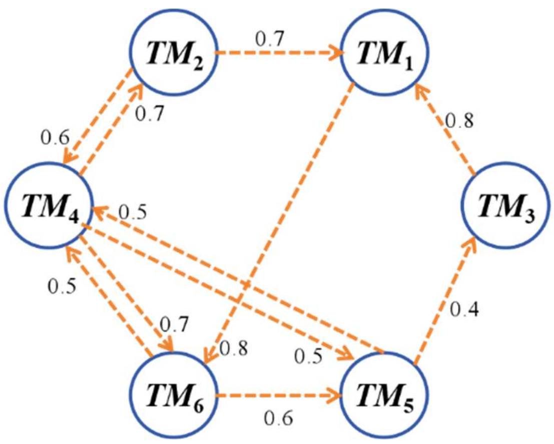


Figure 5: The social network based on trust relationship between TMs

Table 5 Possible hazards and risks in a medical waste management system

No.	Failure mode	Cause of failure	Effect
FM_1	Patient infected	Lack of personal protective equipment; Lack of awareness of the problem; Poor hospital management	Infect other patients and healthcare workers
FM_2	Inadequate education	Inadequate training during employment; Lack of detailed discussion of harmful effects in general education	Mistake in the process of work; Inadequate handling during internal transportation
FM_3	Intentional failure to respect the procedures	Inadequate training during employment; Lack of awareness of harmful effects; Excessive workload	Infection during sterilization
FM_4	Inattention during work, fatigue	Excessive workload; Lack of safety awareness	Not using the protective equipment; Mistake in the process of work; Infection during diagnostics and therapy
FM_5	Injury at work	Inadequate training; Damage to protective equipment; Excessive workload	Infection of staff
FM_6	Not using the protective equipment	Failure to follow procedures; Excessive workload; Insufficient supply of protective equipment	Infection of staff; Inadequate transport conditions
FM_7	Inadequate packaging (waste containers)	Lack of awareness of the importance of packaging	Infection during weighing
FM_8	Inadequate transport conditions	Improper transport organization; Inadequate personnel training; Not using the protective equipment	Infection during transport; Environmental pollution; Infection of the community
FM_9	Spilling or leaking of the waste from container	Lack of awareness of the importance of packaging; Inadequate regular inspections	Infection during unloading
FM_{10}	Contact with infectious waste	Misuse of personnel; Improper segregation of waste; Failure to follow procedures	Infection during transport
FM_{11}	Inadequate packing and handling the waste	Lack of safety awareness	Infection of staff; Environmental pollution; Infection of the community

According to Table A1 in Appendix, the opinion similarity among TMs can be calculated with Eq. (7). Based on the undirected network graph view of the opinion similarity, the weights of TMs are calculated with Eqs. (8) and (9) as follows:

$$\mu_1 = 0.1653; \mu_2 = 0.1634; \mu_3 = 0.1651; \mu_4 = 0.1715; \mu_5 = 0.169; \mu_6 = 0.1657$$

Combining with Figure 5 and according to Eqs. (10) and (11), the weights of TMs are calculated based on the directed network graph view of the trust relationship of TMs as follows:

$$\lambda_1 = 0.2206; \lambda_2 = 0.1029; \lambda_3 = 0.0588; \lambda_4 = 0.2353; \lambda_5 = 0.1618; \lambda_6 = 0.2206$$

According to Eq. (12), the comprehensive weights of TMs are calculated as follows:

$$w_1 = 0.1929; w_2 = 0.1332; w_3 = 0.112; w_4 = 0.2034; w_5 = 0.1654; w_6 = 0.1931$$

According to Eqs. (13) and (14), the integrated belief structure of FM_i with RF_j can be obtained (shown in Table 6.

Table 6 The integrated evaluation of FM_i under RF_j

TM	***FM***	***The integrated evaluation***
O	FM_1	{P},0.0003; {M},0.0001;{G},0.0004;{VG},0.9992
	FM_2	{G},0.0165;{VG},0.9835
	FM_3	{P},0.0008; {M},0.9812;{G},0.0178;{VG},0.0001;{M,G},0.0001
	FM_4	{VP},0.0001;{P},0.0005; {M},0.6969;{G},0.3023;{M,G},0.0003
	FM_5	{M},0.0053;{G},0.1019;{VG},0.8922;{M,G},0.0006
	FM_6	{M},0.0001;{G},0.0026;{VG},0.9974
	FM_7	{G},0.0002;{VG},0.9998
	FM_8	{M},0.006;{G},0.0696;{VG},0.9239;{M,G},0.0005
	FM_9	{M},0.4176;{G},0.3736;{VG},0.1935;{M,G},0.0153
	FM_{10}	{VP},0.0074;{P},0.0204; {M},0.5306;{G},0.4336;{VG},0.0078;{VP,P},0.0001
	FM_{11}	{VP},0.0008;{P},0.001; {M},0.0019; {VG},0.9962;{VP,P},0.0001
S	FM_1	{VP},0.5911;{P},0.4014; {M},0.0076
	FM_2	{P},0.0009; {M},0.9906;{G},0.0081;{VG},0.0003
	FM_3	{VP},0.9666;{P},0.0271; {M},0.0062
	FM_4	{VP},0.9927;{P},0.0004; {M},0.0065;{G},0.0004
	FM_5	{VP},0.0625;{P},0.9054; {M},0.0076;{G},0.0052;{VG},0.0172;{VP,P},0.0021
	FM_6	{P},0.0001; {M},0.9942;{G},0.0031;{VG},0.0026
	FM_7	{M},0.7942;{G},0.2057
	FM_8	{VP},0.0918;{P},0.7607; {M},0.1056;{G},0.0049;{VG},0.034;{VP,M},0.0015;{P,G},0.0012;{M, VG},0.0004
	FM_9	{VP},0.1605;{P},0.0073; {M},0.671;{G},0.1298;{VG},0.0314
	FM_{10}	{VP},0.001;{P},0.1164; {M},0.8825;{G},0.0001
	FM_{11}	{VP},0.0266;{P},0.0062; {M},0.5767;{G},0.0114;{VG},0.3791
D	FM_1	{VP},0.0004;{P},0.9258; {M},0.0739
	FM_2	{P},0.0051; {M},0.5542;{G},0.0131;{VG},0.4275
	FM_3	{P},0.18; {M},0.7613;{G},0.0585;{VG},0.0002
	FM_4	{VP},0.0161;{P},0.9599; {M},0.0119;{G},0.0116;{VG},0.0005
	FM_5	{M},0.9896;{G},0.0102;{VG},0.0002
	FM_6	{P},0.0155; {M},0.8628;{G},0.0011;{VG},0.1206
	FM_7	{P},0.1088; {M},0.8406;{G},0.0222;{VG},0.0284
	FM_8	{P},0.0001; {M},0.9888;{G},0.011;{VG},0.0001
	FM_9	{VP},0.0003;{P},0.7654; {M},0.2317;{G},0.0023;{VG},0.0002
	FM_{10}	{VP},0.0042;{P},0.0065; {M},0.9892;{G},0.0001
	FM_{11}	{P},0.0014; {M},0.6832;{G},0.3154

Table 7 The comprehensive evaluation of FM_i

FM	***The integrated evaluation***
FM_1	{VP},0.1018;{P },0.0.6720; {M},0.0001;{G},0;{VG},0.2261
FM_2	{VP},0.00;{P },0; {M},0.688;{G},0.;{VG},0.312
FM_3	{VP},0.2699;{P },0.0014; {M},0.7287;{G},0.0001;{VG},0.
FM_4	{VP},0.6361;{P },0.2484; {M},0.1065;{G},0.009;{VG},0.
FM_5	{VP},0.0002;{P },0.4806; {M},0.2921;{G},0.0005;{VG},0.2265
FM_6	{VP},0.;{P },0.; {M},0.8749;{G},0.;{VG},0.1251
FM_7	{VP},0.;{P },0.0002; {M},0.8529;{G},0.0032;{VG},0.1438
FM_8	{VP},0.005;{P },0.2917; {M},0.4318;{G},0.0002;{VG},0.2758
FM_9	{VP},0.021;{P },0.1066; {M},0.847;{G},0.0413;{VG},0.003
FM_{10}	{VP},0.;{P },0.0003; {M},0.9953;{G},0.0044;{VG},0.
FM_{11}	{VP},0.;{P },0.; {M},0.4655;{G},0.0056;{VG},0.5289

Based on Table 6, the weights of RFs (O/S/D) are calculated according to Eqs. (15) and (16) as follows:

$\omega_1 = 0.3044; \omega_2 = 0.3942; \omega_3 = 0.3014$

Then, according to Eqs. (17) and (18), the RF's evidence was modified and further integrated to obtain the comprehensive evaluation of each FM, as shown in Table 7.

Finally, the final score of each FM is calculated according to Eq. (19) and Table 8, and the FMs is ranked according to the score size as shown in Table 9.

Table 8. Utility Functions for Assessment Grade

Assessment grade	s_0	s_1	s_2	s_3	s_4
$u(*)$	0	0.3	0.5	0.7	1.0

Table 9 The overall risk priority and the final ranking of FM_i

FM	***The overall risk priority***	***Ranking***
FM_1	0.4278	9
FM_2	0.656	2
FM_3	0.3648	10
FM_4	0.1341	11
FM_5	0.5172	6
FM_6	0.5626	5
FM_7	0.5725	4
FM_8	0.5793	3
FM_9	0.4874	8
FM_{10}	0.5008	7
FM_{11}	0.7656	1

5. Conclusions

The proposed group-based FMEA approach can enhance the reliability of FMEA, which combined with D-S evidence theory and social network for the application of medical waste management, which fully considered the uncertainty of evaluations and the relationship network between TMs. The main advantages can be summarized as follows: (1) Belief structure of D-S evidence theory is used to process and transform evaluation information, which can fully consider the existence of multiple emotional preferences and unknown preferences in evaluation information; (2) A dual network structure is constructed by combining the undirected network graph based on the opinion similarity and the directed network graph based on TMs' trust relationship to obtain the weights of TMs; (3) The weights of RFs from an objective view are calculated by using the maximizing deviation method based on the belief structure; (4) A modified Dempster's combination rule is proposed for evaluation information integration which can distinguish between uncertainty and ignorance while managing conflict between evidence; (5) The effectiveness of the proposed method is verified by a case study, which can get more accurate and practical risk ranking results to help the healthcare industry formulate accurate risk prevention and control plans.

In future work, it is considered to obtain the trust relationship of FMs through data-driven, and further carry out conflict resolution for non-cooperative behavior in FMEA team; Considering complex network relationships in large-scale scenarios, the dynamic opinion interaction mechanism can be further designed and extended to other case backgrounds to provide scientific and effective risk management strategies.

References

Carnero, M. C. (2020). Waste segregation FMEA model integrating intuitionistic fuzzy set and the PAPRIKA method. *Mathematics*, *8*(8): 1375. https://doi.org/10.3390/math8081375

Das, S., Dhalmahapatra, K., & Maiti, J. (2020). Z-number integrated weighted VIKOR technique for hazard prioritization and its application in virtual prototype based EOT crane operations. *Applied Soft Computing*, *94*: 106419. https://doi.org/10.1016/j.asoc.2020.106419

Dempster, A. P. (1967). Upper and Lower Probabilities Induced by a Multivalued Mapping. *Annals of Mathematical Statistics*, *38* (2): 325-339.

Duan, C. Y., Chen, X. Q., Shi, H., & Liu, H. C. (2019). A new model for failure mode and effects analysis based on k-means clustering within hesitant linguistic environment. *IEEE Transactions on Engineering Management*, *69*(5): 1837-1847. https://doi.org/10.1109/TEM.2019.2937579

Fang, H., Li, J., & Song, W. (2020). Failure mode and effects analysis: an integrated approach based on rough set theory and prospect theory. *Soft Computing*, *24*(9): 6673-6685. https://doi.org/10.1007/s00500-019-04305-8

Ghoushchi, S. J., Ghiaci, A. M., Bonab, S. R., & Ranjbarzadeh, R. (2022). Barriers to circular economy implementation in designing of sustainable medical waste management systems using a new extended decision-making and FMEA models. *Environmental Science and Pollution Research*, *29*(53): 79735-79753. https://doi.org/10.1007/s11356-022-19018-z

Ho, C. C., & Liao, C. J. (2011). The use of failure mode and effects analysis to construct an effective disposal and prevention mechanism for infectious hospital waste. *Waste management*, *31*(12): 2631-2637. https://doi.org/10.1016/j.wasman.2011.07.011

Hu, L. (2023). A hybrid generalized TODIM based risk prioritization method for failure mode and effect analysis with linguistic Z-numbers. *Journal of Intelligent & Fuzzy Systems*, *44*(5): 7935-7955. https://doi.org/10.3233/JIFS-223132

Hu, Y. P., You, X. Y., Wang, L., & Liu, H. C. (2019). An integrated approach for failure mode and effect analysis based on uncertain linguistic GRA–TOPSIS method. *Soft Computing*, *23*: 8801-8814. https://doi.org/10.1007/s00500-018-3480-7

Jousselme, A. L., Grenier, D., & Bossé, É. (2001). A new distance between two bodies of evidence. *Information Fusion*, *2* (2): 91-101. https://doi.org/10.1016/S1566-2535(01)00026-4

Li, G. F., Li, Y., Chen, C. H., He, J. L., Hou, T. W., &Chen, J. H. (2020). Advanced FMEA method based on interval 2-tuple linguistic variables and TOPSIS. *Quality Engineering*, *32*(4): 653–662. https://doi.org/10.1080/08982112.2019.1677913

Li, Y., Kou, G., Li, G., & Wang, H. (2021). Multi-attribute group decision making with opinion dynamics based on social trust network. *Information Fusion*, *75*: 102-115. https://doi.org/10.1016/j.inffus.2021.04.010

Li, Y., Liu, P., & Wu, X. (2023). Failure mode and effect analysis approach considering risk attitude of dynamic reference point cumulative prospect theory in uncertainty contexts. *Artificial Intelligence Review*, *56*(12): 14557-14604. https://doi.org/10.1007/s10462-023-10501-8

Liang, D., Li, F., & Xu, Z. (2023). A group-based FMEA approach with dynamic heterogeneous social network consensus reaching model for uncertain reliability assessment. *Journal of the Operational Research Society*, *74*(1): 33-47. https://doi.org/10.1080/01605682.2021.2020694

Liu, P., Wu, Y., Li, Y., & Wu, X. (2024). An improved FMEA method based on the expert trust network for maritime transportation risk management. *Expert Systems with Applications*, *238*: 121705. https://doi.org/10.1016/j.eswa.2023.121705

Liu, Y., Liang, H., Gao, L., & Guo, Z. (2021). Optimizing consensus reaching in the hybrid opinion dynamics in a social network•. *Information Fusion*, *72*: 89-99. https://doi.org/10.1016/j.inffus.2021.02.018

Liu, Z., Bi, Y., & Liu, P. (2023). A conflict elimination-based model for failure mode and effect analysis: A case application in medical waste management system. *Computers & Industrial Engineering*, *178*: 109145. https://doi.org/10.1016/j.cie.2023.109145

Liu, Z., Wang, X., Sun, N., Li, L., Wang, D., & Liu, P. D. (2021). FMEA using the normalized projection-based TODIM-PROMETHEE II model for blood transfusion. *International Journal of Fuzzy Systems*, *23*(6): 1680–1696. https://doi.org/10.1007/s40815-021-01056-3

Luo, X., & Liao, W. (2022). Collaborative reverse logistics network for infectious medical waste management during the COVID-19 outbreak. *International Journal of Environmental Research and Public Health*, *19*(15): 9735. https://doi.org/10.3390/ijerph19159735

Makajic-Nikolic, D., Petrovic, N., Belic, A., Rokvic, M., Radakovic, J. A., & Tubic, V. (2016). The fault tree analysis of infectious medical waste management. *Journal of Cleaner Production*, *113*: 365-373. https://doi.org/10.1016/j.jclepro.2015.11.022

Perumal, S., Aljefri, A. H., & Alsaifi, R. H. (2021). Analyzing the Interactions Among the Barriers for Safe and Effective Medical Waste Management[C]//2021 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM). *IEEE*, 302-307. https://doi.org/10.1109/IEEM50564.2021.9672966

Sarwar, M., Akram, M., & Liu, P. (2021). An integrated rough ELECTRE II approach for risk evaluation and effects analysis in automatic manufacturing process. *Artificial Intelligence Review*, *54*(6): 4449–4481. https://doi.org/10.1007/s10462-021-10003-5

Shafer. (1976). *A mathematical theory of evidence*. Princeton university press, Princeton.

Zheng, Q., Liu, X., & Wang, W. (2022). A likelihood-based ORESTE method for failure mode and effect analysis (FMEA) based risk analysis problem under interval type-2 fuzzy environment. *Quality and Reliability Engineering International*, *38*(1): 304-325. https://doi.org/10.1002/qre.2978

Appendix

Table A1 The Evaluation of FM_i under RF_j Provided by TM_k

TM	FM	O	S	D
TM_1	FM_1	{G},0.1;{VG},0.7;{M,VG},0.2	{VP},0.5;{P},0.2; {M},0.3	{VP},0.2;{P},0.4; {M},0.3;{G},0.1
	FM_2	{P},0.1; {G},0.9	{M},0.4;{VG},0.4;{M,VG},0.2	{VP},0.2;{P},0.2; {M},0.3; VG},0.3
	FM_3	{M},0.8;{P,M},0.2	{VP},0.3;{P},0.2; {M},0.5	{P},0.4; {M},0.2;{G},0.4
	FM_4	{VP},0.1;{P},0.2; {M},0.3;{G},0.4	{VP},0.5;{M},0.2;{VP,P},0.3	{VP},0.1;{P},0.3; {M},0.3;{G},0.2;{VG},0.1
	FM_5	{G},0.2;{VG},0.8	{P},0.4; {M},0.3;{G},0.1;{VG},0.2	{M},0.5;{G},0.2;{VG},0.1
	FM_6	{G},0.2;{VG},0.8	{P},0.3; {M},0.5;{G},0.2	{P},0.3; {M},0.3;{G},0.1;{VG},0.3
	FM_7	{G},0.2;{VG},0.8	{M},0.3;{G},0.4;{VG},0.3	{VP},0.2;{P},0.3; {M},0.3;{G},0.2
	FM_8	{G},0.3;{VG},0.4	{VP},0.5;{P},0.3; {VP,P},0.2	{M},0.5;{G},0.2;{VG},0.1
	FM_9	{M},0.4;{G},0.5	{VP},0.4;{P},0.2; {M},0.2; {VG},0.2	{VP},0.2;{P},0.5; {M},0.2;{G},0.1
	FM_{10}	{VP},0.1; {M},0.5;{VP,P},0.4	{M},0.4;{G},0.2;{M,G},0.4	{VP},0.05;{P},0.05; {M},0.6;{G},0.1;{VG},0.2
	FM_{11}	{VP},0.2;{P},0.2; {VP,P},0.6	{VP},0.2; {M},0.4;{VG},0.2;{VP,M},0.2	{VP},0.1;{P},0.1; {M},0.4;{G},0.3;{M,G},0.1
TM_2	FM_1	{G},0.2;{VG},0.6;{M,G},0.2	{P},0.4;{M},0.2;{P,M},0.4	{P},0.5; {M},0.2; {VG},0.3
	FM_2	{VG},0.9;{M,G},0.1	{M},0.2;{G},0.6;{VG},0.2	{VP},0.2; {P},0.1; {M},0.1; {G},0.3; {VG},0.2; {G,VG},0.1
	FM_3	{VG},1	{M},0.5;{G},0.2;{VG},0.3	{VP},0.2; {P},0.1; {M},0.4; {G},0.2; {M,G},0.1
	FM_4	{VP},0.1;{G},0.3;{VG},0.2	{VP},0.8;{P},0.1;{M},0.1	{VP},0.1; {P},0.4; {M},0.1; {G},0.1; {VG},0.2; {P,VG},0.1
	FM_5	{P},0.1;{G},0.2;{VG},0.7	{VP},0.2;{P},0.3;{VG},0.1;{VP,P},0.4	{P},0.1; {M},0.6; {G},0.3
	FM_6	{P},0.1;{M},0.5;{VG},0.4	{M},0.2;{G},0.2;{VG},0.3;{M,G},0.3	{VP},0.1; {P},0.2; {M},0.4; {G},0.1; {VG},0.2
	FM_7	{VG},0.9;{M,G},0.1	{M},0.2;{G},0.4;{VG},0.3;{M,G},0.1	{P},0.3; {M},0.2; {G},0.1; {VG},0.2; {P,M},0.2
	FM_8	{VG},0.8;{M,G},0.2	{P},0.4;{G},0.1;{P,G},0.5	{VP},0.1; {P},0.05; {M},0.6; {G},0.2; {VG},0.05
	FM_9	{P},0.1; {M},0.3;{VG},0.4;{P,M},0.2	{VP},0.1; {M},0.5;{VG},0.4	{VP},0.2; {P},0.4; {M},0.1; {VG},0.2; {VP,P},0.1
	FM_{10}	{P},0.5; {M},0.5	{P},0.3; {M},0.4; {P,M},0.3	{VP},0.1; {P},0.2; {M},0.5; {G},0.2
	FM_{11}	{VP},0.1;{G},0.3; {VG},0.6	{VP},0.5;{P},0.2; {VP,P},0.3	{P},0.1; {M},0.3; {G},0.5; {VG},0.1
TM_3	FM_1	{VG},0.8;{G,VG},0.2	{VP},0.7;{P},0.3	{VP},0.1; {P},0.3; {M},0.3; {G},0.2; {P,M},0.1
	FM_2	{M},0.2;{G},0.4;{VG},0.4	{G},0.7;{P,M},0.3	{P},0.2; {M},0.3; {G},0.2; {VG},0.3
	FM_3	{G},1	{VP},0.5;{P},0.3; {VP,P},0.2	{VP},0.2; {P},0.1; {M},0.2; {G},0.3; {VG},0.1; {M,G},0.1
	FM_4	{VP},0.5;{P},0.2; {M},0.3	{P},0.2;{M},0.4; {VG},0.1;{P,M},0.3	{VP},0.1; {P},0.35; {G},0.25; {VG},0.1; {P,G},0.2
	FM_5	{G},0.2; {VG},0.4;{M,G},0.4	{VP},0.7; {M},0.3	{P},0.1; {M},0.5; {G},0.2; {VG},0.2
	FM_6	{M},0.2; {VG},0.7;{M,G},0.1	{M},0.4; {G},0.2;{VG},0.4	{P},0.2; {M},0.4; {VG},0.2; {P,M},0.2
	FM_7	{M},0.2; {VG},0.8	{VP},0.1;{P},0.1;{M},0.2; {P,M},0.6	{P},0.3; {M},0.2; {G},0.2; {VG},0.3
	FM_8	{G},0.3;{VG},0.7	{P},0.3;{VG},0.2;{P,VG},0.5	{M},0.5; {G},0.2; {M,G},0.3
	FM_9	{M},0.2; {VG},0.7;{M,G},0.1	{M},0.2; {G},0.3; {VG},0.3;{M,G},0.2	{P},0.5; {M},0.4; {G},0.1
	FM_{10}	{M},0.2;{G},0.2;{VG},0.5;{M,G}, 0.1	{VP},0.6;{P},0.3;{M},0.1	{VP},0.1; {P},0.2; {M},0.4; {G},0.3
	FM_{11}	{M},0.3; {VG},0.7	{M},0.2; {G},0.1; {VG},0.7	{VP},0.05; {P},0.1; {M},0.4; {G},0.4; {M,G},0.05
TM_4	FM_1	{G},0.1; {VG},0.7;{P,M},0.2	{VP},0.6;{P},0.3;{M},0.1	{VP},0.1; {P},0.6; {M},0.3
	FM_2	{G},0.2;{VG},0.8	{VP},0.1;{P},0.3;{M},0.6	{P},0.1; {M},0.4; {G},0.15; {VG},0.35
	FM_3	{M},0.7;{P,M},0.3	{VP},0.7;{P},0.1; {VP,P},0.2	{VP},0.1; {P},0.4; {M},0.3; {VG},0.2
	FM_4	{M},0.3; {G},0.4;{M,G},0.3	{VP},0.6;{G},0.2; {VG},0.1;{G,VG},0.1	{VP},0.4; {P},0.3; {M},0.3
	FM_5	{M},0.1;{G},0.3;{VG},0.6	{G},0.4; {VG},0.6	{VP},0.1; {M},0.55; {G},0.15; {VG},0.2
	FM_6	{G},0.5;{VG},0.5	{P},0.2; {M},0.5;{VG},0.3	{P},0.1; {M},0.4; {G},0.15; {VG},0.35
	FM_7	{G},0.3;{VG},0.7	{P},0.1;{M},0.3;{G},0.2; {M,G},0.4	{P},0.1; {M},0.4; {G},0.25; {VG},0.25
	FM_8	{M},0.2; {VG},0.4;{M,G},0.4	{M},0.1; {VG},0.3;{M,VG},0.6	{VP},0.1; {M},0.55; {G},0.15; {VG},0.2
	FM_9	{M},0.2; {VG},0.6;{M,G},0.2	{VP},0.4;{P},0.2;{M},0.1;{G},0.3	{VP},0.1; {P},0.4; {M},0.4; {G},0.1
	FM_{10}	{G},0.9;{M,G},0.1	{VP},0.2; {P},0.4; {M},0.4	{VP},0.3; {P},0.1; {M},0.5; {G},0.1
	FM_{11}	{P},0.1; {M},0.3; {VG},0.6	{P},0.2;{M},0.1;{VG},0.5; {M,VG},0.2	{P},0.2; {M},0.5; {G},0.3
TM_5	FM_1	{M},0.1;{G},0.3;{VG},0.6	{VP},0.1; {P},0.4; {M},0.3; {G},0.2	{VP},0.2; {P},0.4; {M},0.3; {VG},0.1
	FM_2	{VG},0.7;{M,G},0.3	{M},0.7; {G},0.3	{P},0.1; {M},0.4; {G},0.2; {VG},0.2; {M,G},0.1
	FM_3	{P},0.2; {M},0.5;{VP,P},0.3	{VP},0.7; {P},0.3	{P},0.4; {M},0.3; {G},0.1; {VG},0.1; {P,M},0.1

	FM_4	{P},0.2; {M},0.3;{VG},0.3;{P,M},0.2	{M},0.2; {G},0.2; {VG},0.3; {M,G},0.3	{VP},0.3; {P},0.3; {M},0.2; {G},0.2
	FM_5	{M},0.2;{G},0.4;{VG},0.4	{P},0.5; {VP,P},0.5	{P},0.1; {M},0.6; {G},0.2 {VG},0.1
	FM_6	{VG},0.8;{M,G},0.2	{P},0.1; {M},0.4; {G},0.5	{P},0.1; {M},0.4; {G},0.1; {VG},0.2; {M,G},0.1
	FM_7	{G},0.4;{VG},0.6	{M},0.5; {G},0.5	{P},0.1; {M},0.4; {G},0.1; {VG},0.2; {M,G},0.1
	FM_8	{M},0.3;{G},0.4;{VG},0.3	{P},0.3; {G},0.2; {P,G},0.5	{P},0.1; {M},0.5; {G},0.1; {VG},0.3
	FM_9	{VG},0.4;{M,G},0.6	{M},0.3; {G},0.2; {VG},0.3; {M,G},0.2	{P},0.2; {M},0.3; {G},0.2; {VG},0.3
	FM_{10}	{G},0.4;{VG},0.6	{P},0.4; {M},0.3; {P,M},0.3	{VP},0.1; {P},0.4; {M},0.5
	FM_{11}	{VG},1	{VP},0.1; {P},0.3; {G},0.4; {VG},0.2	{VP},0.1; {P},0.2; {M},0.4; {G},0.3
TM_6	FM_1	{VP},0.4;{P},0.6	{VP},0.5; {P},0.4; {VP,P},0.1	{VP},0.1; {P},0.4; {M},0.3; {G},0.2
	FM_2	{VG},1	{P},0.1; {M},0.4; {G},0.2; {P,M},0.3	{P},0.2; {M},0.3; {G},0.1; {VG},0.4
	FM_3	{G},0.4;{M,G},0.6	{VP},0.4; {P},0.3; {M},0.2; {P,M},0.1	{P},0.1; {M},0.5; {G},0.3; {VG},0.1
	FM_4	{M},0.4;{G},0.3;{M,G},0.3	{VP},0.8; {P},0.1; {M},0.1	{VP},0.1; {P},0.35; {M},0.1; {G},0.15; {VG},0.2; {P,G},0.1
	FM_5	{M},0.1;{VG},0.3;{M,G},0.6	{P},0.2; {M},0.4; {G},0.3; {VG},0.1	{VP},0.1; {M},0.4; {G},0.3; {VG},0.2
	FM_6	{G},0.3;{VG},0.7	{M},0.7; {VG},0.1; {M,VG},0.2	{P},0.2; {M},0.3; {G},0.1; {VG},0.4
	FM_7	{P},0.2; {VG},0.8	{VP},0.2; {M},0.5; {G},0.3	{P},0.3; {M},0.3; {G},0.1; {VG},0.3
	FM_8	{VG},0.7;{G,VG},0.3	{VP},0.1; {M},0.1; {VP,M},0.8	{VP},0.1; {P},0.5; {M},0.2; {G},0.2
	FM_9	{G},0.3;{VG},0.3;{M,G},0.4	{VP},0.3; {P},0.3; {M},0.2; {G},0.2	{P},0.2; {M},0.4; {G},0.3; {VG},0.1
	FM_{10}	{VP},0.3;{P},0.2; {M},0.5	{VP},0.3; {P},0.3; {M},0.4	{VP},0.4; {P},0.3; {M},0.3
	FM_{11}	{M},0.5; {VG},0.5	{M},0.4; {G},0.2; {VG},0.1; {M,G},0.3	{VP},0.15; {P},0.15; {M},0.3; {G},0.3; {VG},0.1

Two-stage Selection of International Flights by Passengers in China for Official Business Purpose using Logit Model

Jiangying Qin [1*], Huan Song [1], Shawei He [1,2] and Rafael Bitoman [1]
[1] College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing, China
[2] College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, China
[*] qinjiangying52@163.com

Abstract

Compared to leisure travel, official business travel tends to contribute to higher revenue per seat, i.e. higher yields. Hence, official business-purpose passengers, being high-value customers for airlines, have been paid high attention for airlines to further extend their revenues in international flights. The objective of this research is to explore the factors influencing the selection of international flights for official business purposes in China constrained by corporate travel policies defined by governments. The study proposes a two-stage framework: in the first stage the corporate travel policies are transformed into attribute threshold constraints for forming a choice set. The second stage incorporates the choice set in a discrete choice model to estimate the behaviour of selecting international flights. The framework can effectively model the behaviour of official business travellers in screening flights that comply with policies before making their selection, enabling an assessment of the impact of such policies on their decision-making behaviour. A survey, based on the stated preference method, is designed to collect passenger preferences in terms of the distribution of choice for connecting flights in the Nanjing Metropolitan Area. This proposed two-stage approach is examined using the survey data for official business-purpose passengers affiliated with universities in the Nanjing Metropolitan Area.

Keywords: Official business passengers; International flight choice; Two-stage model; Error components logit

1. Introduction

Official business passengers are high-value customers for airlines: their behaviour of selecting international flights has been paid increasing attention in the fierce competition along the lucrative international routes. Traditionally, passengers are categorized into three categories, Business, Leisure and VFR (visiting friends and relatives) (Dresner, 2006). According to demand information of the passengers for business purposes, airlines design and sell their tickets by observing the spending behaviour of passengers, with the aim of capturing market share from competitors and maximizing their revenue. In comparison to those traveling for personal reasons, official business passengers often have tighter time constraints and less sensitive to ticket prices. Compared with leisure travel, official business passengers typically possess higher spending capacity and willingness, contributing to higher revenue for airlines. Hence, there is an incentive for airlines to further explore the choice behaviour of passengers in the official business travel category. Moreover, due to the restrictions of strict regulations on official business travel, official business passengers need to consider more factors in the process of choosing routes, compared with private travel, and the range of alternatives for them is also greatly reduced, which makes it difficult for them to quickly and accurately choose the right travel route

The practical motivation of this study is to explore the factors influencing the selection of international flights for official business purposes in the Nanjing Metropolitan Area. The Nanjing Metropolitan Area is an economic region centred around Nanjing. Located in the core area of the cities along the middle and lower reaches of the Yangtze River, it spans the provinces of Jiangsu and Anhui. With Nanjing as the leading city, it includes seven member cities: Zhenjiang, Yangzhou, Huai'an, Ma'anshan, Chuzhou, Wuhu, and Xuancheng. These seven cities are all within approximately 100 kilometres of Nanjing. The total area of the Nanjing Metropolitan Area is 63,000 square kilometres, and it is home to over 90 universities.

In the literature on route selection, most of the contributions in the literature are based on the Random Utility Theory, and may be classified in models that simulate airport choice alone and models that simulate a combination of two (airport and airline; airport and access mode) or three choice dimensions (airport, airline and access mode) (Murakami, 2012; de Luca, 2012; Hess and Polak, 2006a, 2006b). The discrete choice models used have evolved from the simplest structures like the Multinomial Logit models to more complex models capable of handling the cross relations among the various flights (Hierarchical Logit models – HL; Cross-Nested Logit models – CNL), and ones that address heterogeneity between different passenger groups (Mixed Multinomial Logit models – MMNL) (Marcucci, 2011; de Luca and Di Pace, 2012; Pels,2009; Yang et al., 2014; Yang et al., 2017).

As theories have evolved and matured, some scholars have questioned the single-stage utility models. Simon (1956) proposed the need for a choice mechanism that represents the desire of organisms to ‘‘satisfice’’ rather than to optimize in a seminal paper. The satisficing idea of Herbert Simon is present in various choice models that use thresholds and heuristics. Manski (1977) developed the foundational framework for a two-stage model, which posits that the decision-making process of individuals occurs in two stages. In the first stage, decision-makers select a subset of alternatives the universal realm of alternatives to form a choice set. In the second stage, individuals then choose the optimal alternative from the choice set. Building upon Manski's (1977) framework, the two-stage model has evolved from its simplest structure to a semi-compensatory model with a flexible error structure (Kaplan et al.,2012). It has also incorporated the Constraint Multinomial Logit models (CMNL) that consider implicit choice sets (Martínez,2009). Furthermore, Rashedi (2019) proposed a semi-compensatory framework for discrete choice models, combining the Independent Availability Logit (IAL) with the Constrained Multinomial Logit (CMNL) model. These models find widespread application in destination selection for tourism and residential address choices (Kaplan et al.,2012; Rashedi, 2019).

Previous research has predominantly focused on business and leisure passengers. However, the factors influencing the international flight choices of passengers for official business purpose are currently not well understood. Meanwhile, the majority of literature analysing flight selection employs single-stage models such as Multinomial Logit models, Nested Logit models, Mixed Logit models, and others. Due to the constraints imposed by corporate travel policies, official business travellers often filter out flights that comply with the regulations before making their selection. In this scenario, using a single-stage model would be evidently inappropriate. However, two-stage choice models are typically applied in the context of destination selection for tourism and residential address choices, receiving less attention in the context of flight choice behaviour. In this paper, a novel two-stage model is developed for the first time to analyse the selection of international flights by passengers, which is a contribution of this paper. In addition, another contribution in this paper is that in order to conform to the singleness of the constraint on official business travel, the probability formula for an alternative to be taken into account in the choice set is redefined.

There are five sections in this paper. The research motivations and the relevant literature are discussed in the Introduction. In the second section, the methodology proposed in this study is discussed. The third section covers the research framework of SP experiments and the data specifications. The fourth section explains and discusses the model estimation results. The conclusions, study limitations and future research directions are contained in the last section.

2. Two-stage model formulation

In this section, the structure of the two-stage model is illustrated and its formulation is explained. The general form of a two-stage model is presented firstly. Then the formulation in the stage for forming choice set and the choice stage is explained.

The two-stage framework is proposed in this study: in the first stage the corporate travel policies are transformed into attribute threshold constraints for forming a choice set S from the universal realm of alternatives G. In the second stage, passengers then choose the optimal alternative from the choice set. The descriptions of the notations are provided in Table 1.

Table 1.The descriptions of the notations

Notation	Description	Value range
i、j	Alternative number	$1 \leq i$、$j \leq I$
k、l	Attribute number of the alternative	$1 \leq k$、$l \leq K$
q	Passenger number	$1 \leq q \leq Q$
m	Category number	$1 \leq m \leq M$
P	Choice probability	$0 \leq P \leq 1$
S	Choice set	-
G	Universal realm of alternatives	-
w	Scale parameter	$-\infty \leq w \leq +\infty$
t^*	Attribute threshold for the alternative	$-\infty \leq t^* \leq +\infty$

The general form of a two-stage model, is expressed by Manski (1977) as:

$$P_q(i|G) = \sum_{S \sqsubset G} P_q(i|S) P_q(S|G) \,. \tag{1}$$

Here G denotes the universal realm of alternatives and S denotes a choice set, which can be any subset of G. The probability that an individual $q(q = 1,2, \ldots, Q)$ chooses alternative i from G depends on the probability that this individual has the choice set $P_q(S|G)$, and the probability that the alternative i is chosen from that choice set $P_q(i|S)$. The latter probability is generally modelled with the Utility Theory. The random error term introduced in the utility function can account for differences in personal preferences, decision contexts, and other unobserved factors. Therefore, the Utility Maximization model is capable of capturing individual variations, making it the most prominent rational decision-making model. To model the probability of considering alternatives $P_q(S|G)$, usually a non-compensatory heuristic is assumed.

The two-stage process, which is illustrated in Figure 1, largely agrees with the behavioural two-stage choice process described by Kaplan et al. (2012). As official business travel activities are mandatory, the option to cancel travel plans is not considered in this paper.

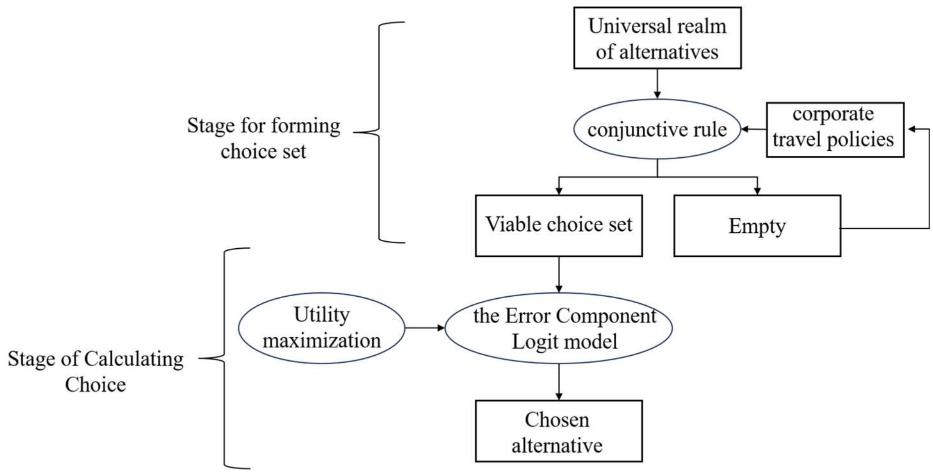


Figure 1: Two-stage decision process

2.1. Choice set formation stage

In official business travel, requirements on selecting the flight routes are issued by governments, such as requiring the layover times for international trips no more than 24 hours. Assuming the requirements consist of K attributes of the alternatives. Designating the workplace-specified standard values as t_k^*, an alternative can be considered by the passenger only if all levels of its attributes fall within the specified standards. Assuming that passengers apply a conjunctive rule to form their viable choice set (Hauser, 2014). This rule states that an alternative is considered if all of its levels are at an acceptable level:

$$I_i = \prod_k I(x_{ik} \le t_k^*) = 1\,, \tag{2}$$

in which x_{ik} denotes the level of attribute k that is present in the alternative i, and t_k^* is the threshold value. The indicator function I equals 1 if the level is below the threshold (i.e. acceptable), and zero otherwise. According to the above assumptions, the probability of the alternative i being included in the choice set S by the passenger can be expressed by the following formula:

$$P_i = \Pi_{l=1}^{L}\theta_{il} = \prod_{l=1}^{L}\frac{1}{1+exp\left(w_l(x_{il}-t_l^*)\right)}, \tag{3}$$

in which x_{il} denotes the level of attribute l that is present in alternative i, and w_l is the scale parameter ($w_l > 0$ if t_l^* is the up bound; else $w_l < 0$). The magnitude of the scale parameter reflects the tolerance for violating the threshold. With an increase in the scale parameter w_l , the likelihood of an attribute being included in the choice set decreases when it violates the threshold. If there is no upper or lower bound for a given attribute i.e. $t_l^* \to +\infty(t_l^* \to -\infty)$,then $\theta_{il} \to 1$ since the value of the exponential function in the denominator is equal to zero. When $(x_{il} - t_l^*) \to \infty$, i.e., when the constrained attribute level is consistently higher or lower than the threshold, $P_i \to 0$. Then the probability of considering a choice set S as $P_q(S|G)$ is defined as (Swait and Ben-Akiva, 1987):

$$P_q(S|G) = \frac{\prod_{i\in S} P_i \prod_{n\in(G-S)}(1-P_n)}{1-\prod_{g\in G}\left(1-P_g\right)}. \tag{4}$$

2.2. Stage of Calculating Choice

In the choice stage, the Error Component Logit model (ECL) is adopted, which can effectively handle correlations between alternative choices. The error term in the utility function for alternative i can be divided into two parts: $\varepsilon_i$and E_m. ε_i is assumed to be an identical extreme value type I distribution, like the MNL or NL model. E_m is the common part of the random error components of the alternatives in group m and is assumed to be normally distributed with $\mathrm{N}(0,1)$. The utility function of the ECL model is expressed as follows (Marcucci and Gatta, 2011):

$$U_{iq} = \boldsymbol{\beta_q}'\ \boldsymbol{X_{iq}} + \varepsilon_{iq} + \sum_{m\in M} d_{im}\delta_m E_{mq}\,, \tag{5}$$

where the constant d_{im} is equal to 1 if E_{mq} appears in the utility function U_{iq} of the alternative i, and 0 otherwise. δ_m is the corresponding parameter to be estimated. X_{iq} is a column vector consisting of factors influencing passenger choices, including alternative attributes, airport attributes, and individual socio-economic attributes, where $\boldsymbol{\beta_q}$ is the corresponding column vector of estimated parameters. The probability expression for the ECL model is expressed as follows (Marcucci and Gatta, 2011):

$$P_q(i) = \int \frac{e^{\left(\boldsymbol{\beta_q}'\ X_{iq}+\sum_{m=1}^{M}\delta_m E_{mq} d_{im}\right)}}{\sum_{j\in S_q} e^{\left(\boldsymbol{\beta_q}'\ X_{jq}+\sum_{m=1}^{M}\delta_m E_{mq} d_{jm}\right)}} f\left(E_{mq}\right) d\mathrm{E}_{\mathrm{mq}}. \tag{6}$$

According to the expression (6), given the choice set S, the conditional choice probability for the alternative i can be formulated as follows:

$$P_q(i|S) = \int \frac{I_i e^{\left(\boldsymbol{\beta_q}' \, \boldsymbol{X_{iq}} + \sum_{m=1}^{M} \delta_m E_{mq} d_{im}\right)}}{\sum_{j \in S_q} I_j e^{\left(\boldsymbol{\beta_q}' \, \boldsymbol{X_{jq}} + \sum_{m=1}^{M} \delta_m E_{mq} d_{jm}\right)}} f(E_{mq}) d\mathrm{E}_{\mathrm{mq}}. \tag{7}$$

The formation of the choice set and the choice of alternatives from that set are distinct cognitive stages. Although passengers' choices are based on the retained choice set, the error terms in the formation of the choice set and the selection stage are uncorrelated. Therefore, the likelihood function for passengers choosing their preferred alternative i from the universal realm of alternatives G can be expressed as:

$$L(\beta; \delta) = \prod_{q=1}^{Q} \prod_{i=1}^{|G|} [P_q(i|G)]^{d_{iq}}, \tag{10}$$

where d_{qi} is an indicator function, equal to 1 if passenger q chooses the alternative i, and 0 otherwise; $|G|$ is the total number of alternatives; Q is the total number of passengers; $(\beta; \delta)$ are the parameters to be estimated. Since the model above lacks a closed-form expression, the Maximum Simulated Likelihood (MSL) method is employed for estimation.

3. Data Collection

This model is applied to the scenario of flight choice among university faculty and staff in the Nanjing Metropolitan Area. The connecting flights from the Nanjing Metropolitan Area to Toulouse, a city in France, are focused on in this paper. There are five civilian airports in the Nanjing Metropolitan Area: Nanjing Lukou International Airport (NKG), Yangzhou Taizhou International Airport, Huai'an Lianshui International Airport, Wuhu Xuanzhou Airport, and Zhenjiang Danyang General Aviation Airport. Only NKG airport is considered in this study due to the limited number of flights from the four airports other than NKG airport to Toulouse. Additionally, as NKG airport currently has few direct flights to European cities, passengers often travel to Shanghai Pudong International Airport (PVG) to take international flights and reduce the number of transfers. PVG airport is one of China's three major hub airports and is the largest hub airport in the eastern part of China, with numerous international routes. NKG airport and PVG airport are considered as the departure airports for official business travellers affiliated with universities in the Nanjing Metropolitan Area., and Toulouse Blagnac Airport (TLS) is considered as the arrival airport. Considering passengers' aversion to connecting flights, the study focuses on flights with the minimum number of transfers from each departure airport. Specifically, for the NKG-TLS, two connecting flights are considered, and for the PVG-TLS, one connecting flight is considered.

The Stated Preference (SP) method is utilized to model the flight choice behaviour of official business travellers. The entire questionnaire can be divided into three parts. The first part investigates the socio-economic characteristics of official business travellers, including age, monthly income, and frequent flyer status. The second part is to investigate the previous experiences of passengers traveling to Europe. The last part involves determining SP scenarios to investigate the preferences of official business passengers among different flights. The structure of the questionnaire is illustrated in Figure 2.

The universal realm of alternatives includes ten flight routes in SP scenario. Each alternative is composed of four flight attributes (air fare, flight departure time, air journey time, and access time) and four transfer attributes (connection time, whether the airlines on the route belong to the same alliance, whether luggage is through-checked, and whether to transfer in the same terminal). Finally, a policy attribute indicating whether the airline is domestic is introduced. This is because official business travellers are usually required to prioritize flights operated by domestic airlines. The attribute values are designed based on the reality of the aviation market, then the orthogonal design will be adopted to generate many scenarios and they will be randomly distributed into different types of questionnaires. The surveyed population consists of faculty and staff from various universities within the Nanjing Metropolitan Area.

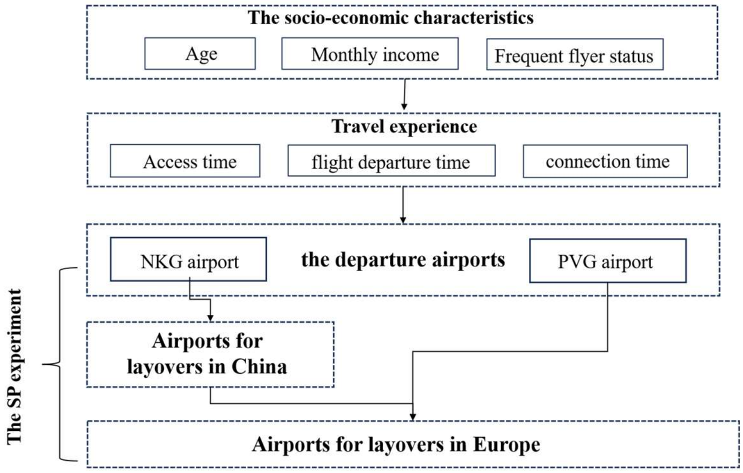


Figure 2: The structure of the questionnaire

4. Results and Discussion

In this section, firstly, the MNL model is employed to determine the combination of exploratory variables and specify them within the two-stage model to estimate the coefficients. According to the estimated results of the model, official business travellers exhibit a higher sensitivity to air journey time and the number of layovers compared to access time, connection time, and air fare. More passengers choose to take flights from Shanghai Pudong International Airport to reduce the number of layovers, even if the access time of that airport is longer. Official business passengers are more inclined to choose routes operated by domestic Chinese airlines and those departing in the morning, aligning with the corporate travel policies defined by governments. Additionally, these three influencing factors: whether the airlines on the route belong to the same alliance, whether luggage is through-checked, and whether to transfer in the same terminal are not significant. This indicates that for international flight routes requiring layovers, official business travellers prioritize whether the route complies with the policies over the convenience of layover services.

Managerial implications of results from this paper mean that airlines can identify those flight attributes that are significant for specific groups of business passengers, and therefore create products with different service levels that are appealing to official business passengers. airport managers and airline operators need to adopt more aggressive strategies to attract official business travellers. For example, Shanghai Pudong International Airport can consider air and rail transport to shorten the access time of passengers.

5. Conclusion and Further Study

This study models the international flight choice behaviour of official business-purpose passengers. A novel application using the Two-stage model is proposed to explore the factors influencing the selection of international flights for official business purposes in China constrained by corporate travel policies defined by governments. An SP experiment with linking variables is designed to collect travellers' preferences for international flight routes in the Nanjing Metropolitan Area. According to the model results of this study, compared to single-stage models, the two-stage method proposed in this study can better simulate the actual situation where official business travellers first filter before making a selection and the fitting degree has improved.

The current model has several limitations that indicate exciting and challenging future research directions towards the realization of the full potential of two-stage models. Firstly, the current study assumes the same model structure for the entire population, while recent studies suggest that the model structure may vary across population segments. Hence, an interesting future research direction is the consideration of latent class models and population segmentation in the two-stage choice. Secondly, the current study postulates utility maximization at the choice stage, while studies indicate the existence of other decision rules such as attribute dominance and regret minimization. Consequently, a natural future research direction is the consideration of alternative decision rules at the stage of calculating choice.

References

De Luca, S., Di Pace, R. (2012). Modelling passenger departure airport choice: implicit vs. explicit approaches[J]. Procedia-Social and Behavioural Sciences 54, 875-885. https://doi:10.1016/j.sbspro.2012.09.803

Dresner, M. (2006). Leisure versus business passengers: Similarities, differences, and implications[J]. Journal of Air Transport Management, 12(1),28-32. https://doi.10.1016/j.jairtraman.2005.09.006

Hauser, J. R. (2014). Consideration-set heuristics. Journal of Business Research, 67(8), 1688-1699. https://doi.10.1016/j.jbusres.2014.02.015

Hess, S., & Polak, W.J. (2006a). Exploring the potential for cross-nesting structures in airport-choice analysis: a case-study of the Greater London area. Transp. Res. Part E 42, 63-81. https://doi.10.1016/j.tre.2005.09.001

Hess, S., & Polak, W.J. (2006b). Airport, airline and access mode choice in the San Francisco Bay Area. Pap. Reg. Sci. 85 (4), 543-567. https://doi.10.1111/j.1435-5957.2006.00097.x

Kaplan, S., Shiftan, Y., & Bekhor, S. (2012). Development and estimation of a semi-compensatory model with a flexible error structure[J]. Transportation Research Part B: Methodological, 46(2), 291-304. https://doi.10.1016/j.trb.2011.10.004

Keumi, C., & Murakami, H. (2012). The role of schedule delays on passengers' choice of access modes: a case study of Japan's international hub airports. Transp. Res. Part E 48, 1023-1031. https://doi.10.1016/j.tre.2012.03.005

Marcucci, E., & Gatta, V. (2011). Regional airport choice: Consumer behaviour and policy implications[J]. Journal of Transport Geography, 19(1), 70-84. https://doi.10.1016/j.jtrangeo.2009.10.001

Marcucci, E., & Gatta, V., (2012). Dissecting preference heterogeneity in consumer stated choices. Transp. Res. Part E 48 (1), 331-339. https://doi.10.1016/j.tre.2011.08.003

Manski, C. F. (1977). The structure of random utility models[J]. Theory and decision, 8(3), 229. https://doi.10.1007/bf00133443

Martínez, F., Aguila, F., & Hurtubia, R. (2009). "The Constrained Multinomial Logit: A Semi-Compensatory Choice Model." Transportation Research Part B: Methodological 43 (3): 365-377. https://doi. 10.1016/j.trb.2008.06.006

Pels, E., Njegovan, N., & Behrens, C. (2009). Low-cost airlines and airport competition[J]. Transportation Research Part E: Logistics and Transportation Review, 45(2), 335-344. https://doi.10.1016/j.tre.2008.09.005

Rashedi, Z., & Nurul Habib, K. (2020). A semi-compensatory choice model with probabilistic choice set: combining implicit choice set within probabilistic choice set formation[J]. Transportmetrica A: Transport Science, 16(3), 1512-1530. https://doi.10.1080/23249935.2020.1821153

Simon, H. A. (1956). Rational choice and the structure of the environment[J]. Psychological review, 63(2), 129. https:// doi.10.1037/h0042769

Swait, J., & Ben-Akiva, M. (1987). "Incorporating Random Constraints in Discrete Models of Choice set Generation." Transportation Research Part B: Methodological 21 (2): 91-102. https://doi:10.1016/01912615(87)90009-9

Yang, C. W., & Liao, P. H. (2016). Modelling the joint choice of access modes and flight routes with parallel structure and random heterogeneity[J]. Transportation Research Part E: Logistics and Transportation Review 95, 19-31. https://doi.10.1016/j.tre.2016.09.001

Yang, C. W., & Wang, H. C. (2017). A comparison of flight routes in a dual-airport region using overlapping error components and a cross-nested structure in GEV models[J]. Transportation Research Part A: Policy and Practice 95, 85-95. https://doi. 10.1016/j.tra.2016.11.006

Equilibrium analysis of selecting network structure for airlines in the foresight of two steps

Peng Sun [1*], Jiangying Qin [1], Huixin Yang [1], Shawei He [1,2], Naiming Xie [1,2]

[1] College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing, China
[2] College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, China
[*] Sunpeng@nuaa.edu.cn

Abstract

The selection of network structures is instrumental for airlines to gain advantage in market competition. In the existing research, game theoretical methodologies are utilized to provide equilibria for suggesting strategies in competition. However, these equilibria only reflect the foresights of decision makers (DMs) in single step, therefore neglecting possible counteractions from opponent. In this paper, the selection of network structures between two competing airlines are investigated using graph model for conflict resolution (GMCR) by considering the perception of not only the own possible actions for respective DMs, but also subsequent moves from their opponents. Theorems are provided to indicate the existence of equilibria reflecting the foresight for DMs of two steps, as well as the comparison with the Nash equilibria calculated by the classical game theoretical approach. As demonstrated in this paper, GMCR can provide wider range of equilibria compared with those calculated by classical game theory, which can suggest enhanced understanding for airlines so as to gaining upper hand in competition.

Keywords: Network structure; equilibrium; graph model for conflict resolution; foresights

1. Introduction

Network for airlines in operation consists of airports and flight routes connecting airports. The selection of network structure by airlines is instrumental for gaining profit and prevailing in the competition with rival airlines. Two types of network structures are generally adopted in operation, Point-to-Point (PP) and Hub-Spoke (HS) networks. In the former network, each pair of airports are connected by routes with same superiority. In comparison, an HS network is organized by spoke routes connecting respective airports with few larger airports called central hubs. Thus, each pair of origin and destination airports can be connected via hub airports. While passengers can enjoy shorter distance of travel by choosing flights in PP network, the HS structure requires fewer routes for connecting the same number of airports compared with PP network. In practice, larger airlines intend to establish international network using HS structure, while smaller carriers, especially low-cost carriers (LCCs), rely on PP network to offer competitive prices and to achieve shorter connection time. By considering the respective advantages and drawbacks of the two networks, emerging airlines in the market ought to carefully decide which type of network it should choose for more profits while gaining an upper hand in competing with rival airlines. In a competitive market particularly, while making the business strategies, airlines often contemplate the counteractions from rival airlines competing over the share of flow along routes. For example, Hainan Airlines of China establishes direct flight to Berlin from Beijing to avoid the disadvantage of competing with larger carriers, such as Lufthansa and Air China, over passengers flying between China and Germany. Their overseas destinations, such as Brussel and Manchester, are not the largest hubs in the respective regions. Thus, the business philosophy for Hainan Airlines of building up their networks is based on the differentiated market strategies by concentrating not only the profitability of routes but also its relative competitiveness with rival airlines.

The study on the selection of air networks for airlines has been carried out from the perspectives of single airlines and multiple competing airlines. The study on Single airlines is primarily conducted from the perspectives of linear programming models and microeconomic models. Strategic interactions between airlines after deregulation have an impact on network choices (Oum et al., 1995). The influence of the network on the demand side and the heterogeneity of the passenger time value are considered when choosing the network structure (Kawasaki, 2008). Monopoly airlines' timing, traffic and aircraft size choices are influenced by network structure (Brueckner, 2004). Carbon charging affects choice of airline network structure (Liao & Wang, 2021). The correlation between the airline network structure and the airport congestion (Fageda, 2012). The motivation of airlines to adopt the multi-hub network from the perspective of horizontal product differentiation (Wang, 2019). Network equilibrium and welfare effects are studied using fares and frequencies as the main factors in network structure selection (Fillol, 2009). The study of Multiple competing airlines can be categorized into two main perspectives, static and dynamic game model. A theoretical model of a two-stage Nash game is built based on the microeconomic theory (Adler, 2001). A two-stage bi-oligopoly model was developed to study airline network competition strategies (Silva et al., 2014). Network structure is chosen by airlines in a duopoly competitive environment (Babići & Kalić, 2018). Although the above studies have considered the competition factor, the dynamic effects of competition were ignored. In a three-city network, a two-stage dynamic game model of network competition is developed to study how two hub firms with established subsidiaries engage in network competition (Lin, 2015). In a duopoly market with successive entries of airlines, a two-layer model has been proposed to describe the behavioral and interactive competition between leaders and followers (Yang et al., 2020). Although the dynamism of competition has been considered in the above studies, the uncertainty factors of the market have been ignored. Demand uncertainty and competition have an impact on the choice of airlines' network structure (Barla &Constantatos, 2005). The hub competition model for non-cooperative games among multiple agents has been established (Hansen, 1990). The effectiveness of different market entry models was examined using data from the U.S. aviation industry (Ciliberto & Zhang, 2017).

Table 1 Summary of Literature Reviews

Topics	Sub-topics	Literature
Non-competitive environment	Demand-side network effects of HS network	Oum et al. (1995)
	Passenger time value heterogeneity	Kawasaki (2008)
	Carbon emissions charges	Liao and Wang (2021)
	Airport congestion	Fageda (2017)
	Horizontal product differentiation	Wang (2019)
	Time-competitive model	Fillol (2009)
Competitive environment	Two-stage oligopoly model	Adler (2001), Silva (2014), Babići and Kalić (2018)
	Two-stage dynamic game model	Lin (2015), Yang et al. (2020)
	Three-stage game model	Barla & Constantatos (2005)
	Hub Competition Model	Hansen (1990)
	market entry model	Ciliberto & Zhang (2017)

Various stability concepts exist for GMCR analysis, each differing in the assumptions made about the behavior of the decision-makers participating in the conflict. As an illustration, under the Nash stability concept, each decision-maker evaluates the feasibility of steering the conflict toward a more favorable scenario, irrespective of potential responses from the opponents (Nash, 1950; 1951). In the concepts of General Metarationality (GMR) (Howard, 1971) and Sequential Stability (SEQ) (Fraser and Hipel 1979; 1984), decision-makers assess whether opponents can respond in a manner that does not improve upon the current state in the conflict. The distinction among these concepts lies in the requirement that, in SEQ, opponents' reactions must be advantageous to them, called focal DM. A solution concept was generalized, called Symmetric Sequential Stability (SSEQ), in the GMCR for conflicts involving n DMs (Rêgo and Vieira, 2017). This concept modifies SEQ stability by allowing a counter-reaction of the focal DM after the opponents' sanction moves. Generalized metarationalities were proposed (Zeng et al., 2006, 2007) and further developed are useful stability concepts that depend on how many steps ahead the focal DM foresees the conflict (Rêgo and Vieira, 2020).

From the above-mentioned literature summarized in Table 1, the equilibria guiding the selection of networks reflect only the perception of the focal airlines on the outcome which would take place in the next step. What would the opponent DMs act subsequently after the initial move from the focal airlines is not considered. These subsequent moves may result in disadvantageous outcomes for the focal airlines, so that variety of equilibria reflecting the interaction among multiple airlines with more than one step of foresight should be introduced. Besides Nash equilibrium, the equilibria reflecting multiple steps of foresight can provide airlines with an enhanced understanding of the market competition in selecting network and an enriched package of guidance for actions.

The concept of SEQ stability can effectively simulate the continuous decision-making process among multiple participants in the aviation network. In the issue of airline network structure selection, airlines are required to consider the behaviors of opponents and respond accordingly in the design of network structure. Sequential stability is specifically designed for the continuous decision-making process involving multiple participants. Participants are required to choose the optimal strategy at each decision-making step, while also considering the possible reactions of other participants. This continuous decision-making process reflects the decisions made by various parties involved in airline network structure selection at different stages, such as flight scheduling and pricing strategies. The resolution of SEQ stability can take into account long-term impacts and seek a solution that maintains stability throughout the continuous decision-making process.

The research in this paper is based on the numerical example in which two competing airlines select two types of networks, either PP or HS (Babić and Kalić, 2018). The contribution of this paper is to investigate the equilibria reflecting the interaction among DMs in two steps of move, by utilizing a type of solution concept called sequential stability (SEQ). The calculation for equilibrium, within the framework of GMCR, can suggest all Nash equilibria from (Babić and Kalić 2018), as well as demonstrating SEQ equilibrium as complimentary but important implications for airlines. In practice, SEQ equilibrium, together with Nash equilibrium, can suggest wider range of options by taking which possible agreements between the two airlines can be reached, or provide solider ground for individual airlines to design market strategy in network competition.

2. A simple Network Selection Model

In this study, the structure of a simple network selection model is introduced according to Babić and Kalić (2018). Assume that two competing airlines, labelled as 1 and 2, operate their fly service among three points, denoted as A, B, and H, seen in Figure. 1. Direct flights are established between A and H, and between H and B. The connection between A and B can be either direct, or indirect via H. If direct, the network is point-to-point, i.e., PP. Otherwise, the network is hub-and-spoke, denoted as HS. Thus, each airline, either 1 or 2, can choose PP or HS for connecting the three points.

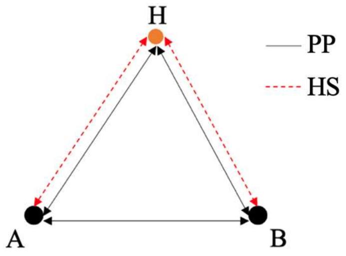


Figure 1: Network structure

3. Framework of Graph Model for Conflict Resolution

A graph model for depicting strategic conflict is said to be written in a 4-tuple set $G = \langle N,\ S,\ A,\ \gtrsim \rangle$, where $N,\ S,\ A$, and $\gtrsim$ represent the sets of DMs, states describing outcomes of conflict, transition of states, and preferene relations. Each DM may control multiple choice of actions, called options. To facilitate finding the matched information from online text, the standard form of option is defined.

In graph model, each DM controls one or more choices of actions, called options. Each option may or may not be selected by a given DM: only binary choices are considered. The selection of an option by the DM can be denoted as Y while the negation N. Note that a DM may control one or more options. A situation at which all DMs select their respective options is defined as a state $s \in S$. A state will transit to another when a single DM changes the selection of one or more of its options. This transition is described as an arc of move which is an element of A. Let $s,\ s^{'} \in S$, the transition from s to $s^{'}$ by DM $i \in N$ can be denoted as $s,\ s^{'} \in\ A_i$, where A is a list of A_i as $(A_i)_{i \in N}$ for each $i \in N$.

Definition 1. (Unilateral Moves, UM) (Fang et al., 1993): Recall that $s,\ s^{'} \in S$, the set of UMs initiated by DM $i \in N$ from s is defined as $R_i(s)$, for $R_i(s) = \{s^{'}: (s,\ s^{'})_i \in\ A_i\}$ and A_i the set of state transition initiated by i.

Preference relations in graph model are described as the comparison between each pair of states. Suppose $s,\ s^{'} \in S$, three basic relations, $\succ, \prec$ and $\sim$, can be written as:

- $s \succ_i s^{'}$ indicating that s is more preferred to $s^{'}$ for $i \in N$;
- $s \prec_i s^{'}$meaning that s is less preferred to $s^{'}$ for i;
- $s \sim_i s^{'}$ denoting that s is indiferente to $s^{'}$ for i.

By considering preference relations, unilateral improvement (UI) is defined as the directed transition of state in which the destination state is more preferred to the starting state for the focal DM.

Definition 2. (Unilateral improvements, UI) (Fang et al., 1993): A UI set for DM $i \in N$ from s is written as $R_i^+(s) = \{s^{'} \in S: (s,\ s^{'})_i \in\ A_i, s^{'} \succ_i s\}$.

Solution concepts, or called stabilities, describes the behavior of DMs reflecting their diversified foresights. Four types of stabilities are commonly utilized in analyzing realworld conflicts, Nash stability (Nash) (Nash, 1950; 1951), Sequential Stability (SEQ) (Fraser and Hipel, 1979; 1984), General Metarationality (GMR) (Howard, 1971), and Symmetric Metarationality (SMR) (Howard, 1971). A state may be stable under one or multiple solution concepts for a single DM, meaning that this DM may refrain from implementing its options so that this state does not transit into another. Equilibrium is then calculated to suggest meaningful guidance of action for DMs. A state is an equilibrium if it is stable for all DMs in graph model, no matter under which solution concept.

4. Network Competition in Graph Model

The DMs and options in network competition are listed in the Table 2. Airline 1 and Airline 2 are denoted by DM1 and DM2, respectively and Each DM has two options, due to the symmetry in the options of Airline Company 1 and Airline Company 2, the description below focuses solely on the options of Airline 1. 1) signifies Point to point network is selected by Airline 1, and the negation of 1) means choosing hub network. 2) indicates Airline 1 opting for HS, and the negation of 2) means choosing PP.

Table 2. DMs and Their Options

DMs	Options
Airline 1（DM1）	1) Select PP 2) Select HS
Airline 2（DM2）	3) Select PP 4) Select HS

According to the Table 2 there are 2^4 outcomes, referred to as states. Infeasible states are identified and deleted according to the actual situation. For instance, PP and HS cannot be selected at the same time for DM1. Then 4 feasible states remain in the Table 3, $S=\{s_1, s_2, s_3, s_4\}$. Each option can be selected by DMs or not selected: consider only binary choices. The choice of options can be represented as Y, while no choice can be represented as N.

Table 3. Feasible States

DMs	Options	s_1	s_2	s_3	s_4
DM1	1）	Y	Y	N	N
	2）	N	N	Y	Y
DM2	3）	Y	N	Y	N
	4）	N	Y	N	Y

The graph model of airline 1 and airline 2 is represented by Figures 2 and 3, respectively. In the graph model, the arc indicates that the decision maker can move from the state of the arrow tail to the state indicated by the arrow.

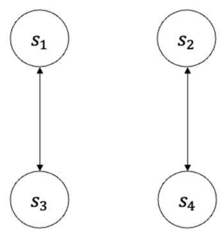


Figure 2: The Graph Model of Airline 1

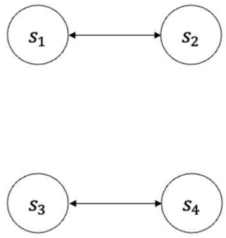


Figure 3: The Graph Model of Airline 2

The profit of DMs in different states is represented by π_{ij}^{k} which is shown in the Table 4, i = PP, HS, j = 1,2; k=s,d. The meanings of these symbols: i =PP denote the choice of point-to-point network structure, i =HS represents the choice of hub network; j =1 represents Airline 1, j =2 represents Airline 2; k=s represents choosing the same network structure, k=d represent choosing different network structures. For instance,
π_{PP1}^{s} represents the profit of Airline 1 when both airlines choose the same network structure PP.

Table 4. Profit of DMs

	s_1	s_2	s_3	s_4
DM1	π_{PP1}^{s}	π_{PP1}^{d}	π_{HS1}^{d}	π_{HS1}^{s}
DM2	π_{PP2}^{s}	π_{HS1}^{d}	π_{PP2}^{d}	π_{HS2}^{s}

DMs in the conflict event rank the advantages and disadvantages of the feasible states in the conflict model according to their own goal requirements and value judgment, forming their own preference order. In this paper, since the conflict model is relatively simple and DMs pursue profit maximization, the direct ordering method is used to find the profit function based on the equilibrium solutions of airlines under different combinations of network structures, which are ranked from left to right, with the most preferred state ranked on the leftmost side, and the worst state ranked on the rightmost side.

All possible ranking cases are divided into 4 scenarios.

Scenario 1: As $\pi_{PP1}^{d} > \pi_{HS1}^{s}$ and $\pi_{PP1}^{s} > \pi_{HS1}^{d}$, The preference order of DM1 and DM2 is:
DM1:$s_1 \succ s_3$， $s_2 \succ s_4$； DM2： $s_1 \succ s_2$， $s_3 \succ s_4$

Scenario 2: As $\pi_{PP1}^{d} < \pi_{HS1}^{s}$ and $\pi_{PP1}^{s} < \pi_{HS1}^{d}$, The preference order of DM1 and DM2 is:
DM1:$s_3 \succ s_1$， $s_4 \succ s_2$； DM2： $s_2 \succ s_1$， $s_4 \succ s_3$

Scenario 3: As $\pi_{PP1}^{d} < \pi_{HS1}^{s}$and $\pi_{PP1}^{s} > \pi_{HS1}^{d}$, The preference order of DM1 and DM2 is:
DM1:$s_1 \succ s_3$， $s_4 \succ s_2$； DM2： $s_1 \succ s_2$， $s_4 \succ s_3$

Scenario 4: As $\pi_{PP1}^{d} > \pi_{HS1}^{s}$and $\pi_{PP1}^{s} < \pi_{HS1}^{d}$, The preference order of DM1 and DM2 is:
DM1:$s_3 \succ s_1$， $s_2 \succ s_4$； DM2： $s_2 \succ s_1$， $s_3 \succ s_4$

5. Equilibrium Solution

Stability analysis is the key to a rational resolution of the conflict problem and aims at finding the equilibrium point of the overall conflict situation. By equilibrium point, we mean a state that is acceptable to all decision makers. For example, for decision maker i, after changing strategies, he will get a new state, and the preference of the new state is inferior to the original state, at this time, decision maker i is not willing to leave the original state. The original state is called an equilibrium state. If all decision makers are stabilized at a certain state point, the state is considered stable for all decision makers, and the state is called a global equilibrium point. If the global equilibrium point is reached, no decision maker is willing to leave the solution he has chosen, so the global equilibrium point may be a satisfactory solution to the conflict.

The solved Nash and SEQ equilibria for each of the four scenarios is shown in the Table 5. The following theorems are obtained and proven.

Theorem 1 If and only if $\pi_{PP1}^{d} > \pi_{HS1}^{s}$ and $\pi_{PP1}^{s} > \pi_{HS1}^{d}$, s_1 are globally unique Nash equilibria.

The Simultaneous selection of PP networks is unique Nash equilibrium solution in the case where the benefits of choosing PP network outweigh the benefits of choosing HS network regardless of the adversary's choice of network.

Theorem 2 Conditional on $\pi_{PP1}^{d} > \pi_{HS1}^{s}$ and $\pi_{PP1}^{s} > \pi_{HS}^{d}$, s_4 is the globally unique strict SEQ equilibrium state if and only if $\pi_{HS1}^{s} \geq \pi_{PP1}^{s}$.

In scenario 1, SEQ equilibrium with multi-step horizons exists in condition $\pi_{HS}^{s} \geq \pi_{PP1}^{s}$ and the profits of the SEQ equilibrium for both airlines is greater than the Nash equilibrium.

Management Insights: In scenario 1, Nash equilibrium is not necessarily globally optimal. If only Nash equilibrium exists, both airlines should choose PP network. If SEQ equilibrium exists, (HS,HS) is the globally unique strict SEQ equilibrium state. As shown in figure 4, in the SEQ equilibrium state (HS,HS), if the airline change its options in pursuit of more interests, the state transfer to s_2. Then the rival also changes its options for more benefits, the state shifts to s_1. However the benefit of state s_1 is less than the initial state (HS,HS), so both airlines should choose HS network for more profit.

Theorem 3 If and only if $\pi_{PP1}^{d} < \pi_{HS1}^{s}$ and $\pi_{PP1}^{s} < \pi_{HS}^{d}$, s_4 are globally unique Nash equilibria.

Similar to Theorem 1.

Theorem 4 Conditional on $\pi_{PP}^{d} < \pi_{HS1}^{s}$ and $\pi_{PP1}^{s} < \pi_{HS1}^{d}$, s_1 is the globally unique strict SEQ equilibrium state if and only if $\pi_{PP}^{s} \geq \pi_{HS1}^{s}$.

Similar to Theorem 2.

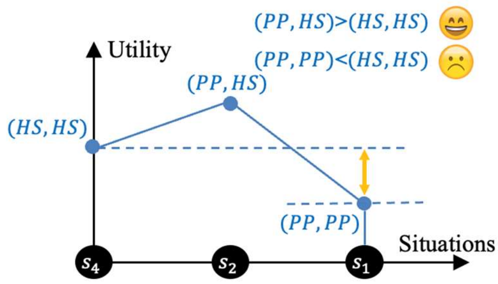


Figure 4: Evolution from (HS,HS)

Table 5. Nash and SEQ Equilibria

	Scenario 1	**Scenario 2**	**Scenario 3**	**Scenario 4**
	$\pi_{PP1}^{d} > \pi_{HS1}^{s}$ and $\pi_{PP1}^{s} > \pi_{HS1}^{d}$	$\pi_{PP1}^{d} < \pi_{HS1}^{s}$ and $\pi_{PP1}^{s} < \pi_{HS1}^{d}$	$\pi_{PP1}^{d} < \pi_{HS1}^{s}$ and $\pi_{PP1}^{s} > \pi_{HS1}^{d}$	$\pi_{PP1}^{d} > \pi_{HS1}^{s}$ and $\pi_{PP1}^{s} < \pi_{HS1}^{d}$
s_1 （**PP,PP**）	Nash	SEQ	Nash	
s_2 （**PP,HS**）				Nash
s_3 （**HS,PP**）				Nash
s_4 （**HS,HS**）	SEQ	Nash	Nash	

6. Sensitivity Analysis

In this section, the range over which the SEQ equilibrium point exists is explored when the relevant parameters are varied through sensitivity analysis and further discuss the impact of the parameters of route length, passenger loyalty, transportation cost, segment-specific cost, frequency delay cost of passengers, and stochastic delay cost on the SEQ equilibrium point.

Due to the space limitation of the article, only one sensitivity analysis for the route length l is presented, seen in Figure. 5.

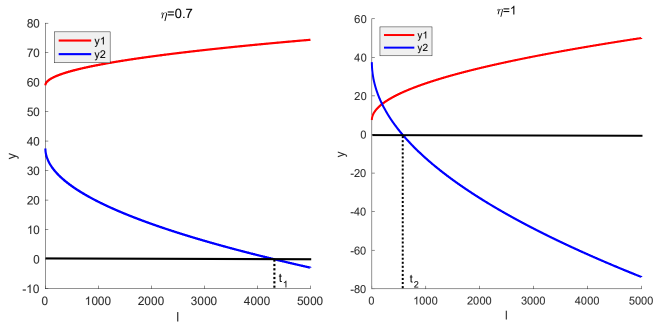


Figure 5: Effect of Routh length l on SEQ equilibrium states

From Figure. 5, it can be seen that the SEQ equilibrium state (HS, HS) exists when η = 0.7 and the route length is less than the critical value t_1, and when η = 1, the critical point t_2 is less than t_1 and the SEQ equilibrium state exists when the segment length is less than the critical value t_2.

Conclusion: in the case of η = 0.7, when the segment length is lower than t_1, the SEQ equilibrium gain is greater than the Nash equilibrium gain, and when η = 1, the exist space for SEQ equilibrium is significantly narrowed; when the segment length is greater than t_2, the SEQ equilibrium solution disappears, and the profit under the combination of different network structures is greater than the profit under the same network structure, and the space of advantages of different network structure combinations is significantly enlarged.

Management insights: when faced with short and medium haul routes, which are usually accompanied by higher unit costs and where passengers are more sensitive to the number of departures, hub networks are more advantageous for offering higher flight frequencies, generating economies of density. Airlines should choose hub networks at the same time; whereas for long-haul routes, as well as higher occupancy rates, airlines should choose a combination of different network structures.

7. Conclusion and Further Study

In this research, network selection between two competing airlines are modelled by Graph Model for Conflict Resolution (GMCR) for the first time. Nash equilibria calculated by GMCR are the same as those by classical game theory. SEQ Equibria under four scenarios with foresight of two steps are calculated by GMCR and computed with Nash Equibria. The conclusion shows that SEQ equilibrium exists only in Scenario 1 and Scenario 2, and SEQ equilibrium suggests more beneficial strategy for both airlines than Nash. This means that the airline should choose the HS network for Scenario 1, the PP network for Scenario 2, and the network in Nash equalization for Scenarios 3 and 4. SEQ Equilibria can suggest enhanced implications to form strategies for airlines.

As further study, equilibria reflecting foresight of more steps need to be investigated, as only two steps foresight is considered in this study. Besides, Cases involving two or more airlines need to be investigated and hierarchies in airlines such as subsidiary airlines should be considered. Due to the simplicity of the model established in this paper, there are many areas worthy of further expansion and improvement. In particular, factors such as travel time, the number of intermediate stops during the journey, additional costs, and delays caused by hub usage, which may affect the selection of network structure, can be incorporated into the model.

References

Babić, D., & Kalić, M. (2018). Modeling the selection of airline network structure in a competitive environment. Journal of Air Transport Management, 66, 42-52. 10.1016/j.jairtraman.2017.10.004

Barla, P., & Constantatos, C. (2005). Strategic interactions and airline network morphology under demand uncertainty. European Economic Review, 49(3), 703-716. 10.1016/S0014-2921(03)00054-0

Chunan Wang Xiaoyu Wang. (2019).Why do airlines prefer multi-hub networks?. Transportation Research Part E56-74. 10.1016/j.tre.2019.02.009

Ciliberto, F. & Zhang, Z. (2017). Multiple equilibria and deterrence in airline markets. Economic Inquiry, 55(1), 319-338. 10.1111/ecin.12381

Fageda, X., & Flores-Fillol, R. (2017). Airport congestion and airline network structure. In The Economics of Airport Operations (pp. 335-359). Emerald Publishing Limited.

Fang, L., Hipel, K. W., & Kilgour, D. M. (1993). Interactive decision making: the graph model for conflict resolution (Vol. 11). John Wiley & Sons. 10.1016/0377-2217(95)90171-x

Flores-Fillol, R. (2009). Airline competition and network structure. Transportation Research Part B, 43(10), 966-983. 10.1016/j.trb.2009.03.002

Fraser, N., Hipel, K. (1984). Conflict Analysis: Models and Resolutions (Vol. 11). North-Holland.

Fraser, N. M., & Hipel, K. W. (1979). Solving complex conflicts. IEEE Transactions on Systems, Man, and Cybernetics, 9(12), 805-816. 10.1109/TSMC.1979.4310131

Hansen, M. (1990). Airline competition in a hub-dominated environment: an application of noncooperative game theory. Transportation Research Part B, 24(1), 27-43. 10.1016/0191-2615(90)90030-3

Howard, N. (1971). Paradoxes of rationality: theory of metagames and political behavior (Vol. 1). MIT press.

Kawasaki, A. (2008). Network effects, heterogeneous time value and network formation in the airline market. Regional Science & Urban Economics,38(4), 388-403. 10.1016/j.regsciurbeco.2008.02.004

Liao, W., & Wang, C. (2021). Airline emissions charges and airline networks. Journal of Air Transport Management, 92, 102028.

Lin, M. H. (2015). Airline network competition with new brand subsidiaries. Journal of Transport Economics and Policy (JTEP), 49(1), 58-78.

Nash Jr, J. F. (1950). Equilibrium points in n-person games. Proceedings of the national academy of sciences, 36(1), 48-49. 10.1073/pnas.36.1.48

Nash, J. (1951). Non-cooperative games. Annals of mathematics, 286-295. 10.2307/1969529

Nicole, & Adler. (2001). Competition in a deregulated air transportation market. European Journal of Operational Research, 129(2), 337-345. 10.1016/S0377-2217(00)00231-9

Oum, T. H., Zhang, A., & Zhang, Y. (1995). Airline network rivalry. Canadian Journal of Economics, 836-857.

Rêgo, L. C., & Vieira, G. I. A. (2017). Symmetric sequential stability in the graph model for conflict resolution with multiple decision makers. Group Decision and Negotiation, 26, 775-792.

Rêgo, L. C., & Vieira, G. I. A. (2020). Alternative generalized metarationalities for multiple decision-maker conflicts. Group Decision and Negotiation, 29(3), 461-490.

Silva, H. E., Verhoef, E. T., & Van Den Berg, V. A. (2014). Airline route structure competition and network policy. Transportation Research Part B: Methodological, 67, 320-343. 10.1016/j.trb.2014.05.012

Yang, T. H., Tang, C. H., & Hsiao, H. C. (2020). Strategic airline network design problem in a duopolistic market. Transportation Planning and Technology, 43(6), 586-601.

Zeng, D. Z., Fang, L., Hipel, K. W., & Kilgour, D. M. (2006). Generalized metarationalities in the graph model for conflict resolution. Discrete Applied Mathematics, 154(16), 2430-2443. 10.1016/ j.dam. 2006.04.021

Zeng, D. Z., Fang, L., Hipel, K. W., & Kilgour, D. M. (2007). Policy equilibrium and generalized metarationalities for multiple decision-maker conflicts. IEEE Transactions on Systems Man & Cybernetics Part A Systems & Humans, 37(4), 456-463. 10.1109/TSMCA.2007.897704

24th International Conference on Group Decision and Negotiation &
10th International Conference on Decision Support System Technology
Human-Centric Decision and Negotiation Support for Societal Transitions

Equilibrium decisions of two delivery service platforms from the perspective of value Co-creation: take “Meituan” and “Ele. me” as an example

Yu Han[1*], Haiyan Xu[1], Jinping Tong [2]
[1] School of Business, Changzhou University, Changzhou, Jiangsu 213159, P. R. China,
[2] College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu 211106, P. R. China,
*yuhan@nuaa.edu.cn, xuhaiyan@nuaa.edu.cn, tjp@cczu.edu.cn

Abstract

Based on the theory and method of the graph model for conflict resolution, this paper analyzes the competition and cooperation between the platforms from the perspective of value co-creation in the first stage, and determines the competition and cooperation strategies of both sides. In the second stage, the balance decision of platform, offline stores and consumers should be discussed, and effective strategic countermeasures are proposed. It also analyzes the negative， riders and positive effects of different factors on platformers. Based on the two-stage value co-creation conflict analysis model, the case study of the two platforms is carried out in an attempt to provide theoretical basis for the future development of the delivery platform, so as to promote the joint sustainable development of multiple parties.

Keywords: graph model for conflict resolution; value co-creation; coordination strategy; delivery service platforms; equilibrium decisions.

1. Introduction

In the operation process of the delivery service platform, consumers participate in the value co-creation process of the delivery platform system through complaints or feedback based on their love of the store's products, the experience and feeling of the service quality of the delivery service staff, and the reasonability or not of the promotion means of the delivery platform. However, there are still some unbalanced decision-making problems that need to be solved urgently. Such as market monopoly, distribution service quality is worrying, food safety is difficult to guarantee, etc. These problems have led to the common occurrence of vicious competition such as "price war" in the process of pursuing development of different platforms, which has damaged the interests of multiple entities. Resulting in consumers' trust in the platform or merchants decreased, consumers and merchants have refund and return disputes, take-out workers unreasonable orders, a certain period of time or a certain area without take-out workers orders and other problems, resulting in value destruction, seriously affecting the value creation of the take-out service industry.

Taking Meituan and Ele. me as an example, the current fierce competition between the two platforms has the following adverse effects: The low-price competition between the two has caused certain pressure on merchants and delivery personnel, and the continuous reduction of commission and delivery fees has led to the difficult survival of merchants and the decline of service quality of delivery personnel. Therefore, in the case of fierce competition in the take-out industry, it is necessary to ensure the stickability of existing users, tap potential customers, and rationally allocate resources from the perspective of co-creating value to help the sustainable development of the take-out service industry. This study aims to take "Meituan" and "Ele. me" as examples, to deeply explore the balanced decision-making mechanism of the two food delivery service platforms from the perspective of value co-creation, explore its influencing factors and realization paths, and provide certain theoretical and practical guidance for solving the existing problems.

In recent years, it can be seen that some well-known delivery platforms, such as "Ele. me Takeaway" and "Meituan Takeaway", not only compete with each other in terms of service price for consumers [5-6], but also compete with each other in terms of service level such as the layout of operation areas and the rebalancing of products and services between different regions [7]. Therefore, in the face of supply is less than demand and supply is more than demand, how the two competing platforms decide their respective service prices and service levels is a research problem that needs attention. Moreover, what is the difference between consumer surplus and social welfare under different supply and demand scenarios? It also needs to be explored further.

Alessandro Crivellari et al. (2022) studied the distribution of short-term food delivery demand in cities and found that a key component of the delivery service platform is logistics, especially the logistics of drivers. Delivery driver delivery logistics. Ideally, the number of drivers operating in urban areas should be just enough to meet the needs of the area. Hsi Tse Wang (2023) studied the value co-creation strategy of stakeholders of the delivery platform based on the tripartite evolutionary game, and reached a balance in three aspects: the adoption of efficient platform, the tendency of restaurants to join the platform, and the use of the platform by consumers. The platform should consider how to increase the positive social evaluation brought by the adoption of efficient platform. To offset the cost of investing in efficient platforms. Xiaoting Dai et al. (2023) explored the influence of collaborative governance participation willingness between restaurants and consumers on delivery platforms from the perspective of cybernetics, and found that perceived risk and online complaints work together to motivate restaurants and consumers to participate in governance activities. Ying Han et al. (2020) chose the three-party evolutionary game to build a mathematical model integrating the government, platform organizations and internal enterprises, analyzed the behaviors of the government, platform organizations and internal enterprises in the platform ecosystem, and found that both formal and informal governance of the government can reduce the opportunist behaviors within enterprises. The effect of formal governance is stronger than informal governance, but the effect of informal governance is more stable. Therefore, most of the previous studies focus on the internal multi-agent game analysis of a single platform, and there is still much room for further research on the equilibrium decision of two platforms in a competitive environment.

The main contribution of this paper is to study the balanced decision-making of food delivery service platform from the perspective of value co-creation, expand the scope of the research on the subject of conflict, and design a two-stage conflict analysis process considering the concurrence relationship, which improves the applicability of the model.

2. The two-stage value co-creation conflict analysis model

The first stage: analysis of competition and cooperation strategy.

Decision Makers	Options
DM1	(1) Antagonistic competition and cooperation strategy (2) Partner-based competition and cooperation strategy
DM2	(1)Antagonistic competition and cooperation strategy (2) Partner-based competition and cooperation strategy

According to their own goal requirements and value judgment, decision makers in conflict events rank the advantages and disadvantages of feasible states in the conflict model and form their own preference order.

The second stage: four decision-makers in the platform game.

Decision Makers	Options	
Delivery service platform (DM1)	a	Adopt a competitive strategy
	b	Adopt a multi-party cooperative strategy
Rider (DM2)	d	Choose multiple platforms to work with
	e	Select only a single platform
Merchants have not entered the platform (DM3)	f	Adhere to the real economy stores
	g	Partnering with delivery service platform
Consumer (DM4)	h	Give feedback to the platform

The preference ranking is based on the factors that the decision-making body participates in value co-creation. Defining the concept of stability of value conflict and share the equilibrium solution is a step we will continue to improve. After this, evaluate the value of the equilibrium solution. When a more desirable balanced outcome is found, the following coordination model is used for preference adjustment.

Minimum cost reverse conflict analysis diagram model of two delivery platforms V：

$$min\sum_{p=s_1}^{s_m}\sum_{q=s_1}^{s_m} VP_i^{+}(p,q)$$

$$\begin{aligned} s.t.\ & VX_i^{\text{Nash}}(s,p)=0, && \forall p\in S,\\ & VP_i^{+}(p,p)=0, && p=s_1,s_2,\cdots,s_m,\\ & VP_i^{+}(p,q)+P_i^{+}(p,q)\le 1, && p,q=s_1,s_2,\cdots,s_m,p\ne q,\\ & VP_i^{+}(p,q)\in\{0,1\}, && p,q=s_1,s_2,\cdots,s_m,p\ne q \end{aligned}$$

In above framework, model V contains two platforms, which is constructed with value co-creation coordination process influenced by risk perception factors.

3. Conclusions

In the competitive environment, the value co-creation of delivery service platform is particularly important. Each profit-maximizing platform should use survey methods, according to the survey report provided by third-party institutions and the development report of product and service sharing industry, or through big data collection and analysis to know the product volume of competitors, as well as the data of competitors' users looking for nearby food delivery products and services, and combined with the user data stored by the platform to estimate the difference Based on the measurement of consumer demand in different periods of time, the supply and demand situation in different periods of time is distinguished. Operators should use big data technology to further estimate the average consumer usage of takeaway products at different times over a period of time. Based on the key information obtained above, the platform dynamically formulates the appropriate service level and service price at different times. For example, during peak hours, the platform should set a higher service level and service price; In the peak period, the platform should reduce the service level and service price.

The theoretical modeling part and case analysis of this paper need to be improved. This work will be continuously updated and improved in order to provide a full draft in the future.

References

Crivellari A, Beinat E, Caetano S, et al. Multi-target CNN-LSTM regressor for predicting urban distribution of short-term food delivery demand[J]. Journal of Business Research, 2022, 144: 844-853.

Chen H S, Liang C H, Liao S Y, et al. Consumer attitudes and purchase intentions toward food delivery platform services[J]. Sustainability, 2020, 12(23): 10177.

Dai D, Ma H, Zhao M, et al. Group Buying Pricing Strategies of O2O Restaurants in Meituan Considering

Service Levels[J]. Systems, 2023, 11(12): 5-62.

Dai X, Qin K, Wu L. Study on effect of collaborative governance participation willingness of online food delivery platform restaurants and consumers from perspective of control theory: Based on moderating effects of perceived risks[J]. Frontiers in Psychology, 2023, 14: 114-9538.

Han Y. A tripartite evolutionary game analysis of enterprises' behaviour in the platform ecosystem[J]. Discrete Dynamics in Nature and Society, 2020: 1-10.

Han Y, Xu H, Fang L, Hipel K W. An Integer Programming Approach to Solving the Inverse Graph Model for Conflict Resolution with Two Decision Makers[J]. Group Decision and Negotiation. 2022, 31:23-48.

Li J, Xu X, Yang Y. Research on Value Co-Creation Strategies for Stakeholders of Takeaway Platforms Based on Tripartite Evolutionary Game[J]. Sustainability, 2023, 15(17): 13010.

Li J, Xu X, Yang Y. Research on the Regulation of Algorithmic Price Discrimination Behaviour of E-Commerce Platform Based on Tripartite Evolutionary Game[J]. Sustainability, 2023, 15(10): 82-94.

Li C, Mirosa M, Bremer P. Review of online food delivery platforms and their impacts on sustainability[J]. Sustainability, 2020, 12(14): 5-528.

Melián-González S. Gig economy delivery services versus professional service companies: Consumers' perceptions of food-delivery services[J]. Technology in Society, 2022, 69: 101-969.

Wang H T. The construction of the strategy selection behavior of online food delivery platform based on the tripartite evolutionary game model[J]. Asia Pacific Management Review, 2023, 28(3): 316-326.

Xu H Y, Hipel K W, Kilgour D M, et al. Conflict resolution using the graph model: strategic interactions in competition and cooperation[M]. Springer, Cham, Switzerland, 2018.

Yu J, Jia W. Research on Price Discrimination Behavior Governance of E-Commerce Platforms—A Bayesian Game Model Based on the Right to Data Portability[J]. Axioms, 2023, 12(10): 9-19.

Zhang P, Ju S, Huang H. Can a Restaurant Benefit from Joining an Online Take-Out Platform?. Mathematics, 2022, 10(9): 13-92.

Group Decision and Negotiation

Poster abstracts

Assessment model and optimization strategies of resilience for Equipment Systems-of-systems

Renjie Xu [1], Jiahao Liu [1], Hang Yang [1], Jiuyao Jiang [1], Jiang Jiang [1] and Jichao Li [1*]

1

[6] National University of Defense Technology, Changsha, China
[*] ljcnudt@hotmail.com

Abstract

The equipment system-of-systems (ESoS) plays a pivotal role in military operations, however, the continuous development and expansion of military confrontations can introduce increased complexity, uncertainty, and vulnerability to the ESoS, particularly when confronted with disruptive events. Although resilience is a pivotal aspect in comprehending and managing the damage and recovery of the ESoS, the current research on the ESoS resilience model, which should encompass comprehensive and effective assessment and optimization, still lacks sufficiency. Hence, this paper proposes a resilience assessment model and explores the optimization of recovery strategies based on this model. Specifically, we use a heterogeneous network model to characterize the ESoS by incorporating diverse types of functional entities and information. Next, developing concepts for operation plans and considering uncertain information, ESoS combat network's performance metric was formulated for ESoS resilience analysis by using rough set theory. It also develops a resilience assessment model based on network topology and network performance. Then, research on the recovery of ESoS resilience process, we present an optimization strategy for resilience based on reinforcement learning and six other recovery strategies from different dimensions to determine link recover sequence and maximum resilience. Finally, the effectiveness of the proposed model and the strategies is demonstrated through resilience analysis of an ESoS combat network model under various conditions, which is built according to the assumed combat scenarios. The results provide valuable insights for designing and building a more robust and resilient ESoS.

Keywords: ESoS combat network; resilience assessment; reinforcement learning; optimization strategy

24th International Conference on Group Decision and Negotiation &
10th International Conference on Decision Support System Technology
Human-Centric Decision and Negotiation Support for Societal Transitions

Hierarchically weighting methods based on ordinal information within MAVT/MAUT

Zhor Chergui [1], Antonio Jiménez-Martín [1*] and María del Carmen Robustillo [2]

[1] Decision Analysis and Statistic Group, Universidad Politécnica de Madrid, Boadilla del Monte, Spain
[2] Departamento de Matemáticas, Universidad de Extremadura, Mérida, Spain
[*] antonio.jimenez@upm.es

Abstract

Different weighting methods based on ordinal information within *multi-attribute utility/value* (MAUT/MAVT) theory are available in the literature, including surrogate weighting methods, the modification of classical decision rules and dominance measuring methods. Several comparison analyses have also been carried out on its performance based on empirical and simulated data using different quality measures, which conclude that the *sum reciprocal* (SR) method is the best-performing one. Besides, weighting methods accounting for additional information regarding the ordinal information have also been proposed by different authors, including a ranking of differences between the weights of consecutive criteria, a semantic scale representing the strength of such differences, and precise/imprecise cardinal information about them.

In all cases, the corresponding weighting method is aimed at simultaneously eliciting the relative importance of all the criteria under consideration. However, many real problems have a multidisciplinary character, sometimes involving a high number of criteria, including criteria of a very different nature or very specialized or technical, such as economic, social, environmental impacts, or other depending on the specific problem being addressed. Moreover, different decision-makers (DMs) could be involved in the multicriteria decision-making problem, with expertise in certain criteria, but without knowledge or experience in others. In that situation, the provision of ordinal information by the DMs for the whole set of attributes may become complex or must be completely discarded.

Note that an objective hierarchy is usually built in a MAUT/MAVT context, with the most general objectives at the upper levels and more specific ones at the lower levels. Attributes are then associated to the lowest level in that hierarchy to indicate the corresponding achievement level by the alternatives under consideration.

We propose hierarchically eliciting weights, i.e., eliciting weights at the different levels and branches of the objective hierarchy, involving in each case the DMs who are experts on such objectives (environmental, economic, social…). Moreover, the corresponding DMs would be able to provide ordinal information on weights but also additional information if desired in the form he/she considers most appropriate (by means of ranking of differences, a semantic scale, or precise/imprecise cardinal information). Thus, different information about weights might be available at the different levels and branches of the objective hierarchy.

Then, the available ordinal information at each level and branch would be transformed into surrogate weight vectors and the attribute weights would be computed by multiplying the weights in the path from the corresponding attribute (lowest-level objective) until the overall objective. Finally, the attribute weights would be incorporated into the multi-attribute utility/value model to derive a ranking of the considered alternatives.

A comparison analyses between the proposed hierarchical weighting method and the *sum reciprocal* method is carried out based on simulated data using the hit ratio and the Kendall's τ as quality measures.

Keywords: MAUT/MAVT; hierarchical weighting method; ordinal information; surrogate weighting methods; comparison analyses

Acknowledgements: Grants PID2021-122209OB-C31 and RED2022-134540-T funded by MICIU/AEI/10.13039/ 501100011033.

A Strategic Analysis of Nagorno-Karabakh Conflict Using Graph Model for Conflict Resolution

Yuming Huang[1], Bingfeng Ge[1], Zeqiang Hou[1], Zihui Liu[1], Jiang Jiang[1], and Kewei Yang [1]

[1] College of Systems Engineering, National University of Defense Technology, Changsha, China
[*] huangyuming14@nudt.edu.cn, bingfengge@nudt.edu.cn,houzeqiang_nudt@163.com, zihuiliu18@163.com, jiangjiangnudt@163.com, and kayyang27@nudt.edu.cn

Abstract

A strategic analysis of the Nagorno-Karabakh conflict dated from September to November 2020, using the Graph Model for Conflict Resolution (GMCR), is presented to formally investigate the strategic decisions and resulting outcomes regarding the dispute between the Azerbaijan and Armenia along with Russian mediation. Formal modeling, stability analysis, and evolution path analysis are carried out to offer a better understanding of the conflict. Analytical results reveal that the Nagorno-Karabakh conflict would not be resolved completely anytime soon, but it is also unlikely to break out into a serious war. Moreover, Russia was a key player in the resolution of the 2020 Nagorno-Karabakh conflict and played an important role in determining the equilibrium states. As demonstrated in this paper, GMCR provides a systematic investigation way to generate farsighted strategic insights with respect to the possible evolution of the conflict and mechanisms for resolving a serious conflict in the future.

Keywords: Conflict resolution; Graph Model for Conflict Resolution; mediation; Nagorno-Karabakh conflict; strategic analysis.

Updating Utility Functions in a Bayesian negotiation model

María Jesús Rufo [1*] and Jacinto Martín [2]

[1] School of Technology, University of Extremadura, Department of Mathematics, Cáceres, Spain
[2] Faculty of Science and ICCAEx, University of Extremadura, Department of Mathematics, Badajoz, Spain
[*] mrufo@unex.es

Abstract

Decision theory plays an important role in reliability issues under uncertainty. Within this context, there are problems that involves two or more actors with different interests. They could be considered as adversarial problems in reliability. A way to address such situations is through a negotiation model. With this idea in mind, this work presents a general negotiation model between two competitors (manufacturer and buyer) who are considered adversaries. They are negotiating the sale of a product based on its lifetime (connected with reliability). It is expected that they reach an agreement. Initially, both actors have their own preferences and beliefs about the product. For the preferences, utility functions based on an observable life length are taken. Under the premise that the manufacturer only has partial knowledge about the buyer's preferences, two uncertainty frameworks are provided. In both schemes, the negotiation process is as follows, at first the manufacturer offers a product to the buyer. Based on his preferences and beliefs, the buyer accepts or rejects it. In the first case, the process finishes. In another case, the manufacturer updates the information on the buyer's utility by modifying the considered prior distribution on the unknown parameter/s in that utility. With this information, the manufacturer either withdraws the proposal or offers the buyer a product batch to make life testing. Thus, he could modify his prior beliefs and accepts or rejects the product. Suitable simulation-based approaches are implemented to decide the optimal sample size in each framework. Finally, an application is presented.

Keywords: Adversarial problems; Decision theory; Negotiation model; Reliability; Utility function

24th International Conference on Group Decision and Negotiation &
10th International Conference on Decision Support System Technology
Human-Centric Decision and Negotiation Support for Societal Transitions

Qualitative and Quantitative Evaluation of Causal Maps created using Group Support Systems

Parmjit Kaur [1*] , Valerio Ficcadenti[2]

[1] Guildhall School of Business and Law, London Metropolitan University, London ,UK.
[2] Business School - London South Bank University , London, UK
[*] corresponding.author@mail.com

Abstract

The research examines the output from a series of group facilitation workshops and the causal maps created using group decision explorer software. The research objective is to compare and contrast the effectiveness of the use of Group Support Systems and the resulting causal maps created to support strategic thinking and to overlay this with a new understanding drawn from a network analysis approach based on words connections, mixing text mining, network analysis and managerial science. This comes together in a mixed-methods approach adopted under the umbrella of soft operational research. The problem structuring methods for the initial workshops leads to a number of text-based causal maps which are studied using quantitative analysis of text data, investigating the connections among words considered crucial according to the field of interest. This latter part utilises advanced techniques to dissect and interpret complex interplay of words used during the workshops. Using these two techniques, we review causal maps through the traditional form using analysis tools from decision explorer and the addition of quantitative analysis of textual data, producing a richer understanding of the maps created. Hence the research aims to assess the depth, scope, and evolution of strategic thinking, proposing an innovative approach from the analysis perspective. This new way of making analysis bridges the gap between qualitative workshop dynamics and quantitative text analysis, offering a comprehensive lens through which to view and understand the intricate processes of strategic decision-making and problem-solving within groups.

Keywords: Causal mapping; Facilitation; Text-mining; Network Analysis; soft operation research

Performance analysis of the JIC2 procedure for the joint improvement of inconsistency and incompatibility

María Teresa Escobar [1*], Juan Aguarón [1] and José María Moreno-Jiménez [1]

[1] Grupo Decisión Multicriterio Zaragoza (GDMZ), Universidad de Zaragoza, Spain
[*] mescobar@unizar.es

Abstract

In a local AHP-GDM context, using the Row Geometric Mean as the prioritisation method, and measuring the inconsistency and incompatibility, respectively, by the Geometric Consistency Index (GCI) and the Geometric Compatibility Index (GCO), the authors have recently proposed a general framework for the joint improvement of inconsistency and incompatibility (JIC2). The associated procedure considers slight modifications in relative terms of the judgements of the collective matrix that further improve the convex combination of the efficiencies of the two indicators. The associated weights reflect the relative importance that decision makers give to improving the validity of the priority vector (inconsistency) and to the representativeness of the collective priority vector (incompatibility). The poster presents a simulation study utilised to analyse the performance of the JIC2 procedure. 10,000 families of matrices were generated for each combination of decision makers (d = 3,…,6) and alternatives (n = 3,…,9). For each family of matrices, an associated consensus matrix was generated by randomly selecting each entry of the matrix from the judgements of the decision makers. The JIC2 procedure was applied using different permissibility values (permissibility is the maximum relative change allowed for the judgements' modifications) and different weights associated to the efficiencies of the two indicators. The performance is analysed by measuring the improvement (efficiency) reached for the GCI and the GCO for each situation. Following a cognitive orientation, this performance analysis aims to provide decision makers with relevant information about the critical points and the decision opportunities of the resolution process.

Keywords: AHP-GDM; Consistency; Compatibility; Joint Improvement; Performance Analysis

Promoting circular economy concerns in medical devices' procurement through structured collaborative processes

Francisco Viterbo[1]*, Ana Carvalho[1] and Ana Vieira[1] and Mónica Oliveira[1,2]

[1] CEG-IST, Centre for Management Studies of Instituto Superior Técnico, Lisbon, Portugal
[2] IBB, Institute for Bioengineering and Biosciences and i4HB – Associated Laboratory Institute for Health and Bioeconomy, Lisbon, Portugal.
* francisco.viterbo@tecnico.ulisboa.pt
anacarvalho@tecnico.ulisboa.pt
ana.lopes.vieira@tecnico.ulisboa.pt
monica.oliveira@tecnico.ulisboa.pt

Abstract

Health expenditure in Portugal has risen significantly and is expected to keep growing, with the health sector contributing 4.6% to global greenhouse emissions. Single-Use Medical Devices (SUMDs) are increasingly used in healthcare and their procurement is mostly price-based with literature recognising the need to implement sustainable procurement practices but studies in the area being scarce. This study aims at identifying, with stakeholder engagement, which aspects related to the evaluation of SUMD should be considered to leverage a more circular healthcare system. A novel multimethodology, integrating Problem Structuring Methods within a Delphi to promote structured collaboration and address aspect interconnectedness, is proposed and tested in a case study with twenty Portuguese stakeholders. In Round 0, participants ideated new aspects to consider in evaluating medical devices to promote circular economy concerns. These aspects were organised into a cognitive map made available to participants in rounds 1 and 2, where participants stated their agreement level with each one and revised their views in light of the group's answers; and provided feedback about the Delphi. Twelve experts completed all rounds and results were validated through individual interviews with a core strategic group. Results suggest a wide range of aspects in evaluating SUMDs, including life cycle costing and assessing SUMD's environmental impact. Portugal-specific recommendations were developed for the short, medium, and long term to leverage SUMD reprocessing and procurement. The adopted multimethodology was effective in depicting interrelationships between value aspects and promoting stakeholder agreement.

Keywords: Health Sector; Green Public Procurement; Medical Devices; Circularity; Delphi

Improving decision making processes in Primary Health Centres through the adaptation of business tools

Victoria Muerza [1*], Maite Escobar [1] and José María Moreno-Jiménez [1]

[1] Grupo Decisión Multicriterio Zaragoza (GDMZ), Zaragoza, Spain
[*] Corresponding author: vmuerza@unizar.es

Abstract

The objective of this research is to improve the quality of public institutions, understood as their effectiveness and innovation, in the health system. Due to the social interest they arouse, we have selected the Primary Health Centres (PHC). The approach followed in our research is closely related to the Knowledge and Artificial Intelligences Society. We adopt a constructive and cognitive perspective characterized by interaction, communication, collaboration, and the relevance of human factor. We apply business tools (technology trees), used in industrial and technological diversification, to identify which services should be provided. To this end, we have designed a methodology in the following phases: (i) Classification of the technologies; (ii) Identification of the relevant (from key to future) technologies for the institution, understanding key technologies as those that guarantee an efficient and effective provision of current services and future technologies as those that provide an effective response to the health challenges and needs of the institution in the medium and long term; (iii) Construction of technological trees based on the relevant technologies; and (iv) Multicriteria selection of the strategy that provides the greatest future social value (the best), understood as the one that provides the greatest added value of knowledge to the considered institution. The results show the technology trees based on the future technologies of PHCs and the selection of the most efficient services that the public administration should promote.

Keywords: Decision-Making Process; Technological Trees; Primary Health Centres; Public Administration; Knowledge Society

24th International Conference on Group Decision and Negotiation &
10th International Conference on Decision Support System Technology
Human-Centric Decision and Negotiation Support for Societal Transitions

Aircraft carbon reduction decision-making from the perspective of game theory

Yuanmao Zeng[1*]

[1] College of Design and Engineering, National University of Singapore, Singapore
[2] Affiliation, City, Country
[*] e1192856@u.nus.edu

Abstract

This research delves into the aircraft operations with the game theory framework, attempting to explore the feasibility of adopting the strategy of cancelling the in-flight meals to reduce the carbon emissions while balancing the interests of both passengers and airlines through a novel approach of giving passengers with points.
Constructing a game theory model which considers the complicated interactions between passengers and airlines, aiming to identify a comprehensive optimized decision-making solution. This holistic optimization strategy will base on the game theory model, in the case of pairwise combinations between passengers who prefer points and those who dislike points, and airlines that provide meals and those that do not, eliciting utility functions under different situations, maximize the utility of both passengers and airlines and minimize the utility difference between them, it is not only promises the environmental sustainability but also presents a win-win cooperation model for both passengers and airlines. Our study provides profound insights and decision support for the future decision-making processes within the aviation industry, emphasizing the potential for a cooperative and sustainable approach to aircraft carbon reduction.

Keywords: utility function; maximize utility; game theory model; decision-making; carbon reduction

Tailored Decision Support Simulation Poster Submission

Hannah K. Johnson

HDR, Inc., Tacoma, United States
hannah.johnson@hdrinc.com

Abstract

With significant amounts of available data, increased social activism and political considerations, and economic uncertainties, decision making in today's world is increasingly complex. These are just a few factors influencing decision quality as companies and government entities strive to make data-driven decisions. However, many organizations are still using traditional decision-making processes without adequately addressing behavioral, social, and/or political uncertainties.

Tailored Decision-Support Simulation (TDSS) is an ideal methodology for evaluating impacts of behavioral activities: equity, environmental justice, and broader community issues. TDSS can be used as a stand-alone methodology to inform decision making, or as the first analytic step towards an informed decision path followed by targeted structured analytic techniques.

TDSS refers to a structured event, or a "simulation," where decision makers are tasked with developing proposed actions in response to a mock scenario. Simulations help people to understand an issue or challenge more deeply, enabling them to make better informed decisions on complex issues. There are many applications of TDSS ranging from testing ideas and concepts, informing investment decisions, and providing training.

Next steps consist of developing pilot tests to determine what is needed to bring simulations to, the architectural and engineering industries. The methodology and plan for these tests will be shared along with intermediate findings.

Keywords: simulation; data-driven; equity; methodology; uncertainty

Where do you come from where do you go – Effects of advice source in negotiation goal setting

Stjepan Jurisic [1*], Markus Voeth [1]

[1]University of Hohenheim, Stuttgart, Germany
[*] stjepan.jurisic@uni-hohenheim.de

Abstract

This research delves into how digital innovations, specifically algorithmic support systems (ASS), are reshaping the domain of negotiation management, focusing on the critical phase of goal setting. Despite the promising capabilities of ASS to enhance negotiation strategies through data analysis and pattern identification there exists a notable hesitance among negotiators towards embracing algorithmic recommendations, identified as "algorithm aversion". The core of this investigation revolves around understanding the differential impacts that the origins of advice (whether algorithmic or human) have on the negotiator's willingness and effectiveness in integrating such guidance into their negotiation strategies. We postulate that advice emanating from human sources is more likely to be utilized than that from algorithms. Through a structured 2x2 experimental setup, the study evaluates the interplay between the type of advice source and the goal's difficulty level. Initial data from 211 participants reveals variances in the acceptance rates of advice, contingent on its source and the associated level of difficulty. The outcomes of this investigation are anticipated to broaden the conceptual and applied understanding of negotiation goal setting, shedding light on the nuances influencing the adoption of algorithmic advice and assisting in the development of more effective negotiation strategies and tools.

The integration of digital technologies, particularly algorithmic support systems (ASS) is transforming negotiation management by offering new avenues for improving negotiation performance. The dynamic nature of negotiations presents an ideal environment for intelligent solutions to assist negotiators in achieving better outcomes. As the presence of ASS in our daily lives grows, their application in negotiation processes reveals numerous investigative opportunities. According to Schulze-Horn et al. (2020), digital technologies can support human negotiators throughout the negotiation process. Furthermore, early research by Kersten & Lai (2007) demonstrated that these technologies can analyze data, identify patterns, and provide strategic recommendations, thereby playing a crucial role in decision-making processes. However, the adoption of algorithmic advice, particularly in the goal-setting phase of negotiations, faces challenges such as skepticism and resistance from negotiators, a phenomenon identified by Patra et al. (2023) as affecting the acceptance and effectiveness of algorithmically generated goals.

Research on goal setting in negotiations utilizes findings from goal-setting theory, suggesting that specific and challenging goals enhance negotiation performance (Locke & Latham, 1990). Different sources of goals (self-set, participatively set, or assigned) impact negotiators' self-efficacy and performance (Erez & Arad, 1986; Latham & Yukl, 1975). With the advent of digital technologies, the role of algorithmic advice in participative goal setting needs examination, as highlighted by Schulze-Horn et al. (2020). Furthermore, "algorithm aversion" a concept introduced by Dietvorst et al. (2014), highlights the inherent skepticism individuals have towards algorithms, influencing their acceptance and utilization in decision-making processes. When measuring individuals' behaviour towards external advice, different formulas can be used. Most of them aim to measure the actual extent of advice weighting in the judgment task. One of those measures to determine the degree of advice utilization is the weight of advice (WOA) formula by Yaniv (2004). The WOA results from the ratio of the difference between the final and initial estimates as well as the advice and initial estimate. To use this formula for the utilization of advice in negotiation goal setting the formula is adjusted in that degree, that the estimates are represented by the goals and are not described in absolute value.

This study aims to understand how the source of advice (human or algorithmic) affects negotiation goal setting, specifically examining negotiators' acceptance and utilization of algorithmically generated goals. We hypothesize that advice from a human expert leads to higher Weight of Advice (WOA) than algorithmic advice, and that self-efficacy and perceived credibility mediate the relationship between advice source and goal advice utilization. In the 2x2 between-subjects experimental design the participants first set a goal for the upcoming negotiation. Afterwards they receive a goal advice from either a human expert or an algorithm. Finally, they can utilize the advice by adjusting their initial goal. Preliminary results from 211 participants suggest differences in advice utilization based on its source and the proposed goal difficulties. Specifically, expert advice leads to higher WOA, particularly for difficult goals. Final results are expected by summer 2024.

Keywords: negotiations; goal setting; algorithmic support systems; advice taking; algorithm aversion

Structuring the problem of Smart and Sustainable Cities and Public Security using Value-Focused Thinking: a case study with the Brazilian Forensic Police

Mateus Müller Franco [1*], Maísa Mendonça Silva [2]

[1] Federal University of Pernambuco, Recife, Pernambuco, Brazil
[2] Federal University of Pernambuco, Recife, Pernambuco, Brazil
[*] mateusmullerf@gmail.com

Abstract

Smart and Sustainable Cities have emerged to support the processes of transformation in cities, seeking solutions to impasses in a wide variety of areas, improving urban spaces, the provision of services and the quality of life of citizens. In this context, security is a dynamic concept that aims to prevent damage to cities and their inhabitants. However, it ends up being neglected in Smart and Sustainable Cities and is considered mainly as a consequence. In this scenario, Public Security is a growing problem in cities, requiring efficient actions and the implementation of improvements to make environments safer for citizens. In this way, the complex problem was structured using Value-Focused Thinking, from Soft Operations Research, with the support of the Brazilian Forensic Police. To this end, four decision-makers took part in this process as a group, using a value-centered approach based on their theoretical and practical experiences in the field. Therefore, in meetings held with the actors involved in this process, the devices of wish list, identification of alternatives, and problems and weaknesses were used. As a result, it was possible to structure the complex research problem, define the strategic objective, as well as the fundamental objectives and means, identify the criteria and their appropriate classifications, build the networks of objectives, and establish the space of actions, with the decision alternatives. It was also possible to provide decision-makers with a holistic understanding of the problem, to connect the two areas of research, to show the importance of the Forensic Police in the area of Public Security in Brazil, and to present the importance and employability of Problem Structuring Methods in real, new, and difficult-to-understand problems. Finally, as future work, it is recommended to use the decision problem constructed in this research with multicriteria decision methods and group decision approaches, seeking to choose, order or rank the alternatives. In this way, it will be possible to significantly improve the processes of the Forensic Police, make collaborative decisions, implement improvements in the area, support the process of transforming cities and the police institutions themselves, and make cities smarter, more sustainable, and safer.

Keywords: Problem Structuring Methods; Value-Focused Thinking; Smart and Sustainable Cities; Public Security; Forensic Police.

The prioritization of alternatives for Smart, Sustainable and Safe Cities with the application of a voting framework and the support of the Brazilian Forensic Police

Mateus Müller Franco [1*], Maísa Mendonça Silva [2]

[1] Federal University of Pernambuco, Recife, Pernambuco, Brazil
[2] Federal University of Pernambuco, Recife, Pernambuco, Brazil
[*] mateusmullerf@gmail.com

Abstract

As cities grow, Public Security challenges increase, and the need arises to create and implement alternatives that make environments more prosperous, healthy, and safe for citizens. At this juncture, Safe Cities emerge by reconciling crime and security rates with the development of Smart and Sustainable Cities, applying a wide range of tools to help engage, impact, and improve different areas and services in cities. However, Public Security is neglected and poorly considered in studies in the area, in the construction of indicators and in theoretical and practical terms. To this end, the importance of the different existing police institutions and the possibility of transforming the current reality with their support and with individual and group multicriteria decision-making processes are highlighted. In this way, the voting framework proposed by De Almeida, Morais and Nurmi (2019) was used, with the support of the five decision-makers belonging to the Brazilian Forensic Police. Methodologically, this framework is proposed in five phases, guiding the presentation of the results of this research. The first phase is based on contextualizing the decision problem, defining the decision-makers and the objectives to be achieved, and constructing the decision matrix by establishing the criteria and alternatives. Part of the resolution of this stage was supported by the use of a Problem Structuring Method, called Value-Focused Thinking. Next, in the second phase, the decision-makers' preferences were elicited using the FITradeoff-ranking method. Thus, through their decision support system, each of the decision-makers individually provided their preferences and Hasse diagrams were generated. Then, in the third phase, sensitivity analysis was carried out using the aforementioned multicriteria method. In the fourth phase, the voting procedures and their properties were evaluated, with the aim of choosing the most appropriate one for the problem worked on in this study. In this stage, the PROMETHEE-ROC multicriteria method was used. Finally, in the fifth phase, with the five rankings for each of the decision-makers, the voting procedure was applied, and an overall result was obtained. Therefore, through the steps followed, it was possible to find a final ranking of the group, presenting an overall result of prioritizing the implementation of alternatives in the area. Furthermore, it will be possible to implement internal and external improvements, including in the Brazilian Forensic Police, improve their processes and contribute to the transformation of police institutions and cities as a whole, making them smarter, more sustainable, and safer cities. As future work, we recommend replicating the steps with other important institutions in the country, such as the Federal Highway Police, the Municipal Guard, and the Fire Brigade, understanding their perception and presenting their preferences, as well as seeking cooperation and joint work between Brazilian police institutions and other actors involved in the process and in this area.

Keywords: Smart and Sustainable Cities; Public Security, Forensic Police; Group Decision; Voting Framework.

Group Decision and Negotiation
- Doctoral consortium -

24th International Conference on Group Decision and Negotiation &
10th International Conference on Decision Support System Technology
Human-Centric Decision and Negotiation Support for Societal Transitions

High-Level Development of Agri-Food Logistics from the Perspective of Policy Tool Combinations: Evidence from the Eastern Region of China

Sha Lai [1], Qingye Han [2] and Ginger Y. Ke [3]

[1] Chongqing University, Chongqing, China. laisha@stu.cqu.edu.cn
[2] Chongqing University, Chongqing, China. hanqy11@cqu.edu.cn
[3] Memorial University of Newfoundland, St. John's, Canada. gke@mun.ca

Abstract

Given the relationship with global food security and sustainable progress, agri-food logistics plays a crucial role in the globalized economy, especially in developing countries where the demand for food is increasing; yet its development is uneven. Although policy interventions have become an effective way of dealing with the corresponding complexity and uncertainty, there is not much research on the mechanisms of policy influence. This study proposes a decision-supporting framework to explore policy combination paths to achieve high levels of agri-food logistics by using fuzzy set qualitative comparative analysis (fsQCA). Specifically, the framework begins with the identification of indicators that represent policy tools and levels of agri-food logistics. Then, the primary information about indicators from 11 provinces in Eastern China is collected through content analysis and principal component analysis. After calibrating with three thresholds (95%, 50%, and 5%), the collected data undergoes the analysis of sufficiency to summarize policy combination policy combination paths and suggestions to reach high level developments (i.e., constituting a membership value of more than 50%) of agri-food logistics by applying fsQCA. The results indicate that four such paths can be found and categorized into demand-pull and factor-driven paths. This presented analytical process can also be referenced by policymakers and administrators in other countries to achieve high levels of agri-food logistics through combinations of policy tools.

Keywords: Agri-food logistics; Fuzzy set qualitative comparative analysis (fsQCA); Policy tool.

1. Introduction

Agri-food logistics, as a key segment in the agri-food supply chain, plays an important role in the safety and value of agri-food, and overall socio-economic development. It significantly reduces the waste and loss of agri-food and ensures the safe transportation of perishable food from farm to fork through a variety of operations dedicated to transportation, storage, and distribution (Allaoui et al., 2018). However, agri-food logistics can easily lead to significant energy consumption and affect sustainable development, given that it involves a multitude of stakeholders and segments and is affected by low profitability and intense competition in the industry (Bhargava & Bansal, 2021). The gradual increase in demand for agri-food in emerging countries because of rapid urbanization, and changes in consumer preferences and eating habits also place higher requirements on the development of agri-food logistics (Naik & Suresh, 2018). In particular, the loss rate of agri-food logistics in China is significantly higher than the general level of 5% in developed countries, resulting in an imbalance between supply and demand.

The industry characteristics of agri-food logistics, that is, strong infrastructure dependence, information asymmetry, and complex externality issues, determine that government intervention is crucial to its development. In reality, there are inherent flaws in market operation mechanisms, such as unfair pricing because of unequal access to information, overlooked environmental costs due to externalities, lack of competition owing to market concentration, high volatility caused by natural factors, and inadequate provision of public goods like safety standards and technology dissemination. Policy intervention, as the main intervention method of the government, is an effective measure to optimize the development of agricultural product logistics. The government can allocate resources from a macro perspective, strengthen the synergy of various links and information sharing, and improve the level of agri-food logistics. Existing research mainly explains "why" policy interventions are needed and "what" the current policy is, yet lacks empirical research on the effects of policy combinations. It is not clear "how" to intervene more effectively. Thus, measuring the intricate causal relationship between various policy tools and the performance of agri-food logistics is crucial for enhancing the precision of empirical results. This importance stems from the dual challenges posed by the growing complexity and uncertainty in agri-food logistics development and the inadequacy of relying solely on a single policy tool.

To support the measurement of policy combination effects, this study first proposes a theoretical framework for mining the policy combination paths to high levels of agri-food logistics by using fsQCA to analyze the complex causal relationships between various policy tools and the level of agri-food logistics. Such paths illustrate the concept of configuration, where outcomes are determined by combinations of multiple factors rather than a single factor; various combinations can produce identical outcomes. As a type of QCA, fsQCA employs fuzzy sets to improve the accuracy of variable measurement (Fiss, 2011). Based on the principle of multiple equivalence, it can identify diverse configurations of policy tools in different spatial contexts and clarify the "core-subsidiary" and "complementary-substitute" relationships among them. It is different from the classical regression analysis that considers the causal effects to be consistent (Yuan et al., 2021). Then, empirical analyses of East China are conducted to derive multiple policy combination paths to achieve high levels of agri-food logistics. The results of this study can fill the gap in quantitative analysis of policies in the field of agricultural product logistics, and help the government to make better decisions in the development of agri-food logistics by assisting policy making.

2. Conceptual framework

The conceptual framework integrates an indicator measurement system and an analytical process for examining causality. Multiple methods are employed in building this framework. Firstly, the literature review method and the Delphi method are employed to identify a set of indicators that represent the types of policy tools and the level of agri-food logistics. The measurement of the indicator system is then obtained by combining the content analysis and principal component analysis methods. By applying the indicator system, the use of policy tools and the level of agricultural logistics in the case can be assessed. In light of the assessment results, the analytical process from FsQCA is used to study the inter-relationships among condition indicators, and then how the occurrence of these indicators sufficiently leads to the occurrence of output indicators. This allows for the generation of equivalent and diversified policy combination paths to achieve high levels of agri-food logistics.

2.1. Identifying and measuring the condition indicators

Condition indicators $(X_1, X_2, X_3, X_4, X_5, X_6)$, i.e., policy tools corresponding to the level of agri-food logistics, are identified through literature review and expert opinions. Specifically, 13 policy tools affecting the level of agri-food logistics are first extracted through a literature review. Then, the 13 policy tools are categorized into three categories: supply-based, environment-based and demand-based policies according to the theoretical framework of policy tools established by Rothwell and Zegveld (1982). Finally, each category of policy tools is generalized and condensed into two condition indicators following the Delphi method as shown in Figure 1, resulting in a total of six condition indicators as shown in Table 1.

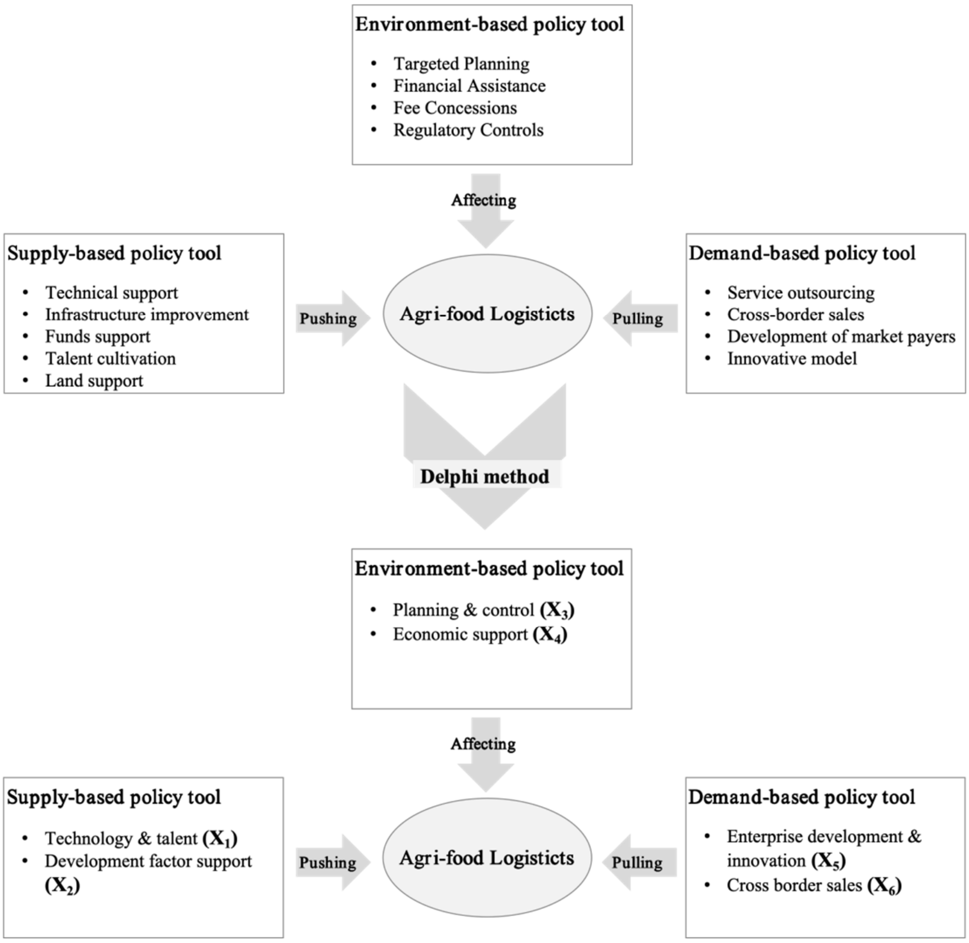


Figure 1. The extraction process for condition indicators

Table 1. Condition indicators

Condition indicators	Symbols	Sub-indicators	Symbols
Technology & talent	X_1	Technical support	X_{1_1}
		Talent cultivation	X_{1_2}
Development factor support	X_2	Infrastructure improvement	X_{2_1}
		Funds support	X_{2_2}
		Land support	X_{2_3}
Planning & control	X_3	Targeted planning	X_{3_1}
		Regulatory controls	X_{3_2}
Economic support	X_4	Financial assistance	X_{4_1}
		Fee concessions	X_{4_2}
Enterprise development & innovation	X_5	Service outsourcing	X_{5_1}
		Development of market players	X_{5_2}
		Innovative models	X_{5_3}
Cross border sales	X_6	Cross-national border sales	X_{6_1}

Based on determining the condition indicators, this study obtains the values of the condition indicators by aggregating the number of policy tools corresponding to the sub-indicators. The measure of the condition indicators can be determined as:

$$\text{Value } (X_{\text{i}}) = \text{Number } \left(X_{\text{ i\_1}}\right) + \text{Number } \left(X_{\text{ i\_2}}\right) + \cdots + \text{Number } \left(X_{\text{ i\_m}}\right), \tag{1}$$

where m denotes the number of sub-indicators for X_i, and X_{i-m} is the sub-indicator for X_i.

2.2. Defining the outcome indicator

The outcome indicator (Y) refers to the level of agri-food logistics. The level of agri-food logistics can be reflected in multiple aspects, both economic and societal, making it difficult to assess with a single indicator. However, it is feasible to measure the level of agri-food logistics comprehensively by constructing an indicator system in the fsQCA research. This study measures this indicator from four dimensions based on the literature review, i.e., macroeconomic environment, supply and demand of agri-food, inputs of the agri-food logistics industry, and information services. Table 2 lists the sub-indicators identified for each dimension. In addition, the contribution of the sub-indicators is determined by applying principal component analysis, and the specific measurement can be calculated by Eq. (2).

$$\text{Value (Y)} = 0.240Y_1 + 0.099Y_2 + 0.287Y_3 + 0.312Y_4 + 0.246Y_5 + 0.172Y_6 + 0.284Y_7 + 0.325Y_8 + 0.262Y_9 + 0.035Y_{10} \tag{2}$$

Table 2. Outcome indicator

Outcome indicator	Symbols	Aspect	sub-indicators
The level of agri-food logistics	Y	Macroeconomic environment	Y_1 : GDP (billion yuan)
			Y_2 : Cargo turnover (billion tons kilometers)
			Y_3 : Freight volume (tons)
		Supply and demand of agri-food	Y_4: Fixed asset investment in agri-food logistics industry (Billions of Yuan)
			Y_5: Transportation infrastructure construction line mileage (kilometers)
			Y_6: Number of employees in agri-food logistics industry (10,000)
		Inputs of agri-food logistics industry	Y_7 : Output of major agri-food (tons)
			Y_8 : Annual consumption of major food products (tons)
		Information services	Y_9: Fixed asset investment in information services (billion yuan)
			Y_{10}: Number of employees in information services (ten thousand)

2.3. Calibration of indicators

After collecting data from the case set based on the indicators and measurements, we transform the data into fuzzy sets, which is a key step known as data calibration in fsQCA. Based on the suggested values given by the Tosmana software (Cronqvist, 2019) and our judgment of the actual situation, the three thresholds are set following Andrews et al. (2016), namely 95% (full in membership: N_1), 0.5 (crossover point: N_2), and 0.05 (full-out membership: N_3) of the cases. Note that the membership function value indicates the degree of cases belonging to the target set. We use $(\cdot)^*$ to represent the calibrated indicators.

2.4. Generating policy combination paths to agri-food logistics by analysing configuration

To empirically complete the generation of paths to high levels of agri-food logistics, fsQCA proceeds in three steps. The first step is using these calibrated set measures to construct a truth table, each row of which corresponds to a specific configuration of condition conditions. The cases are sorted into the truth table based on the set membership of condition conditions. In the second step, the rows of the truth table are reduced constraints: (1) the consistency threshold, i.e., the minimum level of consistency for a combination; (2) the frequency threshold, i.e., the minimum number of cases required for a solution. "Consistency" in this context, calculated by Eq. (3), describes the degree to which the cases correspond to the set-theoretic relationships expressed in the combination.

$$\text{Consistency } (C_h) = \frac{\sum_{k=1}^{qh} \min\left\{\left(Y^k\right)^*, \min_{i=1,\dots,5}\left(X_i^k\right)^*\right\}}{\sum_{k=1}^{qh} \min_{i=1,\dots,5}\left(X_i^k\right)^*}, \tag{3}$$

where qh denotes the number of cases in configuration h (different configurations may have different cases), C_h is the hth conditional configuration, $\min_{i=1,\dots,5}\left(X_i^k\right)^*$ computes the set membership of the conditional configuration C_h in case k.

To improve the effectiveness of the solution, the value of 0.8 is selected as the minimum acceptable consistency, which is higher than the minimum recommended threshold of 0.75 (Ragin & Fiss, 2008). On the other hand, the frequency threshold ensures that a minimum number of empirical observations is obtained given the sample size (Schneider & Wagemann, 2012). The frequency threshold is set to 1 for a rather small case set. In the third step, the truth table is logically analyzed using the Boolean algebra-based algorithm to obtain the solutions. Based on this, three types of solutions can be obtained, namely the complex, intermediate, and parsimonious solutions. Each type of solution has its characteristics (as shown in Table 3). Based on these characteristics, the intermediate solution is chosen as the primary source for summarizing the policy combination paths to agri-food logistics development, and parsimonious solutions as assistance in finding the core explanatory conditions in the policy combination paths.

Table 3. Three types of solutions

Solution Type	Characteristics
Complex solution	Not include analysis of configurations that are not covered in the case set.
Intermediate solution	Include analysis of easily identifiable parts of configurations that are not covered in the case set (e.g. incorporating necessary conditions into the analysis).
Parsimonious solution	Include analysis of all configurations that are not covered in the case set.

3. Case study

The 11 provinces in the Eastern Region of China (excluding Taiwan) are finally selected as case studies for this study. The Eastern Region is the most economically developed in China, and also the frontier region with a better IoT environment, transportation infrastructure, and industrialization. Its development experience is relatively mature and more replicable. Then, the values of the six condition indicators are obtained by coding, analyzing and categorizing the 600 policy tools identified in 215 policy samples from the 11 provinces using content analysis and combining with Eq. (1), as shown in Table 4. Similarly, the outcome indicators for each province are obtained using Eq. (2) and the distribution is shown in Figure 2.

Table 4. Values of condition indicators

Study area	Province	Technology & talent	Development factor support	Planning & control	Economic support	Enterprise development & innovation	Cross border sales
Eastern Region of China	Beijing	6	10	3	1	4	0
	Tianjing	8	12	6	2	8	0
	Hebei	7	12	3	3	13	2
	Liaoning	8	22	4	4	15	1
	Shanghai	4	7	4	5	4	0
	Jiangsu	30	48	8	8	20	2
	Zhejiang	21	42	7	2	17	2
	Fujian	9	40	9	1	12	1
	Shandong	12	31	2	6	11	3
	Guangdong	4	21	2	1	9	6
	Hainan	7	11	4	3	6	3

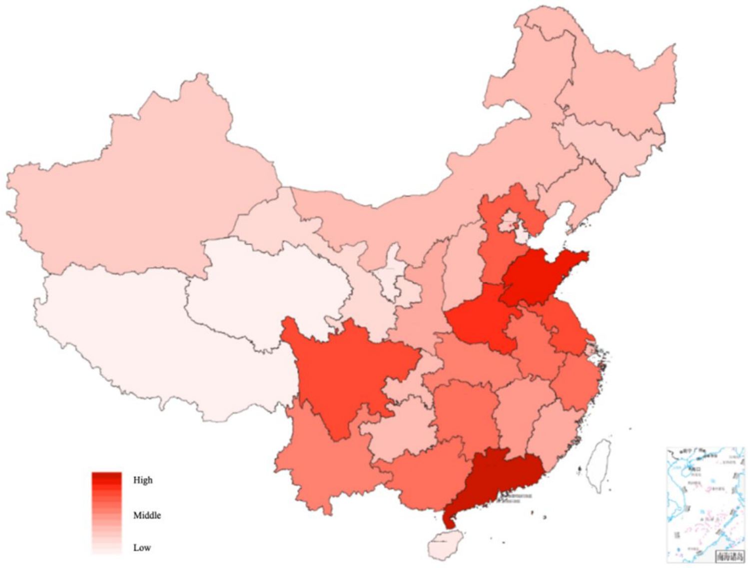


Figure 2. Heat map of outcome indicators

The results facilitate four sufficient policy combination paths after data calibration and configuration analysis (as shown in Table 5). P_1 and P_2 are categorized as demand-pull paths, with two demand-based policy tools at their core conditions, i.e., enterprise development & innovation (X_5) and cross-national border sales (X_6). In this demand-pull path, the government typically employs a range of demand-oriented policy tools to offer varied developmental support to different types of enterprises, thereby encouraging them to engage in the cross-national border import and export of agri-food products. The mutually reinforcing relationship between the two core policy instruments can lead to desired policy implementation. P_3 and P_4 are categorized as factor-driven paths, whose core condition is the development of factor support (X_2). In this path category, the supply of development factors has the most direct role in promoting the development of agri-logistics, but the use of other types of policies varies greatly. For most regions, adequate resources paired with appropriate regulatory instruments can fuel technological research and enterprise development, thus guiding agri-food logistics to high levels of development.

After conducting a thorough analysis, this study provides targeted policymaking suggestions to local governments based on configuration research. First, local governments can issue policies that combine regional characteristics to adapt to local conditions. Secondly, local governments should take advantage of regional complementarity to achieve high-quality industry development. Finally, local governments should recognize the crucial role of technological innovation and talent development in the field of agricultural logistics, and increase their support for technological innovation and talent training. This can be achieved by establishing a comprehensive talent training system and providing incentives for technology research and development. By doing so, they can attract more talented individuals to contribute to the advancement of agricultural logistics.

Table 5. Analytical results

The result of fsQCA	Symbol of the path	Typical case	Path classification	Core condition
Configuration 1	P_1	Hebei	Demand-pull path	Enterprise development & innovation(X_5) and Cross border sales(X_6)
Configuration 2	P_2	Jiangsu Shandong		
Configuration 3	P_3	Zhejiang Fujian	Factor-driven path	Development factor support(X_2)
Configuration 4	P_4	Guangdong		

4. Conclusion

This study aims to assist in determining the optimal combination of policy tools for the development of agricultural logistics. Based on the identified condition indicators of agri-food logistics, i.e., environment-based, supply-based, and demand-based policy tools, including the technology & talent, development factor support, planning & control, economic support, enterprise development & innovation, and cross-national border sales, we used the fsQCA methodology to analyze the conditional configurations of 11 provinces in Eastern China. Our results indicate that different regions with varying resources and development styles have distinct policy combination paths to achieve high levels of agri-food logistics.

In total, four policy combination paths are identified and categorized as demand-pull and factor-driven paths, in which enterprise development & innovation, cross-national border sales and development factor support are the core conditions playing the main role. Therefore, targeted policy suggestions are made to promote high levels of agricultural logistics. Firstly, policymakers should identify the substitution effect of policy tools and focus on regional characteristics. Secondly, policymakers should emphasize the synergistic development between regions by complementing each other's strengths. Finally, policymakers should give attention to the core driving roles of technological innovation and talent cultivation and build a balanced policy system.

This study concludes that combining policy tools is an effective way to promote high levels of agri-food logistics. To provide better guidance for decision-making, we need to understand the processes and influencing mechanisms. Herein, an analytical framework is developed to explore the complex causal relationship between policy instruments and the level of development in agri-food logistics. It is recommended that policymakers and executive leaders pay attention to this study to provide a decision-making basis for future policy practices in agri-food logistics.

Acknowledgement

This research was funded by National Natural Science Foundation of China (Award Number: 72104040), Chongqing Social Science Planning Fund (Award Number: 2020QNGL27), Central University Basic Research Fund of China (Award Number: 2021CDJSKJC21), Postdoctoral Research Foundation of China (Award Number: 2020M683253) and Natural Science Foundation Project of Chongqing, Chongqing Science and Technology Commission (Award Number: cste2020jcyjbshX0038). The third author also appreciates the support from Discovery Grants from the Natural Sciences and Engineering Research Council of Canada (RGPIN-2015-04013, RGPIN-2022-03514).

References

Andrews, R., Beynon, M. J., & McDermott, A. M. (2016). Organizational capability in the public sector: A configurational approach. *Journal of Public Administration Research and Theory*, *26*(2), 239-258.

Allaoui, H., Guo, Y., Choudhary, A., & Bloemhof, J. (2018). Sustainable agri-food supply chain design using two-stage hybrid multi-objective decision-making approach. *Computers & Operations Research*, 89,369–384.

Bhargava, A., Bansal, A., 2021. Fruits and vegetables quality evaluation using computer vision: A review. *J King Saud Univ Comput Inf Sci* 33 (3), 243–257.

Cronqvist, L. (2019). Tosmana [Version 1.61]. *University of Trier. Internet: https://www. tosmana. net*.

Fiss, P. C. (2011). Building better causal theories: A fuzzy set approach to typologies in organization research. *Academy of management journal*, *54*(2), 393-420.

Naik, G., & Suresh, D. (2018). Challenges of creating sustainable Agri-retail supply chains. *IIMB Management Review*, 30(3), 270–282.

Rothwell, R., & Zegveld, W. (1982). Innovation and the small and medium sized firm. *University of Illinois at Urbana-Champaign's Academy* for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship.

Ragin, C. C., & Fiss, P. C. (2008). Net effects analysis versus configurational analysis: An empirical demonstration. *Redesigning social inquiry: Fuzzy sets and beyond*, *240*, 190-212.

Schneider, C. Q., & Wagemann, C. (2012). *Set-theoretic methods for the social sciences: A guide to qualitative comparative analysis*. Cambridge University Press.

Yuan, B., Xu, S., Chen, L., & Niu, M. (2022). How do psychological cognition and institutional environmentaffect t he unsafe behavior of construction workers?—Research on fsQCA method. *Frontiers in Psychology*, 13, 875348.

24th International Conference on Group Decision and Negotiation &
10th International Conference on Decision Support System Technology
Human-Centric Decision and Negotiation Support for Societal Transitions

Group decision making in multiobjective optimization

Juuso Pajasmaa[1*]

[1] Faculty of Information Technology, Jyvaskyla, Finland
[*] juuso.p.pajasmaa@jyu.fi

Abstract

The focus of this thesis is group decision making in multiobjective optimization. The methods of solving problems considering multiple objective functions have focused on decision support for a single decision maker. Therefore, including multiple decision makers to solve complex problems with conflicting objective functions requires research attention. My research aims to develop novel interactive methods to solve these kinds of problems involving multiple decision makers and addressing the challenges solving of real-world problems. The research has started with conducting a systematic literature review on group decision making in multiobjective optimization. Moreover, I am developing a novel interactive method that is based on an improvement concept, discussed in more detail later. In this report, I will describe my next steps.

Keywords: multiobjective optimization; group decision making; group decision making multiobjective optimization methods; interactive methods; GDM-MOO

1. Background on multiobjective optimization

Many real-world problems contain multiple conflicting objectives, such as maximizing profit, while minimizing risks in portfolio optimization or maximizing the speed of the car, while minimizing fuel usage and minimizing the price when buying a car. The objectives are often conflicting, e.g., a faster car consumes more fuel. These kinds of problems with multiple conflicting objectives are known as multiple criteria decision making (MCDM) problems (Ehrgott, 2005; Triantaphyllou, 2000). Depending on the properties of the feasible solutions, this type of problems can be divided into multiobjective optimization (MOO) (Miettinen, 1999) and multicriteria decision analysis (MCDA) problems. In MOO, the set of solutions is not explicitly known in advance. Instead, objective functions to be optimized depend on decision variables. On the other hand, in MCDA, the set of solution alternatives is explicitly predetermined, discrete and finite (Hwang & Yoon, 1981; Miettinen, 1999).

Multiobjective optimization problems (MOPs) contain multiple conflicting objective functions to be optimized simultaneously. One can identify so-called Pareto optimal solutions to such problems. These solutions represent different trade-offs and are mathematically incomparable without additional information. Hence, the expertise of a decision maker (DM) is needed to find and select the most preferred solution (Miettinen, 1999) reflecting the preferences of the DM.

MOPs can be solved by applying MOO methods that incorporate preference information. An example of the type of preference information is an *aspiration level* for each objective function. They represent objective function values that the DM finds desirable. The aspiration levels are components of a *reference point*.

The DM may express preference information before, during, or after optimization. MOO methods that enable the DM to provide preferences during the optimization process, known as interactive methods, provide multiple benefits to the DM. For instance, the DM may learn about the feasibility of their preferences and the trade-offs among the objective functions. Based on this learning, they can then provide well-informed preferences with realistic trade-offs, allowing the optimization method to find solutions that meet the desires of the DM.

The literature on solving MOPs has focused a lot on supporting a DM, or a group of unanimous DMs, acting as a DM. However, many real-world problems have multiple DMs with conflicting preferences on the objective functions and different knowledge, expertise, and attitude to the problem. Hence, in my research, I focus on interactive MOO methods for more than one DM.

2. Motivation and research questions

Group decision making (GDM) involves multiple DMs that are interested in solving a common problem (Hwang & Lin, 1987; Lu & Ruan, 2007; Raiffa et al., 2002). The combination of GDM and MOO can be referred to as GDM-MOO, according to (Xiong et al., 2013). The research on GDM-MOO has gained limited attention in the scientific community (Laengle et al., 2018). The authors of (Laengle et al., 2018) consider papers published between 1992 and 2016 in the *Group Decision and Negotiation* journal. According to them, no information on multiobjective optimization can be found in the top 50 cited papers or the top 40 co-occurrences of author keywords. This is also noted in the emerging literature, see e.g. (Tomczyk & Kadziński, 2022). Despite the need to include the expertise of multiple DMs in various application domains with multiple objective functions, very little attention has been paid to developing methods for GDM-MOPs. My research will introduce new interactive methods to fill this gap.

Next, I describe a few key findings regarding GDM-MOO based on my Master's thesis in (Pajasmaa, 2023), which lays a foundation to my research. There exists no widely accepted terminology or taxonomy for GDM-MOO considering the aspects of GDM relevant to MOO. Consequently, there is no guidance on what type of MOO methods to use with different types of GDM-MOPs, e.g. a group solving a MOP, where they do not share all objective functions or decision variables. Similarly, a group of DMs led by a superior DM may benefit from a different type of GDM-MOO method than a group where all DMs are equal and also treated equally in the solution process.

Additionally, little attention has been paid on handling multiple preferences of several DMs in the solution process and the phase of selecting the final, most preferred solution, is often dismissed. Without selecting one (final) solution, there is nothing to implement in the real world. Regarding the former, the conducted research mainly focuses on different ways of aggregating the DMs' preferences into a group preference. Then, the MOP is solved as if the group was a unanimous DM.

Instead of only focusing on what is familiar from the GDM literature and applying it to MOPs, there must be other ways of finding preferable solutions for multiple DMs. The needs of different types of groups require different MOO methods and these methods should include different ways of incorporating the preferences and different ways of selecting the most preferred solution for the group.

Therefore, the research questions I plan to tackle are:

1. What is the state-of-the-art of GDM and MOP?
2. What are the evident gaps in the literature?
3. How to incorporate the multiple preferences of several DMs into the solution process?
4. How to develop interactive GDM-MOO methods for solving different types of GDM-MOPs? This includes providing answers to how to deal with the challenges of real-world GDM-MOPs.

3. Research approach and preliminary results

My research tackles two separate research fields: GDM and MOO. I plan to publish 5 papers in a 4-year period and complete my Ph.D. thesis as a compilation dissertation. I answer the research questions in three main ways. Firstly, I will assess the existing GDM-MOO literature and detect gaps. Secondly, I will develop novel GDM-MOO methods fulfilling these detected gaps and thirdly, by solving real-world problems I will validate the developed GDM-MOO methods. Additionally, by doing this, I will receive much information on how to further develop these methods to be usable for solving real-world GDM-MOPs. To apply the developed methods and visualizations to address real-world problems, it is crucial to implement them into a practical software framework. The DESDEO framework (Misitano et al., 2021) is a great platform to implement these methods. DESDEO is an open-source software framework for interactive multiobjective optimization and incorporating the GDM-MOO methods in DESDEO will also extend its capabilities to support multiple DMs.

In the first paper, I will conduct a systematic literature review on GDM-MOO aiming to find or create relevant classifications regarding different types of GDM-MOO methods, types of groups of DMs and ways of solving GDM-MOPs. The systematic literature review answers research questions 1 and 2.

The second paper proposes a novel interactive improvement-based GDM-MOO method, called NAUTILI, worked in collaboration with Prof. Kaisa Miettinen, Prof. Francisco Ruiz, Dr. Dmitry Podkopaev, Dr. Babooshka Shavazipour and Dr. Bhupinder Saini. Inspired by the prospect theory (Kahneman & Tversky, 1979), the idea behind improvement-based methods is to start from an inferior solution and improve iteratively until a Pareto optimal solution is reached. In NAUTILI, the search is guided by the DMs, who iteratively express their preferences. In this manner, all the objective functions can improve for all the DMs simultaneously until convergence to a Pareto optimal solution. This can be considered a "win-win" approach and is an extension of the NAUTILUS methods (Miettinen & Ruiz, 2016; Miettinen et al., 2010) developed for a single DM.

The NAUTILI method provides decision support to a collaborative group solving a type of a GDM-MOP, where the group shares the same objective functions and agrees upon whether an objective function has to be minimized or maximized. However, they may have different and (possibly) conflicting preferences on the objective functions. The most preferred solution considers the preferences of the whole group. The method will be implemented and used via a graphical user interface utilizing DESDEO.

A summary of the key steps of the NAUTILI method is given in Figure 1. The NAUTILI method begins with the *1. Initialize* step, where the group determines the number of iterations they want to conduct and the minimum number of votes required to take a step backward. Then, in the *2. Inform* step, the DMs are shown an inferior solution, after which, in *4. Preferences* step, they articulate their preferences as reference points as a suggestion of a direction of improvement. Note, that in the first iteration, we skip the step *3. Vote*.

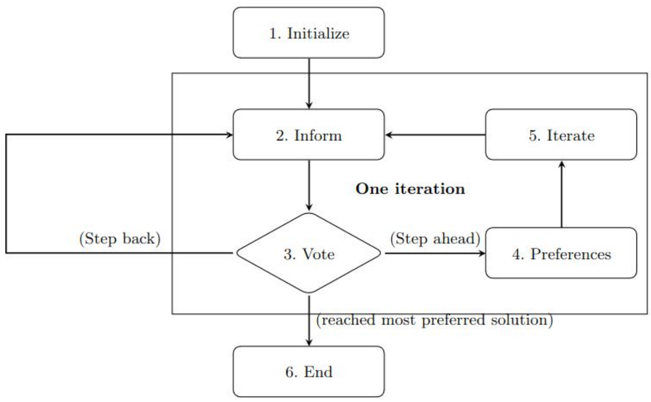


Figure 1: A general NAUTILI scheme describing the solution process. The large square box contains the steps conducted during one iteration of the method.

After getting the multiple reference points converted as multiple direction of improvement, these directions are aggregated as a collective direction of improvement. Then, in the *5. Iterate* step, the method iterates using the collective direction of improvement and simultaneously, we can improve the values of the objective functions for all DMs while taking a step along the collective direction of improvement. The iterative process continues until a Pareto optimal solution is found.

By iteratively approaching the set of Pareto optimal solutions, the DMs see how the values of objective functions that can still be reached without trade-offs shrink. The DMs see a similar visualization as in Figure

2. At any subsequent step (after the first iteration), any DM may start a vote to take a step backward (meaning, going back to an earlier iteration) in the *3. Vote* step before articulating their preferences. In this case, a majority vote will be performed using the graphical user interface to determine whether we take a step back or not. Then, the iterative method proceeds by showing the reachable ranges, current iteration point and their reference point to the DMs which then articulate their preference information.

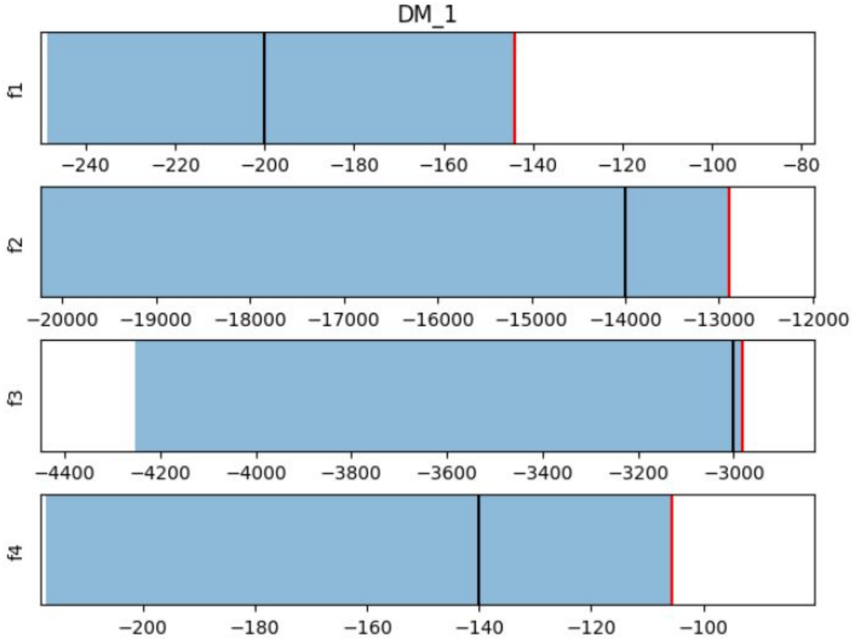


Figure 2: The reachable ranges (in blue), the reference point (the black lines) and the current iteration point (the red lines) are visualized to each DM. Each horizontal bar has an objective function ranging from f_1 to f_4. The numbers below each horizontal bar show the values of the objective functions.

Moreover, the NAUTILI method aims to treat the DMs equally and promotes active participation into the solution process, as the DMs can affect the collective direction of improvement better if they are active. Additionally, this research includes designing relevant visualizations and implementing the method. Moreover, NAUTILI will be tested with forest owners solving a forest management problem. Therefore, the second paper will give an answer to the research questions 3 and 4 regarding one type of GDM-MOP and a GDM-MOO method. Various other types of GDM-MOPs need other GDM-MOO methods.

The conducted work on the NAUTILI method has already raised some additional questions such as why to choose any specific way of aggregating the preferences (e.g. aggregating the directions of improvements versus aggregating the reference points themselves) and how these different ways affect the solution process? Getting answers to these questions is the focus of the third paper. Via collaboration with Prof. Rudolf Vetschera and others, I will consider the advantages and disadvantages of different aggregations of aggregating the reference points and aggregating the directions that the reference points point to. The third paper will answer research question 3 by providing justification and ideas on what kind of aggregation method to use.

The fourth and fifth papers will answer the challenges recognized during the earlier research. The goal is to develop GDM-MOO methods tailored towards different types of GDM-MOPs (e.g. the DMs not sharing the same set of objective functions) and different types of groups compared to those considered in earlier works. The papers will further diversify the applicability of the dissertation. The papers will mainly answer research question 4 and other challenges yet to be known.

Synchronously with the work described above, I will implement the methods and associated visualizations in DESDEO. The rudimentary implementations of the GDM-MOO methods and their respective visualizations from the second, fourth and fifth papers will be refined including the feedback gained from the user experiences. I have already got some experience in using DESDEO during my research.

4. Schedule and current state

Table 1 shows the timeline for my doctoral research. The research contains two phases: developing novel GDM-MOO methods and then implementing them in DESDEO. The development phase will be preceded by assessing the relevant GDM literature. The implementation phase will include testing the implemented software and the methods. All of the phases contain interaction and collaboration with colleagues in the Multiobjective Optimization Group and with the larger international network. The findings and results will be presented at conferences and published in peer-reviewed journals.

Table 1: Gantt chart showing the timeline of the research. The "x" depicts the current progress in the papers.

	Year/Quarter															
	2023		2024				2025				2026				2027	
Tasks	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2
Paper 1: Systematic literature review				x												
Paper 2: NAUTILI				x												
Paper 3: Aggregating reference points				x												
Paper 4: GDM-MOO method I																
Paper 5: GDM-MOO method II																
Experiments with real DMs																
Thesis defence & preparation before defence																

Currently, after the first half year of my doctoral research, I am in the phase of finishing the systematic literature review on GDM-MOO. Additionally, the second paper proposing the NAUTILI method has progressed to the phase of preparing to test the method with forest owners solving a forest problem. Both of these papers are to be finished before the end of spring. Additionally, the work has started on the third paper focusing on different ways of aggregating reference points in improvement-based methods.

References

Ehrgott, M. (2005). *Multicriteria optimization*. Springer.

Hwang, C.-L., & Lin, M. (1987). *Group decision making under multiple criteria: Methods and applications*. Springer.

Hwang, C.-L., & Yoon, K. (1981). *Multiple attribute decision making: Methods and applications: A state-of-the-art survey*. Springer.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47* (2), 263–292.

Laengle, S., Modak, N. M., Merigo, J. M., & Zurita, G. (2018). Twenty-five years of group decision and negotiation: A bibliometric overview. *Group Decision and Negotiation, 27*, 505–542.

Lu, J., & Ruan, D. (2007). *Multi-objective group decision making: Methods, software and applications with fuzzy set techniques*. Imperial College Press.

Miettinen, K. (1999). *Nonlinear multiobjective optimization*. Kluwer Academic Publishers.

Miettinen, K., Eskelinen, P., Ruiz, F., & Luque, M. (2010). NAUTILUS method:An interactive technique in multiobjective optimization based on the nadir point. *European Journal of Operational Research, 206* (2), 426–434.

Miettinen, K., & Ruiz, F. (2016). NAUTILUS framework: Towards trade-off- free interaction in multiobjective optimization. *Journal of Business Economics, 86,* 5–21.

Misitano, G., Saini, B. S., Afsar, B., Shavazipour, B., & Miettinen, K. (2021). DESDEO: The modular and open source framework for interactive multiobjective optimization. *IEEE Access, 9*, 148277–148295.

Pajasmaa, J. (2023). *Group decision making in multiobjective optimization: A systematic literature review* [Master's thesis, University of Jyvaskyla].

Raiffa, H., Richardson, J., & Metcalfe, D. (2002). *Negotiation analysis: The science and art of collaborative decision making.* Harvard University Press.

Tomczyk, M. K., & Kadzi´nski, M. (2022). Interactive co-evolutionary multiple objective optimization algorithms for finding consensus solutions for a group of decision makers. *Information Sciences, 616*, 157–181.

Triantaphyllou, E. (2000). *Multi-criteria decision making methods: A comparative study*. Springer.

Xiong, J., Tan, X., Yang, K.-W., & Chen, Y.-W. (2013). Fuzzy group decision making for multiobjective problems: Tradeoff between consensus and robustness. *Journal of Applied Mathematics, 2013*, article 657978.

Human-Centric Decision and Negotiation Support for Societal Transitions

June 2024

Faculdade de Engenharia da Universidade do Porto
Rua Dr. Roberto Frias s/n, 4200-465 Porto, Portugal