NANYANG
TECHNOLOGICAL
UNIVERSITY
SINGAPORE

Commitment and Coordination in Boundedly Rational Interactions

by

Jakub Černý

School of Computer Science and Engineering

2023

Colophon

This document was typeset in LaTeX using the MIMOSIS template by *Bastian A. Rieck* and sidenote citations inspired by books of *Edward R. Tufte*. The bibliography was processed by Biblatex.

Georg Duffner's EB Garamond acts as both the text and display typeface. Sans-serif and monospaced text is typeset in *Paul D. Hunt*'s Source Code Pro for Adobe Systems.

Commitment and Coordination in Boundedly Rational Interactions

Commitment and Coordination in Boundedly Rational Interactions

by

Jakub Černý

School of Computer Science and Engineering

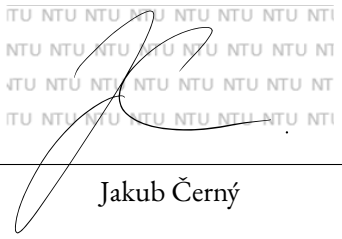
A thesis submitted to the Nanyang Technological University in partial fulfilment of the requirement for the degree of Doctor of Philosophy

2023

Thesis advisers: Prof. Bo An
Dr. Allan Nengsheng Zhang

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarised materials, and has not been submitted for a higher degree to any other University or Institution.

13/1/2023	
Date	Jakub Černý

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

13/1/2023

Date

Bo An

Authorship Attribution Statement

This thesis contains material from 4 papers accepted at the following conferences in which I am listed as an author, and 1 preprint.

- Dinkelbach-type algorithm for computing quantal Stackelberg equilibrium. J. Černý, V. Lisý, B. Bošanský and B. An. In Proceedings of the 29th IJCAI Conference. 2020.
- Computing quantal Stackelberg equilibrium in extensive-form games. J. Černý, V. Lisý, B. Bošanský and B. An. In Proceedings of the 35th AAAI Conference. 2021.

I derived the theory, conducted the experiments and wrote the manuscripts; Profs Lisý and Bošanský motivated the work and helped with writing; and Prof An steered the project towards successful completion.

- Complexity and algorithms for exploiting quantal opponents in large two-player games. D. Milec, J. Černý, V. Lisý and B. An. In Proceedings of the 35th AAAI Conference. 2021.

David designed the algorithms and conducted the experiments; I derived the theory and wrote the manuscript with David; Prof Lisý helped with theory and writing; and Prof An provided insightful comments.

- Quantal correlated equilibrium in normal form games. J. Černý, B. An and A. N. Zhang. In Proceedings of the 23rd ACM EC Conference. 2022.
- Quantal correlated equilibrium in extensive form games. J. Černý, B. An and A. N. Zhang.

I derived the theory and wrote the manuscripts; Prof An and Dr Zhang provided insightful comments.

❖

In addition, during the course of my doctoral studies I (co-)authored the following 6 papers which do not fit into the narrative of this thesis and are hence omitted.

- Finite state machines play extensive-form games. J. Černý, B. Bošanský and B. An. In Proceedings of the 21st ACM EC Conference. 2020.
- Computing ex ante coordinated team-maxmin equilibria in zero-sum multiplayer extensive-form games. Y. Zhang, B. An and J. Černý. In Proceedings of 35th AAAI Conference. 2021.
- Solving pursuit-evasion games using pre-trained strategies. S. Li, X. Wang, Y. Zhang, H. Chan, J. Černý and B. An. In Proceedings of 37th AAAI Conference. 2023.
- Price of anarchy in a double-sided critical goods distribution system. D. Sychrovský, J. Černý, S. Lichau and M. Loebl. In Proceedings of 22nd AAMAS Conference. 2023.
- A unified perspective on deep equilibrium finding. X. Wang, J. Černý, S. Li, C. Yang, Z. Yin, H. Chan and B. An.
- Offline equilibrium finding. S. Li, X. Wang, J. Černý, Y. Zhang, H. Chan and B. An.

13/1/2023

Date

Jakub Černý

Abstract

Computational intelligence has become an integral part of human society. Algorithms recommend us news to read during breakfast, help us choose routes to reach workplaces, or suggest restaurants to dine at; but, in the end, it is us – humans – who have the final word in these interactions. And we choose as we please, driven by copious motives and goals. In many cases, we remain oblivious to the fact that our actions eventually shape the entire society, influencing other people's options and decisions. From this interdependence of our behavior stems the need for a theory of greater human interaction. And while traditional game theory offers an elegant body of work, its results are built upon an inherently flawed assumption – the absolute rationality of decision-makers, which, as many experiments in the past decades demonstrated, human psyché seems to show rather a lack of.

This thesis attempts to venture beyond this assumption and provides a series of results characterizing the implications of deviating from full rationality. To this end, it adopts perhaps the most widely accepted model of limited cognitive abilities called the *quantal response*. Quantal response assumes humans act stochastically, choosing actions with higher utilities more frequently. To aid in human decision-making, we study domain-agnostic integration of quantal response into two celebrated interaction archetypes: the models of *commitment* and *coordination*. These archetypes have been successfully applied to many specific real-world scenarios, and computed strategies were shown to greatly improve in efficacy when behavioral models are incorporated. Yet, the methods developed for solving these scenarios are not transferable to general models of commitment and coordination, where the integration of bounded rationality has never been addressed until now.

Commitment is an ability of a single leading agent to influence the course of play even before the interaction starts by publicly announcing a strategic behavior they will truthfully follow. In order for the commitment to be respected by other agents, the so-called leader has to serve a prominent role in society. The capacity to adopt this role is hence commonly reserved for established market leaders or large governmental institutions. Under mild assumptions, the commitment was shown to benefit the leader greatly, and carefully crafted strategies may hence assist in optimizing social good in the entire society. We show how to integrate quantal response into commitment models in order to account for human-type behavior, we identify the problem's computational complexity, and design algorithms computing commitment strategies with guaranteed convergence and bounded error.

Coordination then further improves the agents' strategic capabilities by letting them act upon their interests not only based on the reasoning about the leader's or their other opponents' behavior, but also by conditioning their strategies on external private signals. The process of selecting and revealing the signals is traditionally entrusted to a mediator mechanism, referred to as the correlation device. By optimizing the device's public distribution over signals, coordination facilitates reaching socially desirable outcomes previously considered unattainable. We investigate how quantal response affects the topology of the solution space, how computationally difficult it is to compute the solutions, and design algorithms that traverse the solution space while optimizing the signaling structure.

We evaluate the scalability and robustness of all the introduced methods on multiple domains characteristic to commitment or coordination scenarios. The reported results indicate that our methods are sufficiently precise and surpass the contemporary state-of-the-art non-convex optimization solvers by several orders of magnitude in terms of computation speed. We hope our efforts may expedite the adoption of game-theoretic methods for designing more efficient and egalitarian systems.

Acknowledgments

I admit that at the time the tenets of doctoral studies were still unbeknownst to me, I used to find lengthy acknowledgments and dedications rather pretentious and vastly exorbitant. Yet it is without surprise – the great amount of effort that goes into doctoral research and the mental toll it takes on a soul of a PhD student is difficult to understand without experiencing it firsthand. At this point in my life, I am better aware that the whole time when I navigated through the rapids of research, I rested on the shoulders of many, who drew me out of deep waters. Thanking them here is the least I can do.

It was an honor to be supervised by two brilliant researchers, Prof. An and Dr. Zhang, who scaffolded my entire PhD journey, providing all the guidance I needed but always willing to let me explore independently. I am grateful for the patience and understanding in the face of my mannerisms and the kindness I was welcomed with in the research group at the university. I would like to extend my thanks also to my other academic mentors, who I regard as friends and who shaped my view on research, Prof. Neufeld, Prof. Xu, and Dr. Bouveret. Working on projects together opened my eyes to broader problems beyond the narrow field of finite game theory I was delving into otherwise. Furthermore, I am indebted to Prof. Bošanský and Prof. Lisý, who introduced me to the aesthetic pleasure of mathematical interactions and the subtleties of bounded rationality.

I owe my gratitude to the research group I shared the lab and dinners with. In particular, to Jiuchuan for taking care of me during my first months in a foreign country on the other side of the globe. But also to Youzhi and Xinrun for our long technical discussions and to Hongxin for letting me rant on the everyday issues a PhD student encounters. My stay in Singapore would also be much more desolate without the outings with our small dinner squad of Tushar, Abhay, Joash, Marcus, and Subhrajit. I will always remember with fondness the philosophical debates with Tushar, and I am grateful we could partake in some of the major life-changing events together.

Over the years in Singapore, I was fortunate to be a part of the amazing Tanglin Kendo Club. I will cherish the memories of the great times we had in perpetuity. Their support and encouragement were invaluable to me, both on and off the bogu. I am especially thankful for the camaraderie of David *sensei*, Jo, and Shaun, who always looked out for me and helped to keep my ego in check. I found much comfort also in the conversations with my friends overseas. Their willingness to stay in touch and keep me updated on their lives made the distance between us feel so much smaller. The daily chats with Dan, funposting with Adam & Adam, and jokes on AI alignment with Andrej were steady constants that allowed me to decompress in times of research despair.

Finally, I wish to take a moment to express my deepest gratitude for the unwavering love and support my parents, my sister, and the rest of my family have given me. I would not accomplish anything in life without them, yet the unquenchable desire for knowledge keeps me away, for which I am sorry. No amount of words can do the influence they have on me justice, and I keep them in my mind at all times. A particular word of thanks to my bosom buddy Bára for fancying me over her mountains yearning.

❖

Furthermore, I want to acknowledge the financial support for traveling and the stipend I received for my doctoral research by the Singaporean Agency for Science, Technology and Research.

Catching the rest of the errors lurking in this manuscript is left as an exercise for the reader.

Contents

Part I

Prologue

1 Introduction

Social interaction is a cornerstone of human society. In contrast to solitary species, the evolution of human civilization led to the development of specialized professions that complement each other, creating a wide range of interdependent roles in our communities. A direct consequence of such a social order is that a daily life of an individual is nowadays largely influenced by many external factors[1], rather than being in hands of the said person. For example, the morality and legality of our actions is determined by local psychological and political climate manifested through a changing legislation, and the values of goods we need to sustain ourselves or labor we produce are affected by inflation and situations on global markets. And since humans could be driven by vastly different motivations, predicting an outcome of different social interactions remains conceptually difficult.

[1] Most importantly, it is affected by a behavior of other humans.

Fascination and desire to understand such complex dynamics motivated the formalization of a mathematical theory of interactions, most commonly referred to as *game theory* due to its focus on the simplest strategic representation of interactions: games. Contrary to the popular belief, (mathematical) games are neither a simple thought experiment, nor just a mean to superficial self-amusement. Characterized by distributing *payoffs* to players based on a combination of all their *actions*[2], mathematical games are capable of modeling many real-world situations. These include, but are not restricted to, scenarios in economics, social sciences, or security.[3] And as contemporary literature shows, leveraging game-theoretic design and analysis techniques in modern applications creates a more efficient and socially just society.

[2] In other words, the players' behavior is interdependent.

[3] T. Roughgarden. "Algorithmic game theory". *Communications of the ACM* 53:7, 2010, pp. 78–86.

Still, even with motivations and preferences reduced to mere payoffs, significant challenges prevail, with the most essential being *how to identify effective strategies*. To this end, game theory builds upon three key aspects of games. First, the specific *properties of domains*; because in many games, using the game-theoretic mathematical apparatus facilitates identifying hidden structures and symmetries within the games that further simplify formal analysis. Second, the *strategic capabilities* of players; for example, being able to commit to a premeditated behavior beforehand or condition the responses on actions of external mediators. Such options significantly widen the spectrum of players' strategic possibilities, enabling them to reach outcomes deemed unfeasible otherwise. Finally, the third aspect is the level of players' *reasoning abilities*. These include various modes of choice, subjective perception of payoffs, or inability to perform certain actions. This thesis focuses on the later two and studies properties and solving methods for games with players of varying degrees of rationality and strategizing capabilities.

1.1 Motivation

The earliest discoveries in game theory considered egalitarian, entirely rational players, based on the neoclassical model of *homo economicus*[4]. Pertaining to the assumptions of this idealized, self-interested economic consumer, the traditional game theory took advantage of the mutual rationality and equal standing of all involved parties to provide a rich and elegant body of work. The seminal result remains the introduction of the concept of Nash equilibrium – a ubiquitous situation when no player has an incentive to unilaterally deviate from their behavior.

[4] It is worth to note that while insisting on full rationality in purely human interactions proved erroneous, the contemporary world relies, to a large extent, also on human-machine interactions. On the side of the machine, this assumption still has its merits.

The postulates of Nash equilibrium are mathematically beautiful, yet at the same time unrealistically strong, oblivious to cognitive limitations of human decision-makers. As Herbert A. Simon, by many considered the forefather of behavioral economics, noted:

> "…in the real world the human behavior is intendedly rational, but only boundedly so, that there is room for a genuine theory of organization and administration."

In his pioneering work,[5] Simon asserted that human decision-making processes are intrinsically molded by limits on the players' knowledge and computational capabilities. These ideas, fusing the seemingly incompatible two-folded nature of human behavior – *intended*, yet *bounded* rationality – gave rise to the Carnegie school of economics. In light of vast evidence accumulated over decades of psychological decision-making experiments,[6] Simon's arguments leave no room for doubt in the contemporary scientific communities. It became increasingly evident that addressing limits of rationality leads to clearer explanations and better predictions of outcomes of human interactions.

[5] H. A. Simon. *Administrative behavior*. Simon and Schuster, 2013.

[6] C. F. Camerer. *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press, 2011.

❖

While the implications of deviating from full rationality have already been extensively studied in the most straightforward, egalitarian situations based on the original definition of Nash equilibrium, other celebrated models of interaction remain mostly unexplored. Two models stand out as the most influential[7]: Stackelberg equilibrium, in which one of the players has *commitment* power to lead the interaction while the opponent(s) respond to the leader's behavior, and correlated equilibrium, in which the players may *coordinate* their behavior through external private signals. These situations enhance the players' strategizing capabilities, which consequently drives the modalities of their intended rationality.

[7] These two models also motivate the title of this thesis.

The limited reasoning abilities of the players are then formalized though their model of bounded rationality. One of the key takeaways from decision-making experiments with human participants is the observation that instead of always choosing the utility-maximizing option, human players tend to systematically discriminate between different alternatives. Among the most renowned models of bounded rationality that replicate this behavior is *quantal response*. Quantal response posits

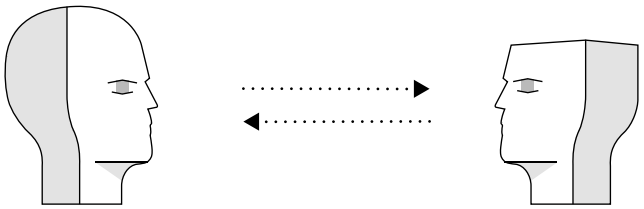
the players act stochastically, choosing higher-utility actions with higher probability, and is consistently regarded as one of the best predictors of human behavior.

❖

This thesis introduces novel solution concepts integrating the model of quantal response into Stackelberg and correlated equilibria in both one-shot (i.e., normal form) and sequential (i.e., extensive form) scenarios. We analyze these concepts' mathematical properties and design scalable methods to compute or approximate them. In the following sections, we give a brief overview of the introduced concepts and motivate them by several potential real world applications. We list our main technical contributions, and conclude by laying out the thesis' structure.

1.2 Boundedly rational equilibria beyond Nash

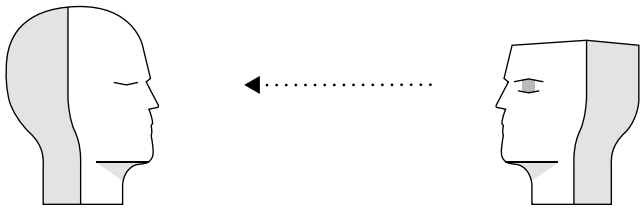
The story of this thesis began with a one-sided concept we refer to as *quantal Nash equilibrium*, which we feel compelled to introduce among its more fundamental siblings. Quantal Nash equilibrium is a unilaterally rational concept describing a situation when an entirely rational player faces a cognitively restricted opponent whose behavior follows the quantal response model[8].



[8] The following figures illustrate the principal differences between the equilibria we study. Here, the round head symbolizes the rational agent, while the cropped head indicates bounded rationality.

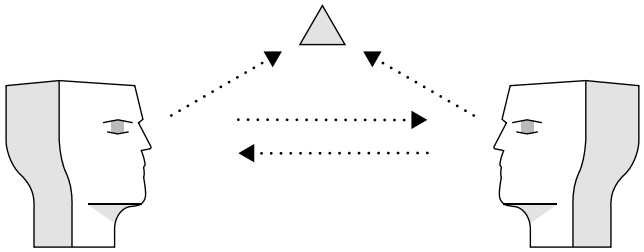
The concept is of a Nashian nature; the agents are assumed to observe each other's strategies and react accordingly. Interactions of this kind are common in many human-machine scenarios ranging from general communications with voice assistants or collaborative work with autonomous systems to specific applications like automated bidding in online auctions. One of the concept's surprising and counter intuitive properties is that the rational agent may be at a loss when adopting a quantal Nash strategy instead of a Nash strategy against a quantal agent. This motivates a generalization into a setting when the rational agent has the ability to commit to a strategy beforehand.

We coin the name *quantal Stackelberg equilibrium* for the corresponding concept. The equilibrium is again unilaterally rational, yet in this case, the committing agent is able to take advantage of their opponent's boundedly rational behavior and steer the interaction towards their preferred outcomes.



Using their privileged status, the leader announces and implements their strategy, letting the quantal agent merely observe their behavior and adapt to the situation at hand. This endeavor may yield a substantially higher payoff for the rational player over the traditional Nashian strategy. Stackelberg-type equilibria are also among the most frequently employed concepts in practice, with quantal commitment being particularly suitable for security applications like long-term continual patrolling or hardening of computer network security. Still, the role of the rational agent remains relatively passive and does not encompass any explicit communication scheme among them and their opponents. Communication then becomes a focal point of the last equilibrium we appraise.

We formalize the aforementioned boundedly rational mediating concept as the *quantal correlated equilibrium*. It characterizes a situation when the agents may contemplate not only the presumed behavior of their opponents but they could also condition their strategies on signals received from an external source.



Similarly to quantal Nash equilibrium, also the quantal correlated equilibrium is an egalitarian concept short of commitment. The process of selecting and revealing the signals is entrusted to a mediator mechanism, a so-called correlation device[9]. Coordinating the agents via signals is beneficial for the system; it enables reaching outcomes that would not be viable otherwise. Typical applications can be sought in automated recommendations, coordination of sovereign entities in umbrella organizations, or implementation of public policies. The role of the correlation device is comparable to that of the committing agent, yet without the immediate option to act. As such, it may suitably describe, e.g., various administrative bodies.

[9] The correlation device is visualized as the small overhead triangle in the figure.

1.3 Aims and contributions of the thesis

This thesis aims to characterize the consequences of deviating from full rationality domain agnostically and for broader notions of equilibria, described in the previous section. The fundamental questions we attempt to answer are

1. how can we compute boundedly rational equilibria, and
2. is their computational complexity lower than of their rational counterparts?

As an answer to the first question, for each of the introduced boundedly rational equilibria we provide an empirically efficient parametric algorithm capable of computing the concept. Neither of the algorithms relies on standard techniques. For example, attempting to find a quantal Stackelberg equilibrium using straightforward gradient-based approaches is bound to fail. As we show, the natural formulations of quantal Stackelberg equilibria are non-concave in general, and the number of local optima may grow linearly in the number of actions. In a style of Dinkelbach,[10] we hence reformulate the fractional representation of quantal Stackelberg equilibrium into an equivalent problem of finding a root of a convex function. Through linearization into a mixed-integer linear program, we arrive at the first main result in normal form games.

[10] W. Dinkelbach. "On nonlinear fractional programming". *Management Science* 13:7, 1967, pp. 492–498.

Theorem 1.1 (formalized in Chapter 4 as Proposition 4.6)**.** *There exists an algorithm approximating the optimal strategy the rational player commits to against a quantal player in normal form games that guarantees a loss in utility upper bounded by $\mathcal{O}(|A|/K^2)$, where $|A|$ is a number of actions of the boundedly rational player and K is a number of segments used to linearize the behavioral model.*

In extensive form games, the linearization is more cumbersome, because of the inherent curse of strategy-space dimensionality in this class of games, and many decision points. Nevertheless, we derive a conceptually similar algorithm.

Theorem 1.2 (formalized in Chapter 5 as Proposition 5.3)**.** *There exists an algorithm approximating the optimal strategy the rational player commits to against a quantal player in extensive form games that guarantees a loss in utility upper bounded by $\mathcal{O}(|A|(1/K^2+1/L))$, where $|A|$ is a number of actions of the boundedly rational player, K is a number of segments used to linearize the behavioral model, and L is a width of an interval controlling the error of bilinear terms estimates.*

To compute a quantal correlated equilibrium, we formulate a robust homotopic algorithm capable of traversing the principal branch of equilibrial correspondence[11]. We employ carefully designed variable substitutions and model reformulations that ameliorate numerical issues precipitated by the steep quantal response functions or wide ranges of utilities. As a consequence, we are able to simultaneously trace the equilibrium and gradiently optimize a probability distribution over the signals while maintaining the homotopy's convergence guarantees.

[11] E. L. Allgower and K. Georg. *Numerical continuation methods: An introduction*. Vol. 13. Springer Science & Business Media, 2012.

Theorem 1.3 (formalized in Chapter 6 as Algorithm 5)**.** *There exists an algorithm that efficiently approaches the coordination equilibrium with quantal players in normal form games from the center of the strategic simplex by continual corrections of increasingly rational strategies via a Gauss-Newton method.*

The innate properties of extensive form games keep causing issues even in coordination scenarios. Continual tracing corrections we employed in normal form games quickly become unfeasible as the number of decision points increases. Tailoring a decomposable corrector allows us to formulate another homotopic method.

Theorem 1.4 (formalized in Chapter 7 as Algorithm 7)**.** *There exists an algorithm that efficiently approaches the coordination equilibrium with quantal players in extensive form games from the center of the strategic simplex by continual corrections of increasingly rational strategies via a conjugate gradient method.*

Due to the ability of the quantal response model to approximate entirely rational behavior up to a desirable error, the general answer to the second question is *negative*: quantal equilibria are at least as difficult to compute as the traditional solution concepts. Still, we are able to identify two special cases when commitment equilibria[12] can be found more efficiently.

[12] I.e., the more computationally complex of the two.

Theorem 1.5 (formalized in Chapters 4 and 5 as Propositions 4.5 and 5.2)**.** *There exist conditions characterizing when the optimal strategy the rational player commits to against a quantal player can be approximated in polynomial time in a form of*

- *two scalar inequalities in normal form games; and*
- *a matrix inequality in extensive form games.*

Moreover, our extensive empirical evaluations provide evidence the designed universal algorithms perform exceptionally well in practice, convincingly outperforming the baseline approaches in computational time by several orders of magnitude.

1.4 Structure of the thesis

The dissertation begins with a discussion of related work in Chapter 2. We review existing solution concepts and behavioral models and classify them according to Simon's taxonomy of the emergence of bounded rationality. We provide numerous viewpoints drawing primarily from operations research, behavioral economy, and cognitive or personality psychology. We briefly deliberate also the historical roots of the quantal response and its construction.

Chapter 3 introduces the fundamental concepts from game theory. We present the rudimentary representations of interactions in forms of normal and extensive form games and construct the associated sequence representation. We provide definitions of the three established entirely rational solutions pertaining to this thesis: the Nash, Stackelberg, and correlated equilibria, and their principal properties and

computational complexities. The rest of the chapter presents boundedly rational solutions, particularly the quantal response and quantal Nash equilibria, and lays down our exploratory results on the latter.

❖

The following first major content part of the thesis is dedicated to optimal commitment in the presence of a quantal adversary. Chapter 4 formalizes the concept as a quantal Stackelberg equilibrium in normal form games and studies its properties. We show that commitment power frequently benefits the leading agent, yet not without exceptions. Finding the optimal strategy then necessitates the development of a specialized method stemming from fractional programming. The chapter demonstrates its efficacy in the experiments. Chapter 5 then attempts to generalize the same approach to extensive form games. It analyzes the concept's topology and computational complexity, highlighting the key differences distinguishing it from its normal form analog. Designing a computational method becomes more involved, and we exert substantial effort to derive its formal guarantees. The method's performance is evaluated in practice at the end of the chapter.

The second major content part presents our contributions to the study of quantal coordination. Chapter 6 defines the concept as a quantal correlated equilibrium in normal form games via a traditional construction of solution concepts with a signaling device. The chapter thoroughly investigates its topological and computational attributes, as well as relations to the precursor equilibria. The convenient continuous geometry of quantal correlation facilitates the introduction of a homotopic algorithm that tends to the equilibrial strategies along the homotopy curve. The viability of the method is again analyzed empirically. Howbeit, extending the results into extensive form games requires a slightly different viewpoint, as following the same approach proves computationally impractical. To this end, Chapter 7 presents the sequential interpretation of quantal correlation and adopts a more suitable iterative method to trace the equilibrium in a homotopic manner still similar to the algorithm's normal form counterpart.

❖

We conclude the dissertation in Chapter 8. It briefly reiterates the principal motivation for studying boundedly rational concepts in game theory and summarizes and discusses the main contributions. The chapter also points out several related directions of research in which the results presented here could be expanded upon to design a new generation of more efficient and socially just systems.

2 Related Work

In this chapter, we briefly review the existing literature on bounded rationality in game theory and bring the models introduced here into broader context. We also explain how these models complement, or relate to, the behavioral model considered in this thesis, and their potential unification.

As hinted in the introduction, the theory of bounded rationality was developed in an effort to generalize the entirely rational, traditional decision-making and conform with the findings of behavior sciences.[13] Initially, the experimental evidence was regarded merely as "special cases", boundary conditions where the standard model did not fully apply. As more and more evidence began to accumulate, behavioral economists started to entertain the thought it may be actually vice versa: it is the bounded, rather than full, rationality that forms the interactions in modern societies.[14] And this imperfect mode of interaction is inextricably linked with our world's intricacy. The emergence of complex social structures would not be possible without interactions between boundedly rational agents. At the same time, our flawed reasoning is in itself a consequence of the complex environment we interact with. It is a natural byproduct of cognitively restricted decision-making in situations beyond our comprehensibility limits.

Building up on the decades of behavioral experiments, Simon proposed the following taxonomy of emergence of bounded rationality:[15]

1. limited knowledge of the world;
2. limited ability to evoke the knowledge;
3. limited ability to work out consequences of actions;
4. limited ability to conjure up possible courses of action;
5. limited ability to cope with uncertainty; and
6. limited ability to adjudicate among competing wants.

The models introduced in single-agent decision making and behavioral economy were soon adopted in multi-agent game-theoretic settings as well. Established solution concepts were refined to explain the empirical findings better or provide more robust solutions against cognitively restricted adversaries. Incorporating different limitations suggested by Simon proved to be a path to more accurate models, better suited for situations where the genuine intentions of the agents are unknown

[13] K. V. Velupillai. "Foundations of boundedly rational choice and satisficing decisions." *Advances in Decision Sciences*, 2010.

[14] C. Lee. "Bounded rationality and the emergence of simplicity amidst complexity". *Journal of Economic Surveys* 25:3, 2011, pp. 507–526.

[15] H. A. Simon. "Bounded rationality in social science: Today and tomorrow". *Mind & Society* 1:1, 2000, pp. 25–39.

or where they systematically fail to pick their utility-maximizing options. We provide an overview of approaches commonly found in the contemporary literature on the topic, divided into six main sections: the generalizations and refinements of the original Nash equilibrium, one-shot subjective utility behavioral models, iterated (hierarchical) models, cognitive and personality models, models based on computation theory, and model-uncertain concepts. Let us start with the first category because of its close connection to Nash equilibrium.

2.1 Nash-related models

Elimination of unsatisfactory choices is one of the most straightforward heuristics rational agents may use to restrict their potentially large action spaces[16]. In a multi-agent setting, a widely used, appropriate measure of suboptimality is *dominance*. A strategy is called (strictly) dominated if another strategy provides (strictly) higher utility for any possible behavior of the agent's opponents. Rationalizable strategies are then a subset of all strategies such that neither (strictly) dominates the other.[17] Nash equilibria always consist of rationalizable strategies; the converse is not true, though. Allowing the agents to play non-rationalizable or even non-Nashian rationalizable strategies hence corresponds to the inclusion of multiple limitations from Simon's taxonomy, perhaps most importantly, the lack of knowledge about the opponents' strategic capabilities.

While rationalizability generalizes Nash equilibrium, other concepts related to relaxing the rationality assumptions further restrict it. An example of such an approach is the trembling hand perfection.[18] The idea of trembles is that agents may choose suboptimal actions with an upper-bounded non-zero probability ϵ. As this upper bound vanishes, the equilibria in these *perturbed games* reach a Nash equilibrium in the original game. Trembling hand perfection amends a specific issue inherent to some entirely rational concepts: the fact that a small "slip of the hand" resulting in accidentally choosing an unintended strategy may substantially change the opponents' behavior. A similar concept is a proper equilibrium which further refines perfection, noting that some strategies may constitute a trembling hand equilibrium just on account of the existence of strictly dominated strategies.[19]

With the strictly positive upper bound ϵ, both perfection and propriety relate to an agent's limited ability to follow a course of action. In this sense, these equilibria are similar to the quantal response. The critical difference lies in the quantal response's systemic discrimination based on the actions' expected payoffs.

2.2 One-shot behavioral models

While the models described so far are characterized by their relation to Nash equilibrium, now we move deeper into the area of behavioral economics and psychology, where predictions may vary greatly and differ substantially from expected Nashian behavior. The models in this area are most frequently driven by either behaviorally

[16] Iterated elimination then leads to a hierarchical model we discuss in Section 2.3.

[17] B. D. Bernheim. "Rationalizable strategic behavior". *Econometrica: Journal of the Econometric Society*, 1984, pp. 1007–1028.

[18] R. Selten. "Reexamination of the perfectness concept for equilibrium points in extensive games". *International Journal of Game Theory* 4, 1975.

[19] R. B. Myerson. "Refinements of the Nash equilibrium concept". *International Journal of Game Theory* 7:2, 1978, pp. 73–80.

constrained optimization or heuristic decision making. Because this thesis deals with computational rationality associated with maximization and calculations of probabilities and expected utilities, we focus predominantly on the first group. We refer the reader inquisitive about the heuristic models to the appropriate literature.[20]

Arguably the most prominent remain the models of human conduct towards *risk and loss*. Risk refers to the ability of agents to cope with uncertainty. The literature recognizes three distinct risk attitudes: risk aversion, risk neutrality, and risk seeking.[21] Risk aversion is an umbrella term for describing the tendency to favor low uncertainty outcomes in lieu of more precarious options, irrespective of the expected payoff. Risk seeking is its corresponding antipode model, preferring higher uncertainty. Risk-neutral agents then act without bias towards either low or high predictability. Both risk aversion and risk seeking may be explained in terms of the traditional expected utility maximization framework via appropriate transformations of the agents' utility functions; their concavity is attributed to risk aversion, while convexity is to risk seeking.

While risk attitudes concentrate on quantitative differences in uncertainty, loss attitudes are concerned with previously experienced or expected future gains and losses. Analogously to risk aversion, when facing an equally plausible profit or loss, loss-averse agents aim to avoid the latter. As a consequence, they become risk-seeking to avert a more considerable loss while risk-averse with respect to potential gains. A standard trope in the literature related to this phenomenon is that "a dollar lost is two dollars gained". Loss aversion is also a cornerstone of the celebrated *prospect theory*,[22] which further refines it by asserting that humans tend to overweight low-probability options while underweighting outcomes that are almost certain.

The postulates of prospect theory, or loss aversion in general, have been challenged in the past two or three decades by several competing theories, often stemming from the *fourfold pattern* of risk preferences.[23] They suggest that boundedly rational agents exhibit opposite behavior under certain circumstances, absent of loss aversion. The agents' diminished sensitivity to values may lead even to a utility function shaped inversely to a conventional prospect theory function. The most apparent difference observed in the experiments is a pronounced concavity in the loss domain, implying risk aversion instead of risk-seeking, especially as losses grow.

Regardless of the concrete shapes of utility functions, both risk and loss attitudes are frequently adopted in game-theoretic analyses. They model the agents' inability to resolve the options framing psychological conflicts arising from the subjective perception of either actions' values or uncertainty in outcomes. And, due to their convenient representation via utility functions, they can be integrated effortlessly into larger decision-making solutions, and we do so later in Chapter 4 as well. This brings us to a behavioral model central to this thesis: the *quantal response*.

The origins of quantal response[24] may be traced to stochastic choice approaches from discrete choice econometrics, and their integration into game-theoretic solution concepts.[25] The key difference between traditional expected utility maximization and quantal response is that in the latter, the agents are not perfect optimizers,

[20] G. Gigerenzer and R. Selten. *Bounded rationality: The adaptive toolbox*. MIT Press, 2002.

[21] J. W. Pratt. "Risk aversion in the small and in the large". In: *Uncertainty in Economics*. Elsevier, 1978, pp. 59–79.

[22] D. Kahneman and A. Tversky. "Prospect theory: An analysis of decision under risk". In: *Handbook of the Fundamentals of Financial Decision Making: Part I*. World Scientific, 2013, pp. 99–127.

[23] H. Markowitz. "The utility of wealth". *Journal of Political Economy* 60:2, 1952, pp. 151–158.

[24] In the literature sometimes also called the "quantal choice".

[25] D. L. McFadden. "Quantal choice analaysis: A survey". In: *Annals of Economic and Social Measurement, Volume 5, number 4*. NBER, 1976, pp. 363–390.

but rather "better-responders", choosing actions with higher expected utility more frequently.[26] Quantal responses are commonly constructed either via structural or reduced-form approaches. In the structural approaches, the stochastic sub-optimal responses arise as rational reactions to unobserved payoff disturbances. In contrast, the reduced-form approaches assume the responses simply follow some particular functional forms that satisfy additional properties akin to continuity, interiority, or independence of irrelevant alternatives. The methods introduced in this thesis apply exclusively to the reduced-form approaches, which appear to be more prevalent in the literature[27].

[26] R. D. McKelvey and T. R. Palfrey. "Quantal response equilibria for normal form games". *Games and Economic Behavior* 10:1, 1995, pp. 6–38.

[27] More specifically, we focus on quantal responses in a form for a univariate function applied to an action's expected utility, normalized over the alternatives.

2.3 Hierarchical behavioral models

The models introduced so far left the process of how the responses or equilibria are attained unexplained, assuming the agents either naturally converge to their behavior via repeated interaction or the predictions alone capture their strategic thinking with sufficient precision. Yet, empirical evidence suggests the agents may at a time reach or approximate these deliberative concepts through a series of "pre-equilibria" with little to no precedent experiences.[28] These processes may hence hardly be attributed solely to learning; rather, they are a consequence of a more involved strategic thinking. Hierarchical models arose as a formalization of such a process, imposing limits on what the agents may infer about their opponents' behavior. As such, they aim to replicate the agents' limitations in conjuring up their actions' consequences or potential future courses.

[28] V. P. Crawford, M. A. Costa-Gomes, and N. Iriberri. "Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications". *Journal of Economic Literature* 51:1, 2013, pp. 5–62.

Rationalizability, which we discussed at the beginning of this chapter, is an aforesaid concept that relaxes the assumption of direct attainment of the equilibrium while maintaining its key component – the (iterated) rationality. As a side product, it induces a finite-level hierarchical model of *k-rationalizability*. In this hierarchy, a level 0 consists of the entire strategy space. The strategies at level k are then those that may act as best responses to an arbitrary $k-1$-level strategy. In interactions among two agents, k-rationalizable strategies correspond to precisely those that survive k rounds of the iterated elimination of (strictly) dominated strategies.

The k-rationalizable strategies impose strong assumptions on the agent's strategic thinking capabilities, aligning them via k at each level of the hierarchy. The *level-k* model is its heterogeneity-tolerant refinement, allowing the k to vary across the agents.[29] In level-k models, the zero level represents the agents' naive expectations of their opponents' potential behavior and is commonly referred to as an "anchor". Similarly to k-rationalizability, a level k is then inductively built from the responses to strategies at level $k-1$. Moreover, each agent's level is commonly assumed to be drawn from an estimated prior distribution.

[29] R. Nagel. "Unraveling in guessing games: An experimental study". *The American Economic Review* 85:5, 1995, pp. 1313–1326.

The model of *cognitive hierarchy* then generalizes the level-k model by assuming the agents respond not merely to behaviors from level $k-1$ but to a mixture of all previous levels.[30] A distribution often employed in the literature is Poisson due to being parsimonious, with the frequency of the higher levels quickly dropping

[30] C. F. Camerer, T.-H. Ho, and J.-K. Chong. "A cognitive hierarchy model of games". *The Quarterly Journal of Economics* 119:3, 2004, pp. 861–898.

off to zero. Contrary to the level-k model, the strategies at level k in the cognitive hierarchy are not k-rationalizable, which makes the model more flexible and better suited for accommodating other behavioral imperfections observed in the data.

The quantal response may be conveniently combined with the hierarchical models by assuming the agents' responses have a quantal nature. The hybrid models are known to outperform their non-quantal alternatives on many scenarios.[31] Coupling the methods introduced in this thesis with the hierarchical models hence has the potential to significantly improve the efficiency of commitment or coordination strategies in practice.

[31] J. R. Wright and K. Leyton-Brown. "Level-0 meta-models for predicting human behavior in games". In: *Proceedings of the 15th ACM Conference on Economics and Computation*. 2014, pp. 857–874.

2.4 Cognitive and personality models

The behavioral models described in the previous two sections replicate well the actions human agents take in the experiments. Still, their rationale for making these decisions is only implicit, hinging on the presumptive limitations of human strategizing processes. Here, we introduce two models that instead strive to reproduce the underlying driving mechanisms. They induce the observed behavior as a mere byproduct of a more complex procedure arising from agents' limited ability to process the knowledge about the world they found themselves in.

The first model aims to justify the substantial dispositional tendencies in rationality or strategic and stealth abilities the agents exhibit[32] by their personality differences. Three distinct deceptive personality traits are dominant in conflict-seeking individuals, and they are referred to as the *Dark Triad* of narcissism, Machiavellianism, and psychopathy.[33] They can be differentiated by the way the individuals high in these traits engage in deceptive behavior. Narcissists are known for self-deception and risk-seeking due to their overconfidence and unrealistic optimism. Machiavellianists are more adaptive and strategic and capable of anticipating the behavior of their opponents. Finally, psychopaths are reckless, they tend not to adhere to rational strategies well, and their abilities to predict penalties are limited. In repeated interactions, the three traits exhibit significantly diverging patterns of behavior. Similarly to risk and loss attitudes, even the Dark Triad traits can be modeled by either modifying the utility functions in accordance with each trait's attributes or applying an appropriate utility discounting.

[32] Especially in highly adversarial situations like cyber crime.

[33] D. L. Paulhus and K. M. Williams. "The dark triad of personality: Narcissism, Machiavellianism, and psychopathy". *Journal of Research in Personality* 36:6, 2002, pp. 556–563.

On the other hand, the *instance-based learning* stems from several essential elements of cognitive psychological mechanisms like memory, learning, and forgetting.[34] This model characterizes a dynamic decision making process of an agent who recalls and associates instances of past decisions from memory on the basis of their similarity. The utility feedback the agents receive is propagated through this memory, influenced by noise and memory decay. When facing a newly encountered situation, the agents estimate their expected utilities from history and act stochastically, proportionally to their expectations. Instance-based learning hence models the agents' limitations in adjudicating among competing options due to their restricted recollection abilities. The association process itself may be integrated into

[34] C. Gonzalez, J. F. Lerch, and C. Lebiere. "Instance-based learning in dynamic decision making". *Cognitive Science* 27:4, 2003, pp. 591–635.

the more considerable decision making models as a subjective, history-dependent utility function.

The stochastic choice in instance-based learning is, by default, implemented as soft maximization over the options and can be hence regarded as a form of quantal response. The model easily permits more elaborate quantal responses. Vice versa, both the dark triad and the instance-based learning may complement this thesis' methods in suitable applications via pertinent utility functions.

2.5 Computational models

Strategies in repeated or sequential interactions are usually represented as mappings from individual situations an agent may encounter to actions to be taken therein. Mappings are mathematically convenient to work with, yet they ignore any structure that may be present in the strategies and inadvertently also the human tendency to identify and favor patterns.[35] One simple approach that captures these implicit structures and characterizes the strategies' implementation complexity is to encode them as results of algorithmic processes in some models of computation. In the context of game theory, the commonly considered models are either deterministic finite automata of Turing machines. Bounding their complexity corresponds to imposing limits on the agents' abilities to evoke knowledge about the interaction, e.g., by restricting their memory to a finite number of past decisions.

Deterministic finite automata is perhaps the most straightforward example of such models. They were introduced to meaningfully reduce the massive strategy space and model strategic complexity in repeated games, where a single base game is played over a period of many[36] iterations.[37] Each state in the automaton has an associated action the agent takes at that state, and the transitions represent the simultaneously performed actions of their opponents. The base game is stationary; this assumption greatly simplifies the constructions as states permit arbitrary actions. The size of a strategy is typically determined by a number of states in a minimal automaton implementing it, but more complex measures reflecting the transitional structure have also been considered. The equilibrial payoffs in repeated games with automata strategies are known to approach individually rational payoffs in the base game in the limit and invoke human behavioral strategies in the finite case. For example, the bounded complexity can justify otherwise rationally unfeasible cooperation in a finitely repeated prisoner's dilemma.

Automata are a simplistic model with low expressibility because they lack explicit memory outside of the states themselves. *Turing machines* constitute a natural generalization that enables making decisions on the basis of the entire history of the interaction with greater ease, receiving the opponents' actions on the input and outputting a desired action to take. In repeated games, Turing machine strategies were shown to exhibit similar properties as automata. They yet again promote collaborative outcomes when the number of states is bounded, as shown on the example of prisoner's dilemma.[38] Turing machines have also been employed as strategies

[35] M. W. Eysenck and M. T. Keane. *Cognitive psychology: A student's handbook*. Taylor & Francis, 2000.

[36] Potentially even infinitely many. In that case, the original strategy space is infinite.

[37] A. Rubinstein. *Modeling bounded rationality*. MIT Press, 1998.

[38] N. Megiddo and A. Wigderson. "On play by means of computing machines: preliminary version". In: *Proceedings of the 1986 Conference on Theoretical Aspects of Reasoning About Knowledge*. Morgan Kaufmann Publishers Inc. 1986, pp. 259–274.

in computable uniform sequences of sequential games.[39] Here, a Nash (resp. sequential) equilibrium of the underlying game reached in the limit translates to a polynomial-time Turing machine implementable Nash (resp. sequential) equilibrium in the computable sequence.

Combining computational models with quantal response appears to be rather tricky. One possibility is to make use of the strategic complexity as a subgame selection method over which the agents act quantally. It remains an open question whether, e.g., Turing machines may implement quantal strategies and how it would affect the topology of the associated quantal equilibria.

[39] J. Y. Halpern, R. Pass, and L. Seeman. "Computational extensive-form games". In: *Proceedings of the 17th ACM Conference on Economics and Computation*. 2016, pp. 681–698.

2.6 Model uncertainty and robustness

Several game-theoretic concepts assumed a situation when the agents' behavioral models are not identified without discrepancies prior to the interaction[40]. Instead, the agents may hold beliefs about their opponents' motives and biases, and they adjust their own behavior accordingly. The resulting strategies are supposed to be robust to a multitude of behavioral profiles.

The most well-known is the model of *Bayesian games*.[41] In Bayesian games, each agent belongs to a particular class of types; the type determines their utility function. The agents are aware of their type but in doubt about their opponents' exact types beyond their class. The uncertainty is expressed via a prior joint probability distribution over the class profiles. The expected outcomes are then calculated using Bayesian probability, which enables to escape the infinite epistemic reasoning.

Another model assumes a commitment scenario where the leader suspects their opponents' are satisficers that may choose any of the responses within an *ϵ-reward interval* centered around their optimal strategies.[42] The aim is hence to optimize against all choices that fall within this interval. The essential advantage of this approach is that it is conceptually simple and easy to implement. The experimental results also indicate that it performs better in situations with human adversaries than entirely rational or overly conservative methods.

Still, the ϵ-deviations considered in this model are subjected to the same criticism as the trembling hand equilibria – they lack more systematic dependency on expected utility. The *monotonic maximin* was proposed as a more structured solution providing theoretical guarantees of optimality against agents acting according to any reduced-form quantal response function.[43] The monotonicity conditions are encodable via linear constraints, which enables to compute (a variant of) the equilibrium efficiently. Again, the empirical results speak favorably of the equilibrium in lieu of less robust concepts though it has not been compared to ϵ-deviations.

In general, model uncertainty methods constitute well-grounded attempts to accommodate the ambivalence arising from the agents' limited knowledge of the world. Some are directly related to quantal response; others show clear potential to enhance this thesis' work, e.g., by conditioning signaling in quantal correlated equilibria on agents' types.

[40] We leave out the minimax decision rule due to being a "no model" concept.

[41] J. C. Harsanyi. "Games with incomplete information played by "Bayesian" players, I–III Part I. The basic model". *Management Science* 14:3, 1967, pp. 159–182.

[42] J. Pita, M. Jain, M. Tambe, F. Ordónez, and S. Kraus. "Robust solutions to Stackelberg games: Addressing bounded rationality and limited observations in human cognition". *Artificial Intelligence* 174:15, 2010, pp. 1142–1171.

[43] A. X. Jiang, T. H. Nguyen, M. Tambe, and A. D. Procaccia. "Monotonic maximin: A robust Stackelberg solution against boundedly rational followers". In: *International Conference on Decision and Game Theory for Security*. Springer. 2013, pp. 119–139.

❖

This chapter described the most important concepts pertaining to bounded rationality in the context of interactions between many self-interested agents. It highlighted their relation to the quantal response model and briefly delineated how they could complement or be directly integrated into the model. Now that we familiarized ourselves with the possible approaches, we may begin to formalize the basics.

3 Game-Theoretic Preliminaries

This chapter introduces the fundamentals of game theory. More specifically, we focus on non-cooperative interactions among independent and self-interested agents. The basic unit therein is an individual, unlike in cooperative game theory, where the agents receive their payoffs as a group.

The first part of this chapter is dedicated to the formal description of the two most common domain agnostic representations of games: the normal form for one-shot games and the extensive form for sequential games. The second part then discusses various solutions of interactions represented using these forms. We analyze their computational complexity and bring to the fore standard methods for finding them. In particular, we emphasize how they relate to methods studied further on in this thesis. The three rational equilibria of our interest are Nash, Stackelberg, and correlated equilibrium. Later we turn our attention to two boundedly rational equilibria – the quantal response and the one-sided quantal Nash equilibrium.

Selected theoretical results from the second half of this chapter were published as D. Milec, J. Černý, V. Lisý, and B. An. "Complexity and algorithms for exploiting quantal opponents in large two-player games". In: *Proceedings of the 35th AAAI Conference on Artificial Intelligence*. 2020.

The definitions in this chapter are well-established notions used consistently across the literature. The formalization, as well as the statements of all theorems, are rephrased from the book of Shoham and Leyton-Brown.[44]

[44] Y. Shoham and K. Leyton-Brown. *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press, 2009.

3.1 Representation forms

Games studied in game theory are well-defined mathematical objects. They are composed of four basic components: the *players*[45] who engage in the game, the *actions* available to them at each decision point, the *payoffs* for each outcome, and the *information* they have about the game.[46] We distinguish between two standard forms.

[45] To distinguish the mathematical games from general interactions, we refer to the actors as players instead of agents.

[46] E. Rasmusen. *Games and information: An introduction to game theory*. Blackwell, 2001.

3.1.1 Normal form

The normal form is a standard game representation for one-shot games. In these games, each player moves only once and all actions are taken simultaneously.

Definition 3.1. *A normal form game (NFG) is a tuple $G = (N, A, u)$, where*

- *N is a set of n players, indexed by i;*
- *$A = A_1 \times \cdots \times A_n$, where A_i is a set of actions for player i; and*
- *$u = (u_1, \ldots, u_n)$ where $u_i : A \to \mathbb{R}$ is a utility function for player i.*

Prisoner 2 \ Prisoner 1	Cooperate	Defect
Cooperate	-1,-1	0,-4
Defect	-4,0	-3,-3

$$
\begin{aligned}
N &= \{\text{Prisoner 1, Prisoner 2}\} \\
A &= \{\text{(C)ooperate, (D)efect}\} \times \{\text{(C)ooperate, (D)efect}\} \\
u &= (\{\text{CC}\rightarrow -1, \text{CD}\rightarrow -4, \text{DC}\rightarrow 0, \text{DD}\rightarrow -3\}, \\
&\quad \{\text{CC}\rightarrow -1, \text{CD}\rightarrow 0, \text{DC}\rightarrow -4, \text{DD}\rightarrow -3\}) \\
\Pi &= A
\end{aligned}
$$

Figure 3.1: (Left) An example of a payoff matrix of a normal form game. Each column of the depicted matrix is labeled by an action/strategy of the first player, while every row is labeled by an action/strategy of the second player. The tuples in the matrix denotes the utilities of the first player and the second player, respectively. (Right) The game's formal NFG representation.

Utility functions in normal form games may be visualized via payoff matrices. For a given game, the dimensionality of the associated payoff matrix is equal to the number of players participating in the game. Each dimension is indexed by the actions of the corresponding player, and the elements of the matrix are tuples of utilities.

Example 3.1. *Consider a two-player game in Figure 3.1. The depicted payoff matrix describes the* prisoner's dilemma*, a standard game modeling a situation when two members of a criminal gang are arrested and kept in isolated confinements with no means of communication. Both are given a chance to testify against the other prisoner in exchange for a lesser charge. If they both decide to* Cooperate *with the other prisoner and remain silent, each of them serves a year in prison. Vice versa, if they* Defect *the gang and testify, they are sentenced to 3 years. The combined choices lead to a situation where the collaborator is absolved of guilt while the other prisoner serves 4 years.*

Strategies can be seen as policies for playing the game. Deterministic strategies are referred to as *pure strategies*. Individual pure strategies of player i are denoted $\pi_i \in \Pi_i$, each assigning one action from A_i. Stochastic strategies are – in game-theoretic lingo – called *mixed strategies*, and denoted $\delta_i \in \Delta_i$ for player i. The set Δ_i is a set of all probability distributions over Π_i. By $\delta_i(a_i)$ we denote the probability of player i taking an action $a_i \in A_i$. A strategy profile is a tuple of pure strategies $\pi = (\pi_1, ..., \pi_n) \in \Pi$ or mixed strategies $\delta = (\delta_1, ..., \delta_n) \in \Delta$ that fully determines how the game will be played. As an example, the entire normal form formalization of the prisoner's dilemma is given on the right of Figure 3.1.

3.1.2 Extensive form

Games in normal form describe singular situations. In contrast, the extensive form represents sequential interactions spanning many situations where players move in turns. A structure of a game in extensive form can be visually represented by a directed rooted tree. Each node in the tree represents a different situation in the game, and the root corresponds to the game's starting point. In every node, exactly one player acts. The edges denote the actions that can be taken by the player who acts in the particular node. All actions are deemed deterministic so that they are always

[47] The situations – the nodes of the tree – are also commonly called game states. We use these terms interchangeably.

correctly executed. Every situation[47] is uniquely determined by a sequence of moves executed by all players since the game began. Limited observations of the players are modeled via grouping nodes among which a player cannot distinguish into information sets. Games with perfect information can be seen as games with a trivial imperfect information structure, i.e., games in which each information set contains only a single node. Extensive form games may also represent stochastic events by introducing random moves of a special *Nature* player.

Definition 3.2. *An extensive form game (EFG) is a tuple $G = (N, H, Z, A, C, \chi, \rho, \sigma, I, u)$, where*

- *N is a set of n players;*
- *H is a set of nodes with $h_0 \in H$ being the root;*
- *$Z \subseteq H$ is a set of terminal nodes;*
- *A is a set of actions;*
- *$C : A \to [0, 1]$ is a probability function for performing a random action;*
- *$\chi : H \setminus Z \to 2^A$ is the action function, which assigns each node a set of actions;*
- *$\rho : H \setminus Z \to N$ is the player function, which assigns to each nonterminal node a player i who acts in that node;*
- *$\sigma : H \setminus Z \times A \to H$ is the successor function, which maps a node and an action to a new node, and graph induced by σ is a rooted tree;*
- *$I = (I_1, \ldots, I_n)$, where $I_i = (I_{i,1}, \ldots, I_{i,k_i})$ is a set of equivalence classes on $\{h \in H : \rho(h) = i\}$[48] with the property that $\chi(h) = \chi(h')$ and $\rho(h) = \rho(h')$ whenever there exists a j for which $h \in I_{i,j}$ and $h' \in I_{i,j}$; and*
- *$u = (u_1, \ldots, u_n)$, where $u_i : Z \to \mathbb{R}$ is a real-valued utility function for player i on the terminal nodes Z.*

[48] In other words, I_i is a partition of player i's nodes.

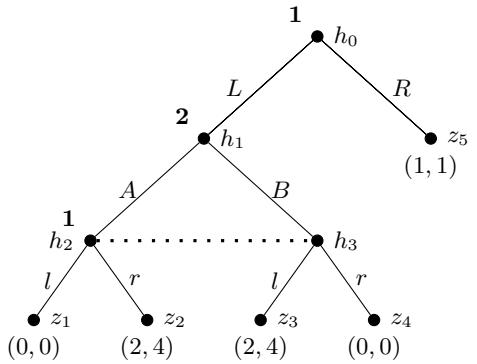
We assume the entire description of the game is public knowledge, available to all players before the game starts. Because the game states form a tree, it holds that whenever $\sigma(h_1, a_1) = \sigma(h_2, a_2)$, then $h_1 = h_2$ and $a_1 = a_2$. As we explained earlier, every node can be hence equated with its history, which is a unique sequence of actions leading from root to that node. In particular, the probability of reaching node h due to Nature[49] is thus the product of probabilities of actions taken by Nature along the path to h. When convenient, we refer to this product as $C(h)$.

[49] Here we assume that all players take actions required to reach h.

As the definition stipulates, in order for situations to be indistinguishable, the set of actions available at nodes that belong to the same information are the same[50]. Since $I_{i,j} \in I_i$ is an equivalence class, we use $\chi(I_{i,j})$ to denote the set of actions available to player i at any node in the information set $I_{i,j}$. We further assume the actions are unique to the information sets and refer to the inverse element function that identifies an information set for action a as $I(a)$ and node h as $I(h)$.

[50] Otherwise, the player could easily discriminate between the nodes on the basis of available actions.

Example 3.2. *Consider the imperfect information extensive form game depicted on the left in Figure 3.2. The game contains a single non-trivial information set consisting of states connected by the dotted line. This information set belongs to player 1, along*



$$
\begin{aligned}
\mathrm{N} &= \{1, 2\} \\
\mathrm{H} &= \{h_0, h_1, h_2, h_3, z_1, z_2, z_3, z_4, z_5\} \\
\mathrm{Z} &= \{z_1, z_2, z_3, z_4, z_5\} \\
\mathrm{A} &= \{A, B, L, R, l, r\} \\
\mathrm{C} &= \emptyset \\
\chi &= \{h_0 \to \{L, R\}, h_1 \to \{A, B\}, \{h_2, h_3\} \to \{l, r\}\} \\
\rho &= \{\{r, h_2, h_3\} \to 1, h_1 \to 2\} \\
\sigma &= \{(h_0, L) \to h_1, (h_0, R) \to z_5, (h_1, A) \to h_2, (h_1, B) \to h_3, \\
&\quad (h_2, l) \to z_1, (h_2, r) \to z_2, (h_3, l) \to z_3, (h_3, r) \to z_4\} \\
\mathrm{I} &= \{(\{h_0\}, \{h_2, h_3\}), \quad (\{h_1\})\} \\
\mathrm{u} &= (\{z_1 \to 0, z_2 \to 2, z_3 \to 2, z_4 \to 0, z_5 \to 1\}, \\
&\quad \{z_1 \to 0, z_2 \to 4, z_3 \to 4, z_4 \to 0, z_5 \to 1\})
\end{aligned}
$$

Figure 3.2: (Left) An example of a game tree of an extensive form game. Each internal node is labeled by a player who acts in this node, while under every terminal node is a tuple of utilities for the first player and the second player, respectively. Every edge is labeled by an action performed on a way from the node above to the node below. The nodes which belong to the same information set are connected by a dashed line. The direction of the edges is omitted, but the tree is assumed to be traversed from top to bottom. (Right) The game's formal EFG representation.

with a trivial set including only the root node. Note that the two bottom nodes belonging to the same information set have the same set of legal actions. The information set reflects the fact that player 1 can not observe whether Player 2 took action A or B in their single state in the game. The complete formalization of this extensive form game is on the right in the same figure.

Similarly, as in the normal form, each player may play deterministic *pure strategies*. To define behavior in the entire game, pure strategies have to assign one action to take in every information set in which the player acts. The set of pure strategies of player i is hence a Cartesian product $\Pi_i = \Pi_{I_{i,j} \in I_i} \chi(I_{i,j})$ and every pure strategy is an element $\pi_i \in \Pi_i$. We further assume the actions prescribed by a strategy π_i can be enumerated as for all $a_i \in \pi_i$.

Example 3.3. *Consider the game in Figure 3.2 again. The pure strategies of player 1 are the combinations of actions in their information sets, namely*

$$\Pi_1 = \{(L, l), (L, r), (R, l), (R, r)\}.$$

Because the second player may act in just one state, their pure strategies are trivially

$$\Pi_2 = \{(A), (B)\}.$$

The fundamental assumption of game theory is that there is no loss of generality in assuming that every player chooses their strategy before the game starts.[51] This follows from the fact that strategies specify players' behavior in every possible situation they might find themselves in, and there is hence no advantage in altering the strategies during the course of play. The notion of sequentiality is strategically irrelevant, and we may reason about extensive form games in the same manner as about

[51] J. von Neumann. "Zur theorie der gesellschaftsspiele". *Mathematische Annalen* 100:1, 1928, pp. 295–320.

		Player 1			
		L, l	L, r	R, l	R, r
Player 2	A	0,0	2,4	1,1	1,1
	B	2,4	0,0	1,1	1,1

Figure 3.3: An equivalent normal form representation of the extensive form game from Figure 3.2. The actions correspond to pure strategies delineated in Example 3.3. The figure follows the standard denotation of a normal form game.

games in normal form. The downside is the associated blowup, exponential in the number of information sets, caused by enumerating the pure strategies in sequential games. An example of the equivalent normal form representation of the game from Figure 3.2 is depicted in Figure 3.3.

The emergence of stochastic behavior is a natural result of the players' attempt to cope with the uncertainty associated with imperfect information. In extensive form games, there exist two ways how to meaningfully introduce randomness into strategizing. The first option is to behave analogically to normal form games, i.e., to specify a probability distribution over pure strategies and later sample this distribution before the game starts. As in normal form games, we call these strategies mixed and denote them by Δ_i for each player i. This approach is justified by the equivalence of extensive and normal form games described in the previous paragraph. Alternatively – and more in the manner of sequential games – the players may randomize independently over actions in each information set. Such strategies are referred to as *behavioral* and denoted as $\beta_i \in B_i$. We write $\beta_i(a_i)$ as the probability of player i taking action a_i in their information set $I \in I_i,\ a_i \in \chi(I)$[52].

The expressive power of behavioral and mixed strategies is incomparable; in certain games, outcomes achievable via mixed strategies may not be reachable with behavioral strategies, and vice versa. Pivotal turns out to be the role of the players' memory. In game-theoretic literature, memory is formalized via *recall*.

[52] For any set of actions $\overline{A}_i$ of player i we write $\prod_{a_i \in \overline{A}_i} \beta_i(a_i)$ as $\beta_i(\overline{A}_i)$.

Definition 3.3. *Player i has perfect recall in an imperfect information extensive form game G if for any two nodes h, h' that belong to the same information set, for any path $h_0, a_0, h_1, a_1, h_2, \ldots, h_m, a_m, h$ from the root of the game to h[53] and for any path $h_0, a'_0, h'_1, a'_1, h'_2, \ldots, h'_{m'}, a'_{m'}, h'$ from the root to h' it must hold that:*

1. $m = m'$;
2. *for all* $0 \leq j \leq m$, *if* $\rho(h_j) = i$ *then* $I(h_j) = I(h'_j)$; *and*
3. *for all* $0 \leq j \leq m$, *if* $\rho(h_j) = i$ *then* $a_j = a'_j$.

G is a game of perfect recall if every player has perfect recall in it.

[53] We assume that h_j's are decision nodes and a_j's are the actions.

In this thesis we consider only games where all players possess perfect recall. In such games, the expressive power of mixed and behavioral strategies coincide.[54]

[54] H. W. Kuhn. "Extensive games and the problem of information". *Annals of Mathematics Studies* 28, 1953.

Theorem 3.1. *In a game of perfect recall, any mixed strategy of a given player can be replaced by an equivalent behavioral strategy, and any behavioral strategy can be replaced by an equivalent mixed strategy. Here two strategies are equivalent in the sense that they induce the same probabilities on outcomes, for any fixed strategy profile (mixed or behavioral) of the remaining players.*

Sequence form

As we explained, the number of pure strategies in an extensive form game is exponential, which makes representing strategies in their mixed form problematic. The behavioral strategies are substantially more efficient as they require storing only one distribution over actions per each information set. The catch is that computing reaching probabilities of leaves now entails multiplications of action probabilities in the information sets, which complicates calculations of expected utilities[55]. If there was a way to represent the reaching probabilities efficiently by behavioral strategies, we would avoid the exponential blowup associated with mixed strategies. It can be done by operating directly on the structure of the extensive form game, and this equivalent representation is called the sequence form.

[55] In particular, computing optimal best responses with behavioral strategies may no longer be done via linear programming.

Definition 3.4. *A sequence form representation of an extensive form game G is an augmentation*[56] *of G by a quintuple $(\Sigma, seq, inf, Ext, g)$, where*

- $\Sigma = \Sigma_1 \times \cdots \times \Sigma_n$*, where Σ_i is a set of ordered lists of actions of player i that lie on the path from the root state h_0 to any state $h \in H$;*
- $seq = (seq_1, \ldots, seq_n)$*, where $seq_i : H \to \Sigma_i$ is a function that identifies a sequence of player i's actions leading to a node;*
- $inf = (inf_1, \ldots, inf_n)$*, where $inf_i : \Sigma_i \to I_i$ is a function that identifies the information set in which player i took their last actions in their sequence;*
- $Ext : \Sigma_i \to 2^{\Sigma}_{i}$ *is a function that identifies extensions of sequences:*

$$Ext(seq_i(h)) = \{seq_i(h)a_j | a_j \in \chi(h)\} \qquad h \in H; and$$

- $g = (g_1, \ldots, g_n)$ *where $g_i : \Sigma \to \mathbb{R}$ is a sequence utility function of player i:*

$$g_i(\sigma_1, \ldots, \sigma_n) = \sum_{z \in Z | \forall i \in N : seq_i(z) = \sigma_i} u_i(z) C(z).$$

[56] We may also construct an entirely separate representation, without retaining the original G, by imposing further linear restrictions on the sequence probabilities from Definition 3.5 as a part of the representation. For details see the book of Shoham and Leyton-Brown.

A sequence may be empty, and in that case we write it as $\emptyset$. For every player, the function $inf_i(\emptyset)$ returns the information set of the root node. Similarly to function χ, we occasionally use $seq_i(I_{i,j})$ to denote a sequence of player i leading to the information set $I_{i,j}$. Also note that if no leaf is reachable with a given tuple of sequences, a value of g_i is 0 for every player i.

Example 3.4. *Consider the extensive form game depicted in Figure 3.2 and its sequence form representation in Figure 3.4. Let us denote the lower non-trivial infor-*

Player 1

Player 2	$\emptyset_1$	L	R	Ll	Lr
$\emptyset_2$	0,0	0,0	1,1	0,0	0,0
A	0,0	0,0	0,0	0,0	2,4
B	0,0	0,0	0,0	2,4	0,0

$$
\begin{aligned}
\Sigma &= \{\emptyset_1, L, R, Ll, Lr\} \times \{\emptyset_2, A, B\} \\
seq &= (\{h_0 \to \emptyset_1, h_1 \to L, h_2 \to L, h_3 \to L, z_1 \to Ll, \\
&\quad z_2 \to Lr, z_3 \to Ll, z_4 \to Lr, z_5 \to R\}, \\
&\quad \{h_0 \to \emptyset_2, h_1 \to \emptyset_2, h_2 \to A, h_3 \to B, z_1 \to A, \\
&\quad z_2 \to A, z_3 \to B, z_4 \to B, z_5 \to \emptyset\}) \\
inf &= (\{\{\emptyset_1, L, R\} \to \{h_0\}, \{Ll, Lr\} \to \{h_2, h_3\}\} \\
&\quad \{\emptyset_2 \to \{h_0\}, \{A, B\} \to \{h_1\}\}) \\
Ext &= \{\emptyset_1 \to \{L, R\}, L \to \{Ll, Lr\}, \emptyset_2 \to \{A, B\}\} \\
g &= (\{(R, \emptyset_2) \to 1, (Ll, B) \to 2, (Lr, A) \to 2\}, \\
&\quad \{(R, \emptyset_2) \to 1, (Ll, B) \to 4, (Lr, A) \to 4\})
\end{aligned}
$$

Figure 3.4: (Left) A payoff matrix of the equivalent sequence form representation of the extensive form game from Figure 3.2. Each column of the depicted matrix is labeled by a sequence of the first player, while every row is labeled by a sequence of the second player. The tuples in the matrix denote the sequence utilities of the first player and the second player, respectively. (Right) The game's formal sequence form representation.

mation set $\{h_2, h_3\}$. For this information set, $seq_1(\{h_2, h_3\})$ is equal to L. Consequently, $inf_1(Ll) = inf_1(Lr) = \{h_2, h_3\}$. The extensions of L form a set Ext(L) $= \{Ll, Lr\}$. Note that in contrast to the normal form representation from Figure 3.3, the utility from each terminal state occurs exactly once in the sequence form table. We may hence represent g sparsely by storing only three[57] *value pairs, as shown on the left in Figure 3.4, instead of eight as in the normal form representation. The entire sequence form representation is on the right in the same figure.*

[57] Since two leaves carry utility $(0, 0)$, they may be omitted.

Using sequences, players may represent reaching probabilities efficiently via linear network flow constraints of the so-called *realization plans*. This strategy representation is equivalent to behavioral strategies; one may recover a behavioral strategy from a realization plan and vice versa. By Theorem 3.1, realization plans have the same expressive power as mixed strategies.

Definition 3.5. *A realization plan of player i is a function $r_i : \Sigma_i \to [0, 1]$ satisfying the following constraints.*

$$
\begin{aligned}
r_i(\emptyset) &= 1 \\
\sum_{\sigma_i' \in Ext_i(I)} r_i(\sigma_i') &= r_i(seq_i(I)) \qquad \forall I \in I_i \\
r_i(\sigma_i) &\geq 0 \qquad \forall \sigma_i \in \Sigma_i
\end{aligned}
$$

❖

For the sake of exposition, we frequently denote the opponents of player i as $-i$. We use this notation for partial strategy profiles, i.e.; $\delta_{-i} = \delta \backslash \delta_i$ or $\pi_{-i} = \pi \backslash \pi_i$, but also for sets akin to sequence profiles as Σ_{-i}, or even functions like seq_{-i} or inf_{-i}. Given a partial action profile $a_{-i} \in A_{-i}$, we further denote the probability of playing a_{-i} given δ_{-i} as $\delta_{-i}(a_{-i}) = \prod_{j \in N, i \neq j} \delta_j(a_j)$, and for partial sequence profiles $\sigma_{-i} \in \Sigma_{-i}$ and behavioral strategies as $\beta_{-i}(\sigma_{-i}) = \prod_{j \in N, i \neq j} \beta_j(\sigma_j)$.

3.2 Rational solution concepts

Solutions of strategic interactions are called *equilibria* or solution concepts. They describe profiles of strategies no player has an intention to deviate from. Deviating may be driven by different motives under different circumstances, depending on players' rationality or strategic abilities. For example, in Nash equilibrium a player will deviate simply upon a chance to increase their utility. This stands in contrast to Stackelberg equilibrium where the leader will consider changing their strategy only when profitable under the condition the followers react accordingly. In most solution concepts, central to the players' reasoning is their utility expectation[58], defined for each player i under a mixed strategy profile $\delta \in \Delta$ directly as

[58] We will give the definitions for normal form games, but all concepts translate to extensive form games naturally as one would expect.

$$u_i(\delta) = \sum_{a \in A_1 \times \ldots \times A_n} u_i(a) \prod_{j=1}^{n} \delta_j(a_j).$$

A rational player then aims to maximize their utility.

Definition 3.6. *Given a normal-form game $G = (N, A, u)$, a strategy δ_i^* of player i is an ϵ-best response to strategies δ_{-i} of player i's opponents if and only if*

$$u_i(\delta_i^*, \delta_{-i}) \geq u_i(\delta_i, \delta_{-i}) - \epsilon \qquad \forall \delta_i \in \Delta_i.$$

We denote the set of all δ_i^* with $\epsilon = 0$ as $BR(\delta_{-i})$.

3.2.1 Nash equilibrium

The standard solution concept of Nash equilibrium assumes players who observe the behavior of their opponents and subsequently choose strategies that optimize their immediate expected utility. In the equilibrial profile, the players gain nothing by unilaterally changing their strategies.

Definition 3.7. *Given a game $G = (N, A, u)$, a strategy profile $\delta^{NE} = (\delta_1, \ldots, \delta_n) \in \Delta$ is a Nash equilibrium if and only if*

$$\delta_i \in BR(\delta_{-i}^{NE}) \qquad \forall i \in N.$$

Example 3.5. *Consider again the extensive form game depicted in Figure 3.2. This game contains four Nash equilibria:*

1. *Ll and B; and*
2. *Lr and A are straightforward. In addition,*
3. *R and $\{A = .5, B = 0.5\}$; and*
4. *$\{Ll = .5, Lr = .5\}$ and $\{A = .5, B = .5\}$ are mixed.*

The seminal result of Nash shows that every game contains a Nash equilibrium.[59]

[59] J. F. Nash. "Non-cooperative games". *Annals of Mathematics*. Second 54:2, 1951.

Theorem 3.2. *Every game with a finite number of players and a finite number action profiles has at least one Nash equilibrium in mixed strategies.*

The fact that Nash equilibrium exists in every finite game distinguishes the problem of finding Nash equilibrium from known NP-complete problems[60]. This hints that Nash equilibrium belongs to another, less familiar complexity class consisting of certain problems for which the solution is guaranteed to exist. This class is called *PPAD*, which stands for "polynomial parity argument, directed version".[61]

Definition 3.8. *A problem A is in PPAD if there is a polynomial time reduction from A to the End-of-Line problem, where End-Of-The-Line is the following search problem: The input consists of a directed graph in which each node has in-degree and out-degree at most 1. The graph is given by a polynomial-time computable function $f(x)$ that returns the predecessor and successor of x. A node v with a successor but no predecessor is also given. Find a node $t \neq v$ that has no successor or no predecessor.*

Computing Nash equilibrium is a complete problem for this class.[62]

Theorem 3.3. *The problem of finding a sample Nash equilibrium of a general-sum game with two or more players up to ϵ-best responses is PPAD-complete.*

3.2.2 Stackelberg equilibrium

Commitment power gives a player who possesses a significant advantage over their opponents by being able to steer the course of play into the direction the committing player favors.[63] Having such power is rare, though, and only major governmental agencies or dominant market leaders are able to acquire and implement it[64]. In the Stackelberg terminology, they are called *leaders* while their opponents are referred to as *followers*. In this thesis, we restrict ourselves to the simplest case with one leader and one follower, and we will write $N = \{l, f\}$ in these scenarios further on.

Leader's rationalization in Stackelberg games is more cumbersome than in Nash scenarios. The reason is they need to take into account the follower's responses to their behavior. The leader will hence adopt a strategy maximizing their utility under the condition the follower best responds to it. A common assumption in the literature is that the follower will act in the leader's favor when indifferent between multiple options. One potential basis of this premise is that the follower's utility is always greater or equal to their Nash equilibrium utility.[65] This form of Stackelberg equilibrium with a compliant follower is called the *Strong Stackelberg equilibrium*.

Definition 3.9. *Given a normal form game $G = (N, A, u)$, the strategy $\delta_l^{SSE} \in \Delta_l$ is the leader's strategy in Strong Stackelberg equilibrium if and only if*

$$\delta_l^{SSE} = \arg\max_{\delta_l \in \Delta_l} u_l(\delta_l, \pi_f^{SSE}),$$

where $\pi_f^{SSE} \in \Pi_f$ is the follower's strategy breaking ties in leader's favor, i.e.,

$$\pi_f^{SSE} = \arg\max_{\pi_f \in BR(\delta_l)} u_l(\delta_l, \pi_f).$$

[60] For the reason these decision problems may not be solvable.

[61] C. H. Papadimitriou. "On the complexity of the parity argument and other inefficient proofs of existence". *Journal of Computer and System Sciences* 48:3, 1994, pp. 498–532.

[62] C. Daskalakis, P. W. Goldberg, and C. H. Papadimitriou. "The complexity of computing a Nash equilibrium". *SIAM Journal on Computing* 39:1, 2009, pp. 195–259.

[63] H. Stackelberg. *Market structure and equilibrium*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 1–10.

[64] Real-world examples of commitment include market leaders being able to set prices for items or services, or law enforcement officials setting up road sobriety checkpoints to check drivers for impairments. For more examples, see the book of M. Tambe. *Security and game theory: Algorithms, deployed systems, lessons learned*. Cambridge University Press, New York, NY, USA, 2011.

[65] B. von Stengel and S. Zamir. *Leadership with commitment to mixed strategies*. Technical report. 2004.

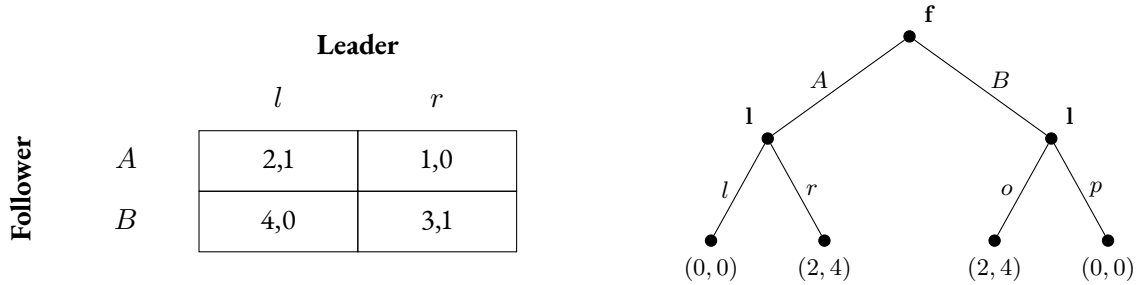


Figure 3.5: The examples of (left) a normal form game and (right) an extensive form game where the leader profits from committing to strategies that make the follower indifferent among their options. The figure follows the standard denotations of normal and extensive forms of games.

Example 3.6. *Consider the payoff matrix of the game depicted on the left in Figure 3.5. Here, the leader's utilities are listed first. In Nash equilibrium, the leader would avoid playing r because no matter how the follower reacts, the leader always obtains a higher utility by taking l. With the commitment power, the leader can take advantage of r, though. In case they commit solely to r*[66]*, the follower will respond with B, granting the leader utility 3. In equilibrium, the leader makes the follower indifferent between A and B by committing to playing both actions with equal probabilities. By virtue of the follower's compliance, the leader ends up with a utility 3.5.*

Now take a look at the extensive form game on the right in Figure 3.5. In Nash equilibrium, the leader plays l and p whilst the follower takes A, resulting in the payoffs $(1, 3)$. In contrast, committing to r in the leader's left state and randomizing uniformly over o and p will again make the follower indifferent between A and B. In this Strong Stackelberg equilibrium, the follower takes B, which will guarantee the leader the expected payoff of 2.5 – a utility strictly higher than in Nash equilibrium.

[66] In some cases, leaders may prefer deterministic commitment. For example, when observing the entire support is not feasible for the followers.

Computing Stackelberg equilibrium

We introduce mathematical programs for computing Stackelberg equilibrium we later use to derive our algorithms for boundedly rational Stackelberg equilibria from. The key insight is that the best response of the follower may be encoded via linear constraints, which enables leveraging standard optimization techniques.[67]

[67] V. Conitzer and T. Sandholm. "Computing the optimal strategy to commit to". In: *Proceedings of the 7th ACM Conference on Electronic Commerce*. New York, NY, USA, 2006, pp. 82–90.

Theorem 3.4. *Given a two-player normal form game $G = (N, A, u)$, the optimal leader's strategy $\delta_l \in \Delta_l$ in Strong Stackelberg equilibrium maximizes the leader's expected utility expressed as*

$$\max_{\delta_l \in \Delta_l, \pi_f \in \Pi_f} \sum_{\pi_l \in \Pi_l} u_l(\pi_l, \pi_f)\delta_l(\pi_l),$$

under the condition the follower's response π_f is optimal with respect to the alternatives:

$$\sum_{\pi_l \in \Pi_l} u_f(\pi_l, \pi_f)\delta_l(\pi_l) \geq \sum_{\pi_l \in \Pi_l} u_f(\pi_l, \pi'_f)\delta_l(\pi_l) \qquad \forall \pi'_f \in \Pi_f.$$

Finding Strong Stackelberg equilibrium in extensive form games is possible via a mathematical program of polynomial size should we employ the realization plans to circumvent the exponential blowup of follower's pure strategies. The downside of this approach is the best response variables are no longer continuous but rather integral, making the formulation a mixed integer linear program.[68]

[68] B. Bošanský and J. Čermák. "Sequence-form algorithm for computing Stackelberg equilibria in extensive-form games". In: *Proceedings of the 29th AAAI Conference on Artificial Intelligence*. 2015, pp. 805–811.

Theorem 3.5. *Given a two-player extensive form game $G = (N, H, Z, A, C, \chi, \rho, \sigma, I, u)$, the realization plan strategies r_l and r_f in Strong Stackelberg equilibrium maximize the leader's expected utility expressed as*

$$\max_{p,r,v,s} \sum_{z \in Z} p(z) u_1(z) C(z),$$

under the follower's best response constraints:

$$\begin{aligned}
v(inf_f(\sigma_f)) &= s_{\sigma_f} + \sum_{I' \in I_f : seq_f(I') = \sigma_f} v(I') \\
&\quad + \sum_{\sigma_l \in \Sigma_l} r_l(\sigma_l) g_f(\sigma_l, \sigma_f) && \forall \sigma_f \in \Sigma_f \\
r_i(\emptyset) &= 1 && \forall i \in N \\
r_i(seq_i(I_i)) &= \sum_{a \in A_i(I_i)} r_i(seq_i(I_i)a) && \forall i \in N,\ \forall I_i \in I_i \\
0 &\leq s_{\sigma_f} \leq (1 - r_f(\sigma_f))M && \forall \sigma_f \in \Sigma_f \\
0 &\leq p(z) \leq r_f(seq_f(z)) && \forall z \in Z \\
0 &\leq p(z) \leq r_l(seq_l(z)) && \forall z \in Z \\
1 &= \sum_{z \in Z} p(z) C(z) \\
r_f(\sigma_f) &\in \{0, 1\} && \forall \sigma_f \in \Sigma_f \\
0 &\leq r_l(\sigma_l) \leq 1 && \forall \sigma_l \in \Sigma_l.
\end{aligned}$$

The variable p in this program represents the probability distribution over leaves induced by r_l and r_f. Variable v then propagates follower's utility through the tree.

❖

Solving mixed integer linear programs akin to that of Theorem 3.5 is difficult yet in line with the complexity of computing Stackelberg equilibria in general.[69]

[69] J. Letchford and V. Conitzer. "Computing optimal strategies to commit to in extensive-form games". In: *Proceedings of the 11th ACM Conference on Electronic Commerce*. 2010, pp. 83–92.

Theorem 3.6. *The problem of finding a Strong Stackelberg equilibrium of games beyond two player normal form games or two player perfect information extensive form games without Nature is NP-hard.*

3.2.3 Correlated equilibrium

Correlated equilibrium describes a situation when the players are able to coordinate their strategies based on external signals. The standard construction of correlated equilibrium assumes a lottery mechanism (a signaling scheme) $\lambda \in \Lambda$ that gives rise to an equilibrium (Nash, refinements thereof, etc.) in the extended signaling game.

Definition 3.10. *An extended signaling game is a tuple $G = (N, A, S, u)$, where*

- *(N, A, u) is a normal form game; and*
- *$S = S_1 \times \cdots \times S_n$, where S_i is a set of signals for player i.*

In an extended game, we denote by Λ the set of distributions over the Cartesian product of individual signal sets S. The extended game can be regarded as a sequential extension of the underlying game in which a correlation device first samples a signal profile according to the signaling scheme λ; the players then learn about their respective signals and play strategies conditioned on the signal. It is imperative that λ is public knowledge so that players are able to reason about their strategies.

Deterministic strategies in the extended game define an action to play for each signal that could be received. In other words, each $\pi_i \in \Pi_i$ in G is a set $\{(a_i, s_i), \forall s_i \in S_i\}$. For this purpose, for a given profile $\pi \in \Pi$, we define $\times\pi$ as a Cartesian product of individual $\pi_i \in \pi$. Note that the elements of $\times\pi$ are tuples of one pair (a_i, s_i) per player. Behavioral[70] strategies B in G describe stochastic strategies in the extended game, where the player's behavior is conditioned upon receiving a signal. B is a Cartesian product of B_i, where B_i is a set of conditional probability mass functions. In the standard definition of correlated equilibrium, the players attain a Nash equilibrium in the extended game[71].

[70] Due to Theorem 3.1, it does not matter whether we use behavioral or mixed strategies.

[71] Note that the standard construction of correlated equilibrium closely resembles Bayesian equilibrium. The difference lies in the definition of utility, which, contrary to the Bayesian equilibrium, is independent of λ in correlated equilibrium.

Definition 3.11. *Given an extended game G, the tuple $(\lambda, (\beta_i)_{i \in N})$, $\lambda \in \Lambda, \beta_i \in B_i$ is a correlated equilibrium if and only if*

$$\beta_i \in BR(\lambda, \beta_{-i}) \qquad \forall i \in N,$$

where the best response $BR(\lambda, \beta_{-i})$ can be expressed via constraints

$$\sum_{a \in A} \sum_{s \in S} \lambda(s)\beta(a|s)u_i(a) \geq \sum_{a_{-i} \in A_{-i}} \sum_{s \in S} \lambda(s)\beta_{-i}(a_{-i}|s)u_i(m(s_i), a_{-i})$$

for all players $i \in N$ and functions $m : S_i \to A_i$.

Example 3.7. *Consider a variant of the Battle of Sexes game shown on the left in Figure 3.6. In this game, the players aim to coordinate their choices but exhibit conflicting preferences over profiles Ao and Br. Assume that each of the players may receive one of two possible signals. The tetrahedron on the right in the figure depicts the simplex of probability distributions on pure profiles in the game, where the corners correspond to trivial distributions. The inner polytope with four vertices is the set of distributions induced by correlated equilibria of the game. Each vertex of the polytope is one Nash equilibrium in the game.*

Player 2 \ Player 1	A	B
o	2,11	0,1
r	0,0	10,2

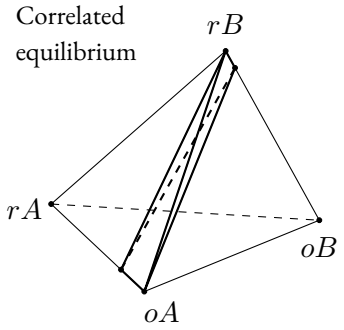


Figure 3.6: Equilibria in a variant of the Battle of Sexes game. (Left) The payoff matrix of the game. The figure follows a standard denotation of normal form games. (Right) The simplex of correlated equilibria.

Computing correlated equilibrium

The probability distribution over pure strategy profiles induced by λ and β we depict in Figure 3.6 is called a *correlated equilibrium distribution*.[72] The Revelation principle in mechanism design posits that there is a close connection between correlated equilibrium distributions and certain signaling schemes.[73] In case the players attain a Nash equilibrium in the extended game, every correlated equilibrium distribution has a *canonical representation*. In a canonical form, the signals are interpreted as actions recommended to the players to play (i.e., $S_i = A_i$), and the behavioral strategies are projections $S \to S_i$ over $S_i = A_i$. As a consequence, λ and β can be merged into one distribution, resulting in a linear formulation called a *direct correlated equilibrium*. Because of its equivalence to correlated equilibrium constructed the standard way and easy characterization, this form of correlated equilibrium is most commonly studied in the literature.

[72] R. J. Aumann. "Correlated equilibrium as an expression of Bayesian rationality". *Econometrica: Journal of the Econometric Society*, 1987, pp. 1–18.

[73] F. Forges. "Correlated equilibria and communication in games". *Complex Social and Behavioral Systems: Game Theory and Agent-Based Models*, 2020, pp. 107–118.

Theorem 3.7. The distribution λ on Π is a correlation device of a correlated equilibrium if and only if for all i, every $\pi_i \in \Pi_i$ with $\lambda(\pi_i) > 0$

$$\sum_{\pi_{-i} \in \Pi_{-i}} u_i(\pi_i, \pi_{-i}) \lambda(\pi_{-i}|\pi_i) \geq \sum_{\pi_{-i} \in \Pi_{-i}} u_i(\pi_i', \pi_{-i}) \lambda(\pi_{-i}|\pi_i) \qquad \forall \pi_i' \in \Pi_i.$$

This linear formulation enables us to more easily characterize the computational complexity and topology of correlated equilibria.

Theorem 3.8. *The space of correlated equilibria is convex and non-empty and contains all Nash equilibria of the underlying game. The problem of finding a sample correlated equilibrium is polynomial.*

However, the choice of the equilibrium in the extended game is a decisive factor in the equivalence. For example, assume the players attain a trembling hand perfect equilibrium described in Section 2.1. In that case, the direct and standard formulations of correlated equilibria do not coincide.[74] As bounded rationality has a similar

[74] A. Dhillon and J. F. Mertens. "Perfect correlated equilibria". *Journal of Economic Theory* 68:2, 1996, pp. 279–302.

effect, we do not introduce formal methods for computing correlated equilibria in extensive form games as they rely entirely on canonical representation.

3.3 Boundedly rational solution concepts

All the equilibria we considered so far assumed perfectly rational strategizing of all players where each player is able to correctly find and execute the optimal utility-maximizing action. Relaxing this assumption leads to solutions to situations that are boundedly, rather than perfectly, rational. Still, even in boundedly rational situations, the players' reasoning keeps being influenced by their strategizing abilities. We provide definitions of two boundedly rational counterparts of Nash equilibrium: quantal response equilibrium and quantal Nash equilibrium. The later chapters are then concerned with the definitions, properties, and methods of computing quantal Stackelberg equilibrium and quantal correlated equilibrium.

As we explained in Chapter 2, one potential source of the emergence of bounded rationality is a subjective perception of utility. We model it via an evaluation function $e_i : \Delta_{-i} \times A_i \to \mathbb{R}$ that assigns values to player i's action against their opponents' strategies. Typically, in normal form games, e_i may have a form of a strictly increasing function applied to the rational expected utility of player i. In extensive form games, the evaluation function may be more cumbersome as it needs to take into account player i's expectation about their own behavior later in the game.

Central to this thesis is then a "statistical version" of the best response called the quantal response, which takes into account the inevitable error-proneness of humans and allows the players to make systematic errors.[75]

[75] J. K. Goeree, C. A. Holt, and T. R. Palfrey. *Quantal Response Equilibrium*. Princeton University Press, 2016.

Definition 3.12. *Given a normal form game $G = (N, A, u)$ and a subjective utility function e_i of player i, their mixed strategy $\delta_i \in \Delta_i$ is a quantal response[76] to their opponents' strategies $\delta_{-i} \in \Delta_{-i}$ if*

$$e(\delta_{-i}, a_i^k) \leq e(\delta_{-i}, a_i^l) \Rightarrow \delta_i(a_i^k) \leq \delta_i(a_i^l) \qquad \forall a_i^k, a_i^l \in A_i.$$

We write such δ_i as values of a function $QR_i : \Delta_{-i} \to \Delta_i$.

[76] Our definition here requires only monotonicity and is hence slightly more general than regular quantal responses, focusing solely on the better-response aspects of decision making.

❖

A quantal function $QR_i : \Delta_{-i} \to \Delta_i$ is a generalized Luce model if there exists a strictly positive, increasing real valued function $q_i : \mathbb{R} \to \mathbb{R}^+$ such that

$$QR_i(\delta_{-i})(a_i) = \frac{q_i(e(\delta_{-i}, a_i))}{\sum_{a_i' \in A_i} q_i(e(\delta_{-i}, a_i'))}.$$

Because q_i is a strictly positive and increasing function, the corresponding QR_i is a valid quantal response function that implements a mixed strategy of player i. We call such functions q_i *generators* of generalized Luce models and frequently write

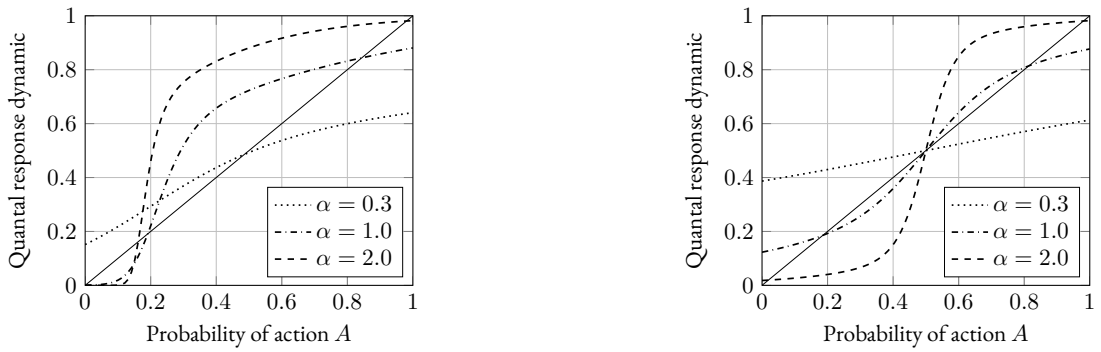


Figure 3.7: The examples of quantal response dynamics in (left) a normal form game from Figure 3.6 and (right) an extensive form game from Figure 3.2. The generators of Luce models have a form $q(x) = exp(\alpha x)$ for different values of α and are shared among the players. The evaluation functions are expected utility functions. The equilibria are found where the dynamics cross the quadrants' axes.

them together with the subjective expected utility function as $[e \circ q]$. All quantal response functions mentioned in this thesis are generalized Luce models unless explicitly stated otherwise.

3.3.1 Quantal response equilibrium

When quantal response functions of all players are continuous[77], by a direct application of the Brouwer's theorem, we know that the quantal response dynamic has a fixed point, called *quantal response equilibrium*. It describes a situation when players observe and quantal respond to their opponents' behavior.

[77] In our case this happens when the evaluation functions and the generators of Luce models are continuous.

Definition 3.13. *Given a normal form game $G = (N, A, u)$ and a profile of quantal response functions $QR_1, \ldots, QR_n$, a strategy profile $\delta^{QRE} = (\delta_1, ..., \delta_n) \in \Delta$ is a quantal response equilibrium if and only if*

$$\delta_i = QR_i(\delta_{-i}^{QRE}) \qquad \forall i \in N.$$

Example 3.8. *On the left in Figure 3.7, we depict three quantal response dynamics in the Battle of Sexes game from Figure 3.6. On the x-axis is the probability p of playing action A by player 1. The y-axis shows the probability of the same action in the quantal response of player 1 to the quantal response of player 2 to mixed strategy $\{A = p, B = 1 - p\}$. The equilibrium is a fixed point of the dynamics attained where the curves cross the quadrant's axis. We consider three logit quantal generators $q(x) = exp(\alpha x), \alpha \in \{0.2, 2.0, 5.0\}$, shared among the players, and the expected utilities as evaluation functions. As rationality increases with α tending to infinity, and the quantal responses approach best responses, the dynamics converge to the pure*[78] *Nash equilibria of the interaction.*

[78] In generalized Luce models, a ratio of two choice probabilities depends only on the expected utilities of those choices – a property known as *independence of irrelevant alternatives*. Choices with the same expected payoffs are hence always played with equal probabilities.

The two most natural ways of introducing quantal behavior in extensive form games are either over the set of pure strategies in the entire game or in each informa-

tion set separately. Since the former is equivalent to quantal response equilibrium in a normal form representation of the game, we give a definition of the latter.

Definition 3.14. *Given an extensive form game $G = (N, H, Z, A, C, \chi, \rho, \sigma, I, u)$, a sequential evaluation function $e_i : B_{-i} \times A_i \to \mathbb{R}$, and a behavioral profile $\beta = (\beta_i, \beta_{-i})$, a behavioral strategy β_i of player i is a quantal response to strategies β_{-i} in G if for any information set $I \in I_i$ the restriction of β_i to I is a valid quantal response in I with respect to utilities of actions $a \in \chi(I)$ evaluated as $e_i(\beta_{-i}, a)$. Again, we write such β_i as values of a function $QR_i : B_{-i} \to B_i$*[79].

❖

Given a profile of quantal response functions $QR_1, \ldots, QR_n$, a strategy profile $\beta^{QRE} = (\beta_1, ..., \beta_n) \in B$ is a quantal response equilibrium if and only if

$$\beta_i = QR_i(\beta_{-i}^{QRE}) \qquad \forall i \in N.$$

Example 3.9. *Figure 3.7 on the right shows three quantal response dynamics in the extensive form game from Figure 3.2. The x-axis depicts the probability p of player 2 playing action A in their sole information set. The y-axis shares the meaning with the graph on the left, i.e., it shows the probability of the same action A when player 2 quantal responds to the quantal response of player 1 to mixed strategy $\{A = p, B = 1 - p\}$. We consider the same three logit quantal generators as in Example 3.8 and assume the players evaluate their actions' values as expected utilities*[80]*. The game has a unique Nash equilibrium in mixed strategies in the lower subgame consisting of uniformly random choices. As α increases, the quantal response equilibria tend to the other two pure Nash equilibria – the outcomes $(2, 4)$. Note that the equilibrial behavior of player 2 is independent of the player 1's response in the root.*

The intuition from the examples that for certain sequences of generators the corresponding quantal response equilibria tend to Nash equilibria is indeed correct.

Theorem 3.9. *Let $\{\alpha_t\}_{t=1}^{\infty}$ be a sequence such that $\lim_{t\to\infty} \alpha_t = \infty$, and let β_t^{QRE} denote the quantal response equilibrium with a generator $q(x) = exp(\alpha_t x)$. Then any accumulation point of $\{\beta_t^{QRE}\}_{t=1}^{\infty}$ is a Nash equilibrium.*

This correspondence between α's and quantal response equilibria can be used to trace the equilibria from $\alpha = 0$ to the so-called *limiting logit equilibrium* via a homotopy method.[81] In other cases are quantal response equilibria typically difficult to compute beyond the straightforward class of two-player zero sum games because of their nonconvex formulation.

3.3.2 Quantal Nash equilibrium

Players engaging in interaction described by the quantal response equilibrium are assumed to be all boundedly rational and behaving according to the quantal response

[79] Typically, a player's evaluation of an action's value would relate to the action's expected utility. As we mentioned earlier, such utility depends not only on the opponents' strategies, but also on the player's own (expected) behavior in their lower information sets, e.g., whether they believe they act quantally or entirely rationally. We will discuss the possibilities and implications of this in more detail later in Chapter 5.

[80] Because the equilibrium does not depend on player 1's behavior in their upper information set, we do not face the perception issue described earlier.

[81] T. L. Turocy. "A dynamic homotopy interpretation of the logistic quantal response equilibrium correspondence". *Games and Economic Behavior* 51:2, 2005, pp. 243–263.

model. Yet the contemporary world is full of situations where humans act in the presence of computers and algorithms which are capable of performing entirely rationally. We call the arising intermediate solution the quantal Nash equilibrium[82].

[82] We remark the rest of this chapter comprises original theoretical results.

Definition 3.15. *Given a normal form game G = (N, A, u), a partition of N into N_b, N_r, and a profile of quantal response functions $QR_{i_1}, \dots, QR_{i_b}$, $i_j \in N_b$, a strategy profile $\delta^{QNE} = (\delta_1, ..., \delta_n) \in \Delta$ is a quantal Nash equilibrium if and only if*

$$\delta_i = QR_i(\delta_{-i}^{QNE}) \qquad \forall i \in N_b,$$

and

$$\delta_i \in BR_i(\delta_{-i}^{QNE}) \qquad \forall i \in N_r.$$

Example 3.10. *Figure 3.8 shows on the left a normal form game with a unique mixed Nash equilibrium. In the equilibrium, player 1 takes both their actions with equal probability, making player 2 indifferent between their actions. Their equilibrial strategy is to play action o with probability ≈ 0.77. The graph on the right depicts a quantal response dynamic, where player 2 acts according to a Luce model with generator $q_2(x) = exp(x)$ and player 1 responds in accordance with one of three Luce models $q_1(x) = exp(\alpha x), \alpha \in \{0.5, 1.0, 5.0\}$. Again, we assume the player evaluates their actions' values via expected utilities. As α approaches infinity, the corresponding quantal response equilibria tend to the quantal Nash strategy of player 1, where they play A with probability ≈ 0.4.*

Player 1's utility in the quantal Nash equilibrium is ≈ 2.31. Since their Nash strategy induces equal expected utilities on both their opponent's actions, every quantal response will prescribe a uniformly random strategy. Playing Nash strategy against any quantal opponent will hence result in utility 3.25 for player 1.

In the example above, player 1 needs to accommodate for player 2's bounded rationality to reach equilibrium. In case they do not have the commitment power, they may lose a substantial fraction of their utility simply on account of their opponent's flawed reasoning.

Computing quantal Nash equilibrium

As a partially rational intermediate solution concept standing in between quantal response and Nash equilibrium, quantal Nash equilibrium inherits properties of both. The previous example showed it could be reached in a limit as quantal responses approach the best response. Likewise, for computing the equilibrium, we may employ regret minimization methods for finding Nash or correlated equilibria.[83]

[83] S. Hart and A. Mas-Colell. "A simple adaptive procedure leading to correlated equilibrium". *Econometrica* 68:5, 2000, pp. 1127–1150.

Definition 3.16. *Given a two-player normal form game $G = (N, A, u)$ and a quantal response function QR of player 2, the regret matching – quantal response (RM-QR) method is a dynamic defined inductively as follows:*

Player 2 \ Player 1	A	B
o	3,0	0,6
r	0,6	10,0

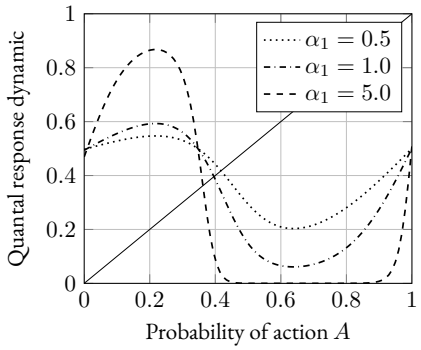


Figure 3.8: Equilibria in a game with conflicting interests. (Left) The payoff matrix of the game. The figure follows a standard denotation of normal form games. (Right) The quantal response dynamics in the game. The generators of Luce models of player 1 have a form $q_1(x) = exp(\alpha x)$ for different values of α. For player 2, $q_2(x) = exp(x)$. The evaluation functions are expected utilities. Quantal Nash equilibrium is found in the limit, as α approaches infinity. The equilibria correspond to the points where the dynamics cross the quadrants' axes.

Base case: *Let δ_1^0 be an arbitrary mixed strategy of player 1, $(r_1^0, \ldots, r_{|A_1|}^0) = 0^{|A_1|}$ be the initial regret values and $\delta_2^0 = QR(\delta_1^0)$ be the response of player 2.*

Induction: *In step $i > 0$, player 1 adopts a strategy δ_1^i defined as*

$$\delta_1^i(a_1^j) = \frac{\max(0, r_j^i)}{\sum_{a_1^k \in A_1} \max(0, r_k^i)} \qquad \forall a_1^j \in A_1,$$

where r_j^i are regret values calculated as

$$r_j^i = r_j^{i-1} + u_1(a_i^j, \delta_2^{i-1}) - u_1(\delta_1^{i-1}, \delta_2^{i-1}),$$

and player 2's response δ_2^i is

$$\delta_2^i = QR(\delta_1^{i-1}).$$

It turns out that for certain quantal response functions, *RM-QR* could be used to find a sample quantal Nash equilibrium.

Proposition 3.1. *Let $G = (N, A, u)$ be a two-player zero sum normal form game and QR be a quantal response function of player 2 that depends only on ordering of expected utilities of individual actions*[84]*. Then the RM-QR dynamic converges to a quantal Nash equilibrium.*

[84] In particular, the boundedly rational player evaluates their actions using expected utilities or a strictly increasing function of thereof.

Proof. A response function f is called a pretty-good-response if it satisfies

$$u_2(\delta_1, f(\delta_1)) \geq u_2(\delta_1, f(\delta_1')) \quad \forall \delta_1, \delta_1' \in \delta_1.$$

Consider two different $\delta_1, \delta_1' \in \Delta_1$. In case δ_1 induces the same ordering of expected utilities of player 2 as δ_1', then

$$u_2(\delta_1, QR(\delta_1)) = u_2(\delta_1, QR(\delta_1')),$$

and QR is a pretty-good-response. As an alternative, let δ_1 induce an ordering of action indices $i_1, i_2, \ldots, i_{|A_2|}$ and δ_1' a different ordering $j_1, j_2, \ldots, j_{|A_2|}$. By Definition 3.12 of a quantal response function,

$$QR(\delta_1, a_2^{i_1}) \geq QR(\delta_1, a_2^{i_2}) \geq \cdots \geq QR(\delta_1, a_2^{i_{|A_2|}})$$

and

$$QR(\delta_1', a_2^{j_1}) \geq QR(\delta_1', a_2^{j_2}) \geq \cdots \geq QR(\delta_1', a_2^{j_{|A_2|}}).$$

For each $k \in [|A_2|]$ it holds that

$$u_2(\delta_1, a_2^{i_k}) QR(\delta_1, a_2^{i_k}) \geq u_2(\delta_1, a_2^{i_k}) QR(\delta_1', a_2^{j_k})$$

and therefore

$$u_2(\delta_1, QR(\delta_1)) \geq u_2(\delta_1, QR(\delta_1')).$$

Simple QR's are hence pretty-good-responses and the dynamics of regret matching coupled with a pretty-good-response are known to converge to a strategy of player 1 exploiting the pretty-good-response the most.[85] The *RM-QR* hence converges to a quantal Nash equilibrium. □

[85] T. Davis, N. Burch, and M. Bowling. "Using response functions to measure strategy strength." In: *Proceedings of the 28th AAAI Conference on Artificial Intelligence*. 2014, pp. 630–636.

Unfortunately, computing a quantal Nash equilibrium is hard in general.

Theorem 3.10. *The problem of finding a sample quantal Nash equilibrium of a general-sum finite game with two or more players is PPAD-complete.*

We begin by proving an auxiliary lemma.

Lemma 3.1. *Let* $\mathcal{A} = \{a_1, a_2, \ldots, a_m\}, a_i \in \mathbb{R}$, $a_1 = \max(\mathcal{A}) > 0$. *Then it holds that*

$$\max(\mathcal{A}) - \text{soft}\max_\alpha(\mathcal{A}) \leq \frac{W(1/e)}{\alpha} + \frac{m-2}{\alpha e},$$

where

$$\text{soft}\max_\alpha(\mathcal{A}) = \frac{e^{\alpha a_1}}{\sum_{a \in \mathcal{A}} e^{\alpha a}},$$

and W is the Lambert W function.

Proof. We proceed by induction on the size of the set $\mathcal{A}$.

Base case: Let $m = 2$. Because $a_2 \leq a_1$, any a_2 can be written as $a_1 x, x \leq 1$. For a given α, the difference between max and soft max can be written as

$$d(x) = a_1 - \frac{a_1 e^{\alpha a_1} + a_1 x e^{\alpha a_1 x}}{e^{\alpha a_1} + e^{\alpha a_1 x}}.$$

To find a maximum of this function, we differentiate it by x, which yields

$$d'(x) = -\frac{a_1 e^{\alpha a_1 x}(e^{\alpha a_1}(\alpha a_1(x-1)+1) + e^{\alpha a_1 x})}{(e^{\alpha a_1 x} + e^{\alpha a_1})^2}.$$

For $a_1 > 0$, the function d' has a root

$$r = \frac{a_1\alpha - W(1/e) - 1}{a_1\alpha}.$$

The root is unique, because the inner function $e^{\alpha a_1}(\alpha a_1(x-1) + 1) + e^{\alpha a_1 x}$ is increasing as its derivative is positive for all $x \leq 1$. It is a maximum of d, because $d''(r) < 0$. By plugging the root into the function d, we obtain the upper bound on the distance between $\max$ and soft $\max$:

$$d(r) = \frac{W(1/e)}{\alpha},$$

which is independent on a_1, a_2.

Induction: For a given $|\mathcal{A}| = m$, assume $\max(\mathcal{A}) - \text{soft}\max_\alpha(\mathcal{A}) \leq C$. Consider a new $a_{m+1} \leq a_1$. Again, we set $a_{m+1} = a_1 x, x \leq 1$. For a given α, the difference between $\max$ and soft $\max$ can be written as

$$a_1 - \frac{\sum_{i=1}^n a_i e^{\alpha a_i} + a_1 x e^{\alpha a_1 x}}{\sum_{i=1}^n e^{\alpha a_i} + e^{\alpha a_1 x}} \leq C + \frac{(a_1 - a_1 x)e^{\alpha a_1 x}}{e^{\alpha a_1}},$$

because the exp function is strictly greater than zero. To find a maximum of the second term, we differentiate it by x:

$$\left(\frac{(a_1 - a_1 x)e^{\alpha a_1 x}}{e^{\alpha a_1}}\right)' = a_1^2\alpha(1-x)e^{a_1\alpha(x-1)} - a_1 e^{a_1\alpha(x-1)}.$$

As in the base case, for $a_1 > 0$ the derivative has a root

$$r = 1 - \frac{1}{\alpha a_1}, \qquad \frac{(a_1 - a_1 r)e^{\alpha a_1 r}}{e^{\alpha a_1}} = \frac{1}{\alpha e}.$$

The root is unique, because the derivative is positive increasing on $(-\infty, 1 - 2/a_1\alpha)$ and decreasing on $(1 - 2/a_1\alpha, 1]$, as differentiating it for the second time reveals. Therefore, we obtain the upper bound

$$\max(\mathcal{A} \cup a_{m+1}) - \text{soft}\max_\alpha(\mathcal{A} \cup a_{m+1}) \leq C + \frac{1}{\alpha e}.$$

The result follows from the induction. Note that the upper bound goes to zero as α approaches infinity. □

Now we can move to the proof of Theorem 3.10.

Proof. Let $\tilde{G}$ be a two-player normal form game with strictly positive utilities, in which one of the players has m actions to play. Computing a Nash equilibrium in $\tilde{G}$ is PPAD-complete by Theorem 3.3. We show that computing a quantal Nash equilibrium is PPAD-hard by reducing the problem of finding a Nash equilibrium in $\tilde{G}$ to a problem of computing a specific quantal Nash equilibrium in $\tilde{G}$.

We construct the reduced game as follows:

- Let the games coincide,
- let the player with m actions be the boundedly rational player 2, and
- let this player use expected utilities to evaluate their actions and act according to a logit generator q_2 of a Luce model of quantal response defined as $q_2(x) = exp(\alpha x)$ for some $\alpha \geq 0$.

Assume that there exists α^*, such that for each ϵ and each strategy δ_1 of player 1

$$u_2(\delta_1, BR(\delta_1)) - u_2(\delta_1, QR(\delta_1)) \leq \epsilon.$$

Because player 1 is fully rational, their quantal Nash strategy is a best response. By the definition of α^*, the quantal response of player 2 is an ϵ-best response. Therefore, by solving for a quantal Nash equilibrium with the generator of the second player $q_2(x) = exp(\alpha^* x)$, we find an approximate Nash equilibrium up to ϵ-best responses in $\tilde{G}$.

Each strategy δ_1 of player 1 generates a set of expected utilities for player 2. A player best responding to δ_1 achieves a maximum utility from this set whilst quantal responding according to the generator q_2 guarantees a $\mathrm{soft\,max}$ utility. Because the game we reduce from has m actions, setting $\alpha^* = \frac{W(1/e)}{\epsilon} + \frac{m-2}{\alpha e}$ as in Lemma 3.1 concludes the proof. □

❖

In summary, this chapter introduced the necessary technical background consisting of notation and solution concepts this thesis builds upon. Now we are ready to delve into the description of the fundamental contributions the thesis makes, starting with the optimal commitment against a quantal response player.

Part II

Commitment

4 Quantal Stackelberg Equilibrium in Normal Form Games

Game theoretic algorithms computing Stackelberg equilibrium have been used for improving physical security or protecting wildlife in natural parks.[86] The lesson learned from these deployments is that following the standard mathematical assumption of perfectly rational players is not optimal against human opponents. The strategies against boundedly rational opponents prove to be more effective in case the algorithms take into account the quantal response nature of human decision makers and their subjective utility functions.[87]

The game model of Stackelberg security games used by many such deployed real-world applications stems from normal form games, but it poses certain modeling limitations. The problem must be formulated in terms of allocating limited resources to a set of targets. This is often impossible, e.g., in classical games from economics like prisoner's or traveler's dilemma, location games, or algorithmic mechanism design. In order to solve real-world problems beyond security games, we study optimal behavior against a quantal response opponent in more general class of normal form games. An optimal strategy of the rational player in such a scenario is described by a leader-follower solution concept we call the quantal Stackelberg equilibrium. The leader computes and implements a fixed strategy and the follower responds to this strategy based on their quantal response function.

The natural formulation of the optimization problem describing the quantal Stackelberg equilibrium is non-concave. Moreover, we show that the number of local minima may grow linearly with the number of actions even in a zero sum games. Hence, simple gradient ascent approaches are not likely to be effective. It is difficult to further analyze the problem of computing the equilibrium directly, so we derive a Dinkelbach-type formulation of the problem and use it to identify sufficient conditions for concavity of the problem. If the conditions are satisfied, the optimal solution can be found by gradient ascent. Furthermore, we generalize the linear relaxation developed for security games to general normal form games and use it to formulate a mixed integer linear program for approximately solving the problem. The approach of security games cannot be applied directly, because it uses specific properties of such games.

In the experiments, we compare the convergence speed of gradient ascent in direct formulation to approximated Dinkelbach-type integer linear program formu-

The results in this chapter were published as J. Černý, V. Lisý, B. Bošanský, and B. An. "Dinkelbach-type algorithm for computing Quantal Stackelberg equilibrium". In: *Proceedings of the 29th International Joint Conference on Artificial Intelligence*. Ed. by C. Bessiere. International Joint Conferences on Artificial Intelligence Organization, 2020, pp. 246–253.

[86] M. Tambe. *Security and game theory: Algorithms, deployed systems, lessons learned*. Cambridge University Press, New York, NY, USA, 2011.

[87] T. H. Nguyen, A. Sinha, S. Gholami, A. Plumptre, L. Joppa, M. Tambe, M. Driciru, F. Wanyama, A. Rwetsiba, and R. Critchlow. "Capture: A new predictive anti-poaching tool for wildlife protection". In: *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems*. 2016, pp. 767–775.

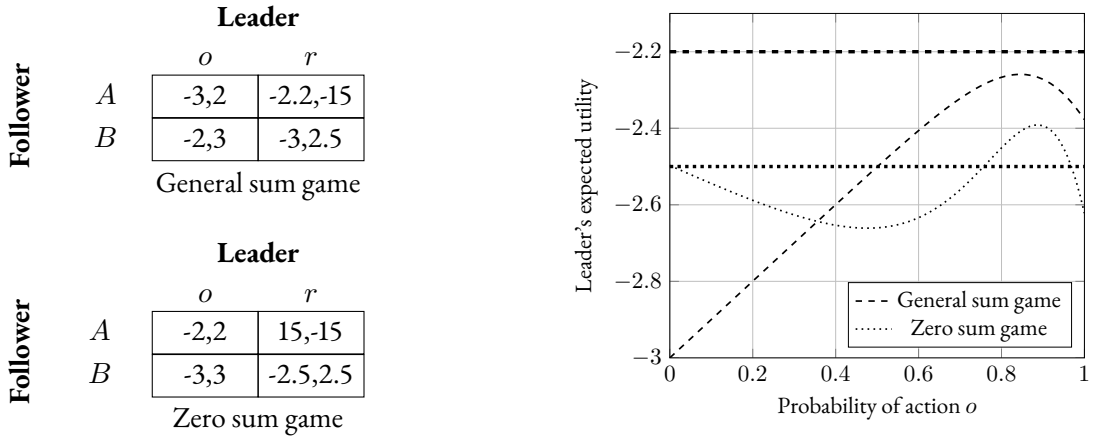


Figure 4.1: An example of a quantal Stackelberg equilibrium in a normal form game. (Left) The payoff matrices of a general-sum game and its zero-sum variant. The figure follows a standard denotation of normal form games. (Right) The criterion functions of the general-sum game's quantal Stackelberg equilibria (dashed curve) and its zero-sum variant's (dotted curve). The quantal Stackelberg equilibrium maximizes the leader's expected utility. The values of Nash equilibria are depicted in corresponding bold straight lines.

lation of quantal Stackelberg equilibrium with various quantal functions. We show that the Dinkelbach-type algorithm is up to 25.5-times faster than one restart of gradient ascent and it also provides solutions of higher quality. For the most common generators of quantal functions[88] and sufficiently large games, the gradient ascent could not find a solution of the same quality even when running 153-times longer than the Dinkelbach-type algorithm.

[88] For example, the exponential logit function.

4.1 Problem definition and properties

We begin our analysis by formally defining the equilibrium.

Definition 4.1. *Given a normal-form game $G = (N, A, u)$ and a quantal response function QR of the follower, a mixed strategy $\delta_l^{QSE} \in \Delta_l$ describes a quantal Stackelberg equilibrium if and only if*

$$\delta_l^{QSE} = \arg\max_{\delta_l \in \Delta_l} u_l(\delta_l, QR(\delta_l)). \qquad \text{(QSE-NFG)}$$

In this situation the leader is fully rational. They commit to an optimal strategy given that the follower quantal responds according to given quantal and evaluation functions[89]. Throughout this chapter, we refer to the expression in Definition 4.1 as to the criterion function of quantal Stackelberg equilibrium.

[89] In other words, the quantal Stackelberg equilibrium is an exact analogue of the standard Stackelberg equilibrium in the presence of a quantal adversary.

Example 4.1. *Consider the general sum game shown on the left in Figure 4.1. Let the follower behave according to a Luce model with generator $q(x) = exp(x/2)$ and let them use the expected utility as an evaluation function. On the right of the figure is depicted a criterion function of the game's quantal Stackelberg equilibrium (dashed)*

and a criterion function of the game's zero sum variant's quantal Stackelberg equilibrium (dotted), in which we set $u_l = -u_f$. The optimal strategy of the leader in the general sum game is to commit to playing action o with probability ≈ 0.85. To achieve the equilibrium in the zero sum variant the leader plays o with probability ≈ 0.89.

The example shows that in general, finding the quantal Stackelberg equilibrium is a non-linear non-concave problem. To understand if standard methods could be of use in computing the equilibrium, we need to analyze its properties in more detail. We begin by studying the relation of quantal Stackelberg equilibrium to Nash equilibrium. Then we show that gradient ascent methods often fail to compute the global optimum.

Proposition 4.1. *Every normal form game with a continuous quantal response function of the follower has at least one quantal Stackelberg equilibrium.*

Proof. The statement is a consequence of the fact the set of mixed strategies is convex and compact and because the utilities are finite, the quantal Stackelberg criterion QSE-NFG is continuous and bounded. □

Moreover, in zero sum normal form games, the leader's utility in quantal Stackelberg equilibrium is always greater than or equal to their utility in Nash equilibrium.

Proposition 4.2. *Let $G = (N, A, u)$ be a zero sum normal form game, QR be a quantal response function of the follower and $\delta_l^{NE}, \delta_l^{QSE}$ be the mixed strategies of the leader in Nash and quantal Stackelberg equilibria, respectively. Then*[90]

$$u_l(\delta_l^{NE}, BR(\delta_l^{NE})) \leq u_l(\delta_l^{QSE}, QR(\delta^{QSE})).$$

Proof. From the definition of quantal response function, for all $\delta_l \in \Delta_l$

$$u_f(\delta_l, BR(\delta_l))) \geq u_f(\delta_l, QR(\delta_l)).$$

Therefore, in zero-sum games it holds that

$$u_l(\delta_l, BR(\delta_l))) \leq u_l(\delta_l, QR(\delta_l)).$$

From the definition of quantal Stackelberg equilibrium

$$u_l(\delta_l, QR(\delta_l))) \leq u_l(\delta_l^{QSE}, QR(\delta_l^{QSE})).$$

The result follows from setting $\delta_l = \delta_l^{NE}$. □

The same is no longer true for general sum games. As shown in Figure 4.1, the expected utility of the leader in the general sum game is ≈ -2.26. In Nash equilibrium, the leader plays action o and the follower responds with B, resulting in the expected utility of the leader -2.2 – strictly higher than in quantal Stackelberg equilibrium. In other words, facing a boundedly rational opponent does not have

[90] In zero sum games, all the best responses of the follower result in the same utility for the leader. For the sake of exposition, we abuse the notation a little and write the utility as $u_l(\delta_l, BR(\delta_l))$ for $\delta_l \in \Delta_l$.

to be an advantage in a general sum game. From now on, we will focus on generalized Luce models and subjective expected utility evaluation functions[91].

[91] I.e., the evaluation function has a form of a strictly increasing function applied to the expected utility. For simplicity, and to emphasize its meaning, we will denote such function still e and write the evaluation value as $e(u_f(\delta_l, \pi_f))$, or, in a generalized Luce quantal response, as $[e \circ q](u_f(\delta_l, \pi_f))$.

Observation 4.3. *Let $G = (N, A, u)$ be a normal form game and q be a generator of a generalized Luce model with a subjective expected utility function e. Then the leader's strategy δ_l^{QSE} of quantal Stackelberg equilibrium in G can be formulated as the following non-concave problem:*

$$\delta_l^{QSE} = \arg\max_{\delta_l \in \Delta_l} \frac{\sum_{\pi_f \in \Pi_f} u_l(\delta_l, \pi_f)[e \circ q](u_f(\delta_l, \pi_f))}{\sum_{\pi_f \in \Pi_f} [e \circ q](u_f(\delta_l, \pi_f))}. \tag{4.1}$$

Non-concave maximization problems can be solved by gradient ascent. The gradient ascents are an efficient class of algorithms with a guarantee to approach a local maximum in a polynomial number of steps in case the gradient of the criterion function is Lipschitz-continuous.[92] Unfortunately, the criterion function of quantal Stackelberg equilibrium can have a large number of such local optima.

[92] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.

Conjecture 4.4. *For every m, $|A_f| = m$, there exists a zero sum normal form game G and a generator q of a generalized Luce model with a subjective expected utility function e, such that a criterion function of quantal Stackelberg equilibrium in G has at least $m - 1$ local maxima.*

We were able to verify the conjecture experimentally up to games with $m = 100$, then the optima became too close to zero to verify their existence numerically. We constructed the instances for every m as follows:

- The game is defined as $G_m = (\{l, f\}, \{A_l, A_f^m\}, u_f)$,
- with action spaces $A_l = \{a^1, a^2\}$, $A_f^m = \{b^1, b^2, \ldots, b^m\}$
- and utilities $u_l = -u_f$, where

$$u_f(a^1, b^1) = 0.5, u_f(a^1, b^i) = -10^i \qquad i \in [2, m]$$

and

$$u_f(a^2, b^i) = 0 \qquad i \in [m].$$

- Let $q(x) = exp(x)$ and $e(x) = x$ if $x < 0$ and $e(x) = 0$ for $x \geq 0$.

Applying restarted gradient ascent for finding quantal Stackelberg equilibrium is thus impractical and can often lead to globally suboptimal local optima. Moreover, analysing the conditions when the equilibrial criterion function from Observation 4.3 is concave is difficult, as concavity is not preserved under division[93].

[93] Generalized concavity (e.g., pseudoconvavity or quasiconcavity) is preserved under division, but it imposes too strong assumptions on the game utilities to be used in practice.

4.2 Dinkelbach-type equilibrium formulation

Now we introduce a transformation of the equilibrium formulation (4.1) into an equivalent optimization problem. This new formulation allows us to derive conditions for approximating quantal Stackelberg equilibrium in polynomial time.

Algorithm 1: Dinkelbach-type algorithm for approximating the optimal leader's strategy of quantal Stackelberg equilibrium in normal form games

$UB \leftarrow \max_{\pi \in \Pi} u_l(\pi)$, $LB \leftarrow \min_{\pi \in \Pi} u_l(\pi)$
$\delta^* \leftarrow \arg\max_{\delta_l \in \Delta_l} \sum_{\pi_f \in \Pi_f} (u_l(\delta_l, \pi_f) - LB)[e \circ q](u_f(\delta_l, \pi_f))$
repeat
  $p \leftarrow (UB - LB)/2$
  $v \leftarrow \max_{\delta_l \in \Delta_l} \sum_{\pi_f \in \Pi_f} (u_l(\delta_l, \pi_f) - p)[e \circ q](u_f(\delta_l, \pi_f))$
  $\delta_p \leftarrow \arg\max_{\delta_l \in \Delta_l} \sum_{\pi_f \in \Pi_f} (u_l(\delta_l, \pi_f) - p)[e \circ q](u_f(\delta_l, \pi_f))$
  if $v < 0$ **then** $LB \leftarrow p, \delta^* \leftarrow \delta_p$ **else** $UB \leftarrow p$
until $UB - LB < \epsilon$
return δ^*

The transformation employs the Dinkelbach's method for solving nonlinear fractional programming problems and we can adapt it for finding the quantal Stackelberg equilibrium because formulation (4.1) is a fractional program[94]. The key idea of the Dinkelbach's method is to express a problem in a form of

$$\max_{x \in M} f(x)/g(x), g(x) > 0$$

for some convex set M and continuous, real-valued functions f and g as an equivalent problem of finding a unique root of function

$$F(p) = \max_{x \in M} \{f(x) - pg(x)\} \qquad p \in \mathbb{R}.$$

Then it holds that

$$\max_{x \in M} f(x)/g(x) = p^* \iff F(p^*) = 0.$$

Because F is a maximum of functions affine in p, it is convex. Finding the root is hence straightforward[95] and it can be done efficiently if and only if we are able to effectively calculate the value of function F for any p. Computing the maximum of the original formulation of quantal Stackelberg equilibrium (4.1) is hence equivalent to solving a sequence of optimization subproblems

$$\max_{\delta \in \Delta_l} \sum_{\pi \in \Pi_f} (u_l(\delta, \pi) - p)[e \circ q](u_f(\delta, \pi)), \tag{4.2}$$

as described in Algorithm 1. We refer to expression (4.2) as the Dinkelbach subproblem of the Dinkelbach formulation of quantal Stackelberg equilibrium. Algorithm 1 iteratively updates the upper bound (UB) and lower bound (LB) on the value of the equilibrium according to a binary search method for finding a root of a function. Note that because of Proposition 4.2, we are able to set the initial lower bound

[94] W. Dinkelbach. "On nonlinear fractional programming". *Management Science* 13:7, 1967, pp. 492–498.

[95] For example, a simple binary search method can be used for this purpose.

to the value of Nash equilibrium in case the game is zero sum. This binary search approach generalizes the same technique used for computing logit quantal Stackelberg equilibrium in security games.[96] The following proposition derives sufficient conditions for Algorithm 1 to run in polynomial time.

[96] R. Yang, F. Ordonez, and M. Tambe. "Computing optimal strategy against quantal response in security games". In: *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems*. 2012, pp. 847–854.

Proposition 4.5. *Let $G = (N, A, u)$ be a normal form game, q be a twice differentiable generator of a generalized Luce model and e be a twice differentiable subjective expected utility function, such that the gradient of the Dinkelbach formulation is Lipschitz continuous. For each pure strategy $\pi \in \Pi_f$, denote*

$$\underline{u}^i_\pi = \min_{\pi_l \in \Pi_l} u_i(\pi, \pi_l), \text{ and} \qquad \overline{u}^i_\pi = \max_{\pi_l \in \Pi_l} u_i(\pi, \pi_l).$$

Moreover, let

$$\underline{u}_i = \min_{\pi \in \Pi_f} \underline{u}^i_\pi, \text{ and} \qquad \overline{u}_i = \max_{\pi \in \Pi_f} \overline{u}^i_\pi.$$

Then the quantal Stackelberg equilibrium in G is polynomially approximable[97] if for all $\pi \in \Pi_f$ and any $x \in [\underline{u}^l_\pi, \overline{u}^l_\pi]$, $y \in [\underline{u}^f_\pi, \overline{u}^f_\pi]$ and $p \in [\underline{u}_l, \overline{u}_l]$:

$$\begin{aligned} (p - x)\Big(q''(e(y))e'^2(y) + q'(e(y))e''(y)\Big) &\geq 0 \\ u_l(\pi)^T u_f(\pi) &\leq 0, \end{aligned} \tag{4.3}$$

where for any function g, g' denotes a derivative of g.

[97] The objective of this proposition is to illustrate that achieving polynomial approximability is inherently difficult, as these conditions often remain unfulfilled. This serves as a motivation for introducing a linearization method that is further developed in the rest of the chapter. An example of a setting that satisfies these conditions are games with negatively correlated utilities and traditional linear Luce quantal generators and expected utility functions.

Proof. If the Dinkelbach subproblem is concave[98], the gradient ascent methods are guaranteed to reach a global optimum. We hence analyse the conditions when an equivalent minimization formulation is convex, i.e., its Hessian matrix is positive semidefinite (PSD). The Hessian matrix is of a form

$$c \cdot u_f(\pi)u_f(\pi)^T - \big[[e \circ q](u_f(\pi)^T\delta))\big]'\big(u_l(\pi)u_f(\pi)^T + u_f(\pi)u_l(\pi)^T\big),$$

where

$$\begin{aligned} c = &\big(p - u_l(\pi)^T\delta\big)\big(q'(e(u_f(\pi)^T\delta))e''(u_f(\pi)^T\delta) \\ &+ q''(e(u_f(\pi)^T\delta))e'^2(u_f(\pi)^T\delta)\big), \end{aligned}$$

and $u_l(\pi)^T\delta$ is the expected utility $u_l(\delta, \pi)$ (and similarly for the follower). A well-known fact from linear algebra states that for non-zero n-dimensional real-valued vectors u, v, the matrix uv^T is positive semidefinite iff $u^Tv \geq 0$. By this fact, $u_f(\pi)u_f(\pi)^T$ is PSD. The matrix $c \cdot u_f(\pi)u_f(\pi)^T$ is then PSD if $c \geq 0$. Because q, e are increasing functions, matrices $-u_l(\pi)u_f(\pi)^T, -u_f(\pi)u_l(\pi)^T$ are PSD if $u_l(\pi)^Tu_f(\pi) \leq 0$. Because the expected utility $u_i(\pi)^T\delta$ always lies in the interval $[\underline{u}^i_\pi, \overline{u}^i_\pi]$ and p is upper (lower) bounded by leader's maximal (minimal) utility in the game, by substituting for c we conclude that the Dinkelbach formulation of quantal Stackelberg equilibrium is concave if conditions (4.3) holds. □

[98] Incorporating additional assumptions concerning the functions q and e, such as concavity or convexity, enables reformulating the conditions outlined herein solely in terms of the game's maximum and minimum utility. This renders the verification process easier.

4.3 Solving the Dinkelbach subproblem

In case a game and the quantal and subjective utility functions do satisfy the condition stated in Proposition 4.5, the gradient ascent methods converge to a globally optimal strategy of the leader. In the opposite case the guarantees are lost. For such cases we introduce efficient linearization methods with bounds on the solution quality to approximate the global optimum through a piece-wise linear approximation of the Dinkelbach subproblem in any game. The linear approximation is later turned into a mixed integer linear program and solved using standard methods.

4.3.1 Limitations of linear approximations in general games

In security games the formulation of the logit quantal Stackelberg equilibrium has separable variables[99] and permits efficient linearization. In contrast, the Dinkelbach formulation in normal form games is not separable in general and the probability simplex must be linearized into multivariate polytopes. The most straightforward method is to split the simplex into an exponential number of hypercubes of dimension $|A_l|$ using a fixed step size. To obtain a continuous approximation, each hypercube must be split into an exponential number of triangulations. To escape the curse of dimensionality, instead of splitting a hypercube, we compute the approximation in the whole hypercube by linear least squares. The resulting formulation is transformed to an integer linear program using the DLog construction.[100] The DLog construction is known to be the fastest because it introduces a number of binary variables only logarithmic in the number of polytopes.[101] Since we construct one polytope for every hypercube and the number of hypercubes is exponential in $|A_l|$, the final mixed integer linear program is of a polynomial size in $|A_l|$.

We implemented a method finding the piece-wise linear approximation and experimentally verified that it will not scale to games with more than 5 actions of the leader. To improve its efficiency, it is necessary to find a precise enough separable approximation of the Dinkelbach subproblem in a form of

$$\max_{\delta_l \in \Delta_l} d_p^1(\delta_l) - p d_p^2(\delta_l),$$

where the functions d_p^1 and d_p^2 are approximated as

$$d_p^1(\delta_l) = \sum_{\pi_f \in \Pi_f} u_l(\delta_l, \pi)[e \circ q](u_f(\delta_l, \pi_f)) \approx \sum_{j=1}^{|A_l|} \tilde{d}_{p,j}^1(\delta(a_l^j)), \text{ and}$$

$$d_p^2(\delta_l) = \sum_{\pi_f \in \Pi_f} [e \circ q](u_f(\delta_l, \pi_f))) \approx \sum_{j=1}^{|A_l|} \tilde{d}_{p,j}^2(\delta_l(a_l^j)),$$

where $\tilde{d}_{p,j}^1$, $\tilde{d}_{p,j}^2$ are (possibly) non-linear univariate functions.

[99] In other words, the multivariate criterion function can be written as a sum of univariate functions.

[100] T. Ibaraki. "Integer programming formulation of combinatorial optimization problems". *Discrete Mathematics* 16:1, 1976, pp. 39–52.

[101] J. P. Vielma, S. Ahmed, and G. Nemhauser. "Mixed-integer models for nonseparable piecewise-linear optimization: Unifying framework and extensions". *Operations Research* 58:2, 2010, pp. 303–315.

4.3.2 Separation via substitutions

There exists a subclass of general sum normal form games in which separable approximation can be constructed more easily than in the entire class. The key property turns out to be the linear dependency of leader's utilities on the utilities of the follower. We call such games linearly dependent[102].

[102] And we refer to the complementary class as linearly independent games.

Definition 4.2. *A normal form game G is called linearly dependent if for each strategy $\pi \in \Pi_f$ there exists a constant $c_\pi \in \mathbb{R}$, such that for each action $a_l \in \mathcal{A}_l$ it holds that $u_l(a_l, \pi) = c_\pi u_f(a_l, \pi)$.*

Subproblem formulation in linearly dependent games

In linearly dependent games[103], we can introduce a new variable y_π for each pure strategy $\pi \in \Pi_f$ with a linear substitution $u_f(\delta, \pi) = y_\pi$. The problem of solving the Dinkelbach subproblem then has the form

$$\begin{aligned} \max_{\delta_l \in \Delta_l} & \sum_{\pi \in \Pi_f} c_\pi y_\pi [e \circ q](y_\pi) - p[e \circ q](y_\pi) \\ & y_\pi = u_f(\delta_l, \pi) = \sum_{a_l \in A_l} u_f(a_l, \pi)\delta_l(a_l) \qquad \forall \pi \in \Pi_f. \end{aligned} \tag{4.4}$$

[103] Note that all zero sum games are linearly dependent since we may set $c_\pi = -1$.

Both functions $[e \circ q](y_\pi)$ and $y_\pi[e \circ q](y_\pi)$ are easily linearizable univariate functions. In case each y_π is divided uniformly into K segments, the final mixed integer linear program has $(K-1) \times |\Pi_f|$ binary variables. The following proposition posits that the error tends to zero as ϵ goes to zero and K approaches infinity.

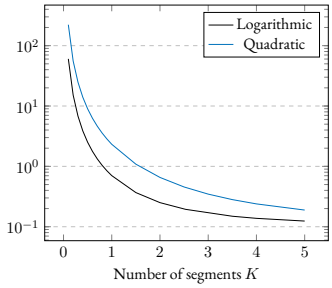


Figure 4.2: An example of the upper bound derived in Proposition 4.6 for a game with a quadratic generator $(x + 2.5)^2$, a logarithmic generator $ln(x + 2.5)$, utility range $[-1, 1]$, $c = -0.2$, $\epsilon = 0.1$, and $|A_f| = 5$.

Proposition 4.6. *Let δ_l^* be a strategy computed by Algorithm 1 with precision ϵ in a linearly dependent game G, where the subproblems are solved via linearization of formulation (4.4) with K segments. Then the quality of solution δ_l^* with respect to a mixed strategy δ_l^{QSE} of the leader in quantal Stackelberg equilibrium is upper bounded as*

$$|u_l(\delta_l^*, QR(\delta_l^*)) - u_l(\delta_l^{QSE}, QR(\delta_l^{QSE}))| \leq \epsilon + |A_f| \frac{C_1 + C_2}{4K^2 [e \circ q](\underline{u}_f)},$$

where

$$C_1 = \overline{u}_l(\overline{u}_f - \underline{u}_f)^2 \max_{x \in [\underline{u}_f, \overline{u}_f]} \left| [e \circ q]''(x) \right|$$

and

$$C_2 = (\overline{u}_f - \underline{u}_f)^2 \max_{x \in [\underline{u}_f, \overline{u}_f], \pi \in \Pi_f} |c_\pi| \left| [I \cdot (e \circ q)]''(x) \right|,$$

where I is the identity function $I(x) = x$.

Proof. For any twice differentiable univariate function g, let $\overline{g}$ be a uniform piecewise linearization of g into K segments on interval $[a, b]$. The maximum difference in values of g and $\overline{g}$ can be then upper bounded as follows:[104]

[104] M. Yano, J. D. Penn, G. Konidaris, and A. T. Patera. *Math, numerics & programming (for mechanical engineers)*. MIT Press, 2013.

$$|g(x) - \overline{g}(x)| \leq \overline{C}\frac{(b-a)^2}{8K^2}, \tag{4.5}$$

where $\overline{C} = \max_{x\in[a,b]} |g''(x)|$. The fractional criterion function (4.1) of the direct formulation of quantal Stackelberg equilibrium can be expressed as $QSE(\delta_l) = u_l(\delta_l, QR(\delta_l)) = N(\delta_l)/D(\delta_l)$, where

$$\begin{aligned} N(\delta_l) &= \sum_{\pi\in\Pi_f} c_\pi[I\cdot(e\circ q)](z_\pi) \\ D(\delta_l) &= \sum_{\pi\in\Pi_f} [e\circ q](z_\pi) \\ z_\pi &= u_f(\delta_l,\pi) \qquad\qquad \forall\pi\in\Pi_f. \end{aligned}$$

The linearized criterion function has the form $\overline{QSE}(\delta_l) = \overline{N}(\delta_l)/\overline{D}(\delta_l)$, where

$$\begin{aligned} \overline{N}(\delta_l) &= \sum_{\pi\in\Pi_f} c_\pi\overline{[I\cdot(e\circ q)]}(z_\pi) \\ \overline{D}(\delta_l) &= \sum_{\pi\in\Pi_f} \overline{[e\circ q]}(z_\pi) \\ z_\pi &= u_f(\delta_l,\pi) \qquad\qquad \forall\pi\in\Pi_f. \end{aligned}$$

Denote $\delta_l^{\overline{QSE}}$ the optimal solution of $\max_{\delta_l\in\Delta_l} \overline{QSE}(\delta_l)$. Then we have

$$\begin{aligned} d =& \left|\frac{N(\delta_l^{QSE})}{D(\delta_l^{QSE})} - \frac{N(\delta_l^*)}{D(\delta_l^*)}\right| \leq \left|\frac{N(\delta_l^{QSE})}{D(\delta_l^{QSE})} - \frac{\overline{N}(\delta_l^{\overline{QSE}})}{\overline{D}(\delta_l^{\overline{QSE}})}\right| \\ &+ \left|\frac{\overline{N}(\delta_l^{\overline{QSE}})}{\overline{D}(\delta_l^{\overline{QSE}})} - \frac{\overline{N}(\delta_l{}^*)}{\overline{D}(\delta_l{}^*)}\right| + \left|\frac{\overline{N}(\delta_l{}^*)}{\overline{D}(\delta_l{}^*)} - \frac{N(\delta_l^*)}{D(\delta_l^*)}\right|. \end{aligned}$$

We bound each of these terms separately. For the first and the last term we use Lemma 7 of Yang, Ordonez, and Tambe[96]. By this lemma, for each $\delta_l \in \Delta_l$:

$$\left|\frac{\overline{N}(\delta_l)}{\overline{D}(\delta_l)} - \frac{N(\delta_l)}{D(\delta_l)}\right| \leq \frac{1}{\overline{D}(\delta_l)}\left(\frac{N(\delta_l)}{D(\delta_l)}d_D + d_N\right), \tag{4.6}$$

where $d_D = \left|D(\delta_l) - \overline{D}(\delta_l)\right|$ and $d_N = \left|N(\delta_l) - \overline{N}(\delta_l)\right|$.

Now we may employ the linearization bound (4.5) to obtain that

$$d_D \leq |A_f| \frac{(\overline{u}_f - \underline{u}_f)^2}{8K^2} \max_{x \in [\underline{u}_f, \overline{u}_f]} \left| [e \circ q]''(x) \right| = \tilde{C}_1, \text{ and}$$

$$d_N \leq |A_f| \frac{(\overline{u}_f - \underline{u}_f)^2}{8K^2} \max_{\substack{x \in [\underline{u}_f, \overline{u}_f] \\ \pi \in \Pi_f}} |c_\pi| \left| [I \cdot (e \circ q)]''(x) \right| = \tilde{C}_2.$$

Because $\overline{D}(\delta_l) \geq [e \circ q](\underline{u}_f)$ and $QSE(\delta_l) \leq \overline{u}_l$, together with inequality (4.6) we conclude that

$$\left| \frac{\overline{N}(\delta_l)}{\overline{D}(\delta_l)} - \frac{N(\delta)}{D(\delta_l)} \right| \leq \frac{\overline{u}_l \tilde{C}_1 + \tilde{C}_2}{[e \circ q](\underline{u}_f)}. \tag{4.7}$$

Because $\delta_l^{\overline{QSE}}$ and δ_l^{QSE} maximize the formulations $\overline{QSE}$ and QSE, respectively, the difference in their values is also at most $(\overline{u}_l \tilde{C}_2 + \tilde{C}_1)/[e \circ q](\underline{u}_f)$, which provides the bound for the first term. To see why this is true, for contradiction, assume that $|\overline{QSE}(x) - QSE(x)| \leq c$, but without loss of generality $\overline{QSE}(\delta_l^{\overline{QSE}}) - QSE(\delta_l^{QSE}) > c$. By combining both inequalities we have $QSE(\delta_l^{\overline{QSE}}) > QSE(\delta_l^{QSE})$. Strategy δ_l^{QS} is hence not the argument of the maximum.

For bounding the second term we use the properties of the Dinkelbach formulation. From Algorithm 1, for each $\delta_l \in \Delta_l$, we have $\overline{N}(\delta_l) - UB\overline{D}(\delta_l) \leq 0$ and $\overline{N}(\delta_l^*) - LB\overline{D}(\delta_l^*) \geq 0$. Therefore, for the optimal strategy $\delta_l^{\overline{QSE}}$ it holds that $LB \leq \overline{N}(\delta_l^{\overline{QSE}})/\overline{D}(\delta_l^{\overline{QSE}}) \leq UB$ and hence

$$\left| \frac{\overline{N}(\delta_l^{\overline{QSE}})}{\overline{D}(\delta_l^{\overline{QSE}})} - \frac{\overline{N}(\delta_l{}^*)}{\overline{D}(\delta_l{}^*)} \right| \leq UB - LB \leq \epsilon. \tag{4.8}$$

By combining the bounds (4.7) and (4.8) we conclude that

$$|u_l(\delta_l^*, QR(\delta_l^*)) - u_l(\delta_l^{QSE}, QR(\delta_l^{QSE}))| \leq \epsilon + 2\frac{\overline{u}_l \tilde{C}_1 + \tilde{C}_2}{[e \circ q](\underline{u}_f)}.$$

□

Figure 4.2 illustrates the bound applied to a game with two potential generators.

Subproblem formulation in linearly independent games

In normal form games which are linearly independent we can take advantage of the fact that all quantal and subjective expected utility functions are invertible because they are strictly monotone. Let g^{-1} be an inverse of a function g. The substitution $[e \circ q](u_f(\delta_l, \pi)) = y_\pi$[105] leads to an easily linearizable constraint

$$u_f(\delta_l, \pi) = \sum_{a_l \in A_l} u_f(a_l, \pi)\delta_l(a_l) = [e \circ q]^{-1}(y_\pi). \tag{4.9}$$

[105] Note that if we would use the same substitution as in the case of linearly dependent games, the number of bilinear terms would be $K \times |\Pi_f|$, i.e., K-times larger than in formulation (4.10).

With an additional substitution $u_l(\delta, \pi)) = z_\pi$, the problem of solving the Dinkelbach subproblem is expressed as

$$\begin{aligned} &\max_{\delta_l \in \Delta_l} \sum_{\pi \in \Pi_f} z_\pi y_\pi - p y_\pi \\ [e \circ q]^{-1}(y_\pi) &= \sum_{a_l \in A_l} u_f(a_l, \pi)\delta_l(a_l) \qquad \forall \pi \in \Pi_f \\ z_\pi &= \sum_{a_l \in A_l} u_l(a_l, \pi)\delta(a_l) \qquad \forall \pi \in \Pi_f, \end{aligned} \tag{4.10}$$

where $z_\pi y_\pi$ are $|\Pi_f|$ bilinear terms. For linearizing the bilinear terms we can use either McCormick's envelopes,[106] which are fully linear, or the MDT method.[107] The size of the mixed integer linear program for the McCormick's envelopes is $|\Pi_f|$ real variables and $(K-1)\times|\Pi_f|$ binary variables. Experimental evaluation of the MDT method showed that while it provides close to optimal solutions, the additional binary variables required by the method significantly slow down the computation. We hence focus more on the McCormick's envelopes.

Proposition 4.7. *Let δ_l^* be a strategy computed by Algorithm 1 with precision ϵ in a linearly independent game G, where the subproblems solved via linearization of formulation (4.10) with K segments and bilinear terms relaxed using McCormick's envelopes. Then the quality of solution δ_l^* with respect to a mixed strategy δ_l^{QSE} of the leader in quantal Stackelberg equilibrium is upper bounded as*

$$|u_l(\delta_l^*, QR(\delta_l^*)) - u_l(\delta_l^{QSE}, QR(\delta_l^{QSE}))| \leq \epsilon + |A_f|\frac{C + M}{4[e \circ q](\underline{u}_f)},$$

where

$$\begin{aligned} C &= (\overline{u}_l + 1)h^2 \max_{x \in [\overline{u}_f, \underline{u}_f]} |[e \circ q]''(x)|, \\ M &= (\overline{u}_l - \underline{u}_l)(\overline{u}_f - \underline{u}_f)/4, \end{aligned}$$

and

$$h = \frac{\left([e \circ q](\overline{u}_f) - [e \circ q](\underline{u}_f)\right)}{K \max_{x \in [[e \circ q](\underline{u}_f), [e \circ q](\overline{u}_f)]} \left|[[e \circ q]^{-1}]'(x)\right|}.$$

Proof. We follow the reasoning of proof of Proposition 4.6 up to the inequality (4.6). In linearly independent games we linearize the inverse function $[e \circ q]^{-1}$ in formulation 4.10 instead of function $[e \circ q]$. To estimate the differences $|N(\delta) - \overline{N}(\delta)|$ and $|D(\delta) - \overline{D}(\delta)|$ we hence seek to bound the linearization error

$$\left|[e \circ q](x) - \overline{[e \circ q]^{-1}}^{-1}(x)\right|.$$

Assume that $[e \circ q]^{-1}$ is uniformly linearized in points $x_0, x_1, \ldots, x_{K-1}$[108]. Ev-

[106] G. P. McCormick. "Computability of global solutions to factorable nonconvex programs: Part I - Convex underestimating problems". *Mathematical Programming* 10:1, 1976, pp. 147–175.

[107] S. Kolodziej, P. M. Castro, and I. E. Grossmann. "Global optimization of bilinear programs with a multiparametric disaggregation technique". *Journal of Global Optimization* 57:4, 2013, pp. 1039–1063.

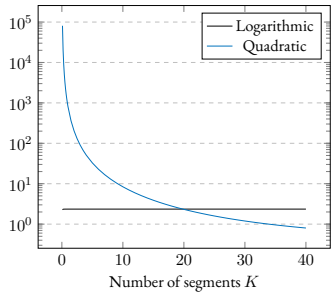


Figure 4.3: An example of the upper bound derived in Proposition 4.7 for a game with a quadratic generator $(x+3)^2$, a logarithmic generator $ln(x+3)$, utility range $[0, 10]$ and $[1, 2]$ for the leader and the follower, respectively, $\epsilon = 0.1$, and $|A_f| = 5$.

[108] In other words, for every i in $\{1, \ldots, K\}$, $[e \circ q]^{-1}(x_i) = \overline{[e \circ q]^{-1}}(x_i) = y_i$.

ery piece-wise linear function is invertible. Since $[e \circ q](y_i) = x_i = \overline{[e \circ q]^{-1}}^{-1}(y_i)$, function $\overline{[e \circ q]^{-1}}^{-1}$ is a linearization of function $[e \circ q]$. Unless the function $[e \circ q]^{-1}$ is an identity[109], the linearization is non-uniform. Since the distance between any x_i, x_{i+1} is $\big([e \circ q](\overline{u}_f) - [e \circ q](\underline{u}_f)\big)/K$, the maximum difference in values of any consecutive y_i, y_{i+1} is upper bounded by

$$h = \frac{(\overline{eq} - \underline{eq})}{K} \max_{x \in [\underline{eq}, \overline{eq}]} \Big| \big[[e \circ q]^{-1}\big]'(x) \Big|,$$

where $\underline{eq} = [e \circ q](\underline{u}_f)$ and $\overline{eq} = [e \circ q](\overline{u}_f)$. Using the bound on a maximum difference between $[e \circ q](x)$ and $\overline{[e \circ q]^{-1}}^{-1}(x)$ from inequality (4.5), we posit

$$d_D \leq |A_f| \frac{h^2}{8} \max_{x \in [\underline{u}_f, \overline{u}_f]} \big|[e \circ q]''(x)\big|. \tag{4.11}$$

For bounding the difference $|N(\delta_l) - \overline{N}(\delta_l)|$ we need to take into account also the error arising from relaxing the bilinear terms using McCormick's envelopes. The envelope is a convex polytope defined by four linear inequalities. Using simple algebra, it can be shown that the difference between a bilinear term $z_\pi y_\pi$ in formulation (4.10) and its approximation by a McCormick's envelope is at most M, where

$$M = \frac{(\overline{u}_l - \underline{u}_l)(\overline{u}_f - \underline{u}_f)}{4}. \tag{4.12}$$

Using the bound (4.12), the difference between $N(\delta_l)$ and $\overline{N}(\delta_l)$ can be finally upper bounded via

$$d_N \leq |A_f| \left(\frac{h^2}{8} \max_{x \in [\underline{u}_f, \overline{u}_f]} |[e \circ q]''(x)| + M \right). \tag{4.13}$$

We proceed as in Proposition 4.6, using inequalities (4.6), (4.8), and a modified inequality (4.7) that uses bounds (4.11) and (4.13). The result follows. □

Due to McCormick's envelopes, the bound does not approach zero with increasing K and decreasing ϵ as in Proposition 4.6. However, it can be relatively small in games where $[e \circ q](\underline{u}_f)$ is higher, as illustrated in Figure 4.3.

4.3.3 Separation via additive approximations

Besides substitution, we considered computing separable approximations also through generalized additive models.[110] Generalized additive models can be used to approximately solve the Dinkelbach subproblem in case the link function of the model is an identity. We implemented two methods for fitting additive models in the subproblem: a spline method (here, $\tilde{d}_p^1$ and $\tilde{d}_p^2$ are smoothing splines[111]) and via deep

[109] Which, in most cases, it is not, unless we assume special cases akin to q and e being identities that would further simply the analysis.

[110] T. J. Hastie. "Generalized additive models". In: *Statistical Models in S*. Routledge, 2017, pp. 249–307.

[111] G. Wahba. *Spline models for observational data*. Vol. 59. SIAM, 1990.

neural networks (here $\tilde{d}_p^1$ and $\tilde{d}_p^2$ are constructed from activation functions[112]). Experimental evaluation showed that piece-wise linear separable approximations of the Dinkelbach subproblem lead to fast computation of a solution of the approximation (especially for linearly independent games), but neither the spline method nor the neural networks provide approximations precise enough for the Dinkelbach method to converge to an optimum.

[112] W. J. Potts. "Generalized additive neural networks". In: *Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1999, pp. 194–200.

4.4 Empirical evaluation

Finally, we demonstrate practical aspects of proposed algorithms for computing quantal Stackelberg equilibrium in normal form games. As a benchmark, we use the original formulation solved by gradient ascent (GA). We compare it to the Dinkelbach-type algorithm (DTA) – Algorithm 1 with subproblems solved via substitutional piece wise linear approximations. All implementations were done in C++17. We used an implementation of the SLSQP GA algorithm in the NLOPT 2.6.1 library for non-linear optimization. A single-threaded IBM CPLEX 12.8 computed solutions of all mixed integer linear programs. The experiments were performed on a 3.2GHz CPU with 16GB RAM. Because the algorithm is domain independent, we use randomly generated games for baseline evaluation. To test the algorithm's behavior on structured domains, we use a normal form representation of a search game that serves as a standard benchmark for comparing strong Stackelberg solving methods in sequential games[113].

[113] B. Bošanský and J. Čermák. "Sequence-form algorithm for computing Stackelberg equilibria in extensive-form games". In: *Proceedings of the 29th AAAI Conference on Artificial Intelligence*. 2015, pp. 805–811.

4.4.1 Experimental domains and their instance generation

We consider six different generators of Luce models. Four of them are convex functions: $exp(0.4x)$, $exp(0.8x)$, $exp(1.2x)$ and $-1050/(x+20)$, and two are concave functions: $\ln x + 12$ and $\sqrt{x+11}$). As subjective expected utility functions we employ one risk-neutral (RN) function: x, one loss-attentive (LAt) function: $x^3/100$ and one loss-averse (LAv) function: $20/(1+e^{-x}) - 10$). The tolerance parameter for the GA algorithm in NLOPT was set to 10^{-6}[114] and $\epsilon = 2\%$ of the leader's utility range for the DTA's binary search. The linearization uses $K = 5$. For each combination of game size $\times$ generator function $\times$ expected-utility function, we constructed and solved 5 instances.

[114] To match the default CPLEX tolerance parameter.

Randomly generated normal form games

We construct random games with action spaces of sizes from $3,000 \times 50$ up to $7,500 \times 50$. The utilities of the follower are generated uniformly randomly from interval $[-10, 10]$, while the constants c_π for linearly dependent random games are generated from interval $[-3, 3]$.

Search game

The search game is played on a directed graph, depicted in Figure 4.4. The follower's goal is to reach one of the destination nodes (D_0 – D_2) from the starting node (S),

strategy space \ steps	4	5	6	7	8	9	10
# of leader's actions	256	1024	4096	16384	65536	262144	1048576
# of follower's actions	13	39	78	130	195	273	364

Table 4.1: The absolute sizes of strategy spaces in the search game as a function of the number of steps.

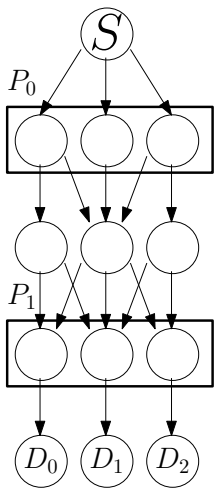


Figure 4.4: Graph for the search game.

while the leader aims to catch the follower with one of the two units operating in the marked areas of the graph (P_0 and P_1). The follower receives a different reward for reaching a different destination node. In node D_0 the reward 3.0, in D_1 4.0, and in D_2 it is 5.0. The leader can move their units freely over the neighboring nodes in the areas. In case the leader captures the follower, the follower is penalized by receiving a utility -2.0. We consider a version of the game in which neither player perceives any information and the leader's utility is complementary to the follower's. Both players can make N steps. The size of the strategy spaces for different number of steps is depicted in Table 4.1.

4.4.2 Experimental results

In Figure 4.5 we depict the runtimes achieved with the DTA in linearly dependent randomly generated games. The x-axis varies the game size, while the y-axis shows the mean runtime for computing the solutions in games of a given size. Every point in the graph corresponds to the mean over the sampled instances and shows also the achieved standard error. The results show that with the increasing size of the game, the speedup of the DTA also increases, being up to 25.5-times faster than one restart of the GA for games with 7500 leader's actions. It suggests that for even larger games, the Dinkelbach type algorithms should perform significantly better than the gradient ascent. The results on the search game are then shown in Figure 4.6. The superiority of DTA over GA is even more profound here, with the gradient ascent not being able to compute solutions even for games with 7 steps within the limit.

The relative errors of computed solutions in random games are presented in Figure 4.7. The x-axis again varies the game size. The y-axis shows the mean ratio of the difference in the leader's expected utility computed using one restart of the GA and the DTA to the absolute difference between the maximal and minimal utility in the game. While the DTA is guaranteed to approximate a global solution, there is no guarantee of a solution quality for the GA. The results show that gradient ascent never finds a better solution than the algorithm based on Dinkelbach reformulation[115]. However, the GA's solution quality is comparable to the DTA for concave functions and the hyperbole. The quality degrades rapidly for the exponential generators with all three evaluation functions, especially for higher coefficients. For the loss-averse function, the quality of the solutions computed by gradient ascent with exponential quantal function with coefficient 1.2 is on average at least 42.5-times worse than the solution of the Dinkelbach type algorithm. We observe a similar behavior also in the search game, as depicted in Figure 4.8, except for the higher-α

[115] Otherwise the ratio would be negative.

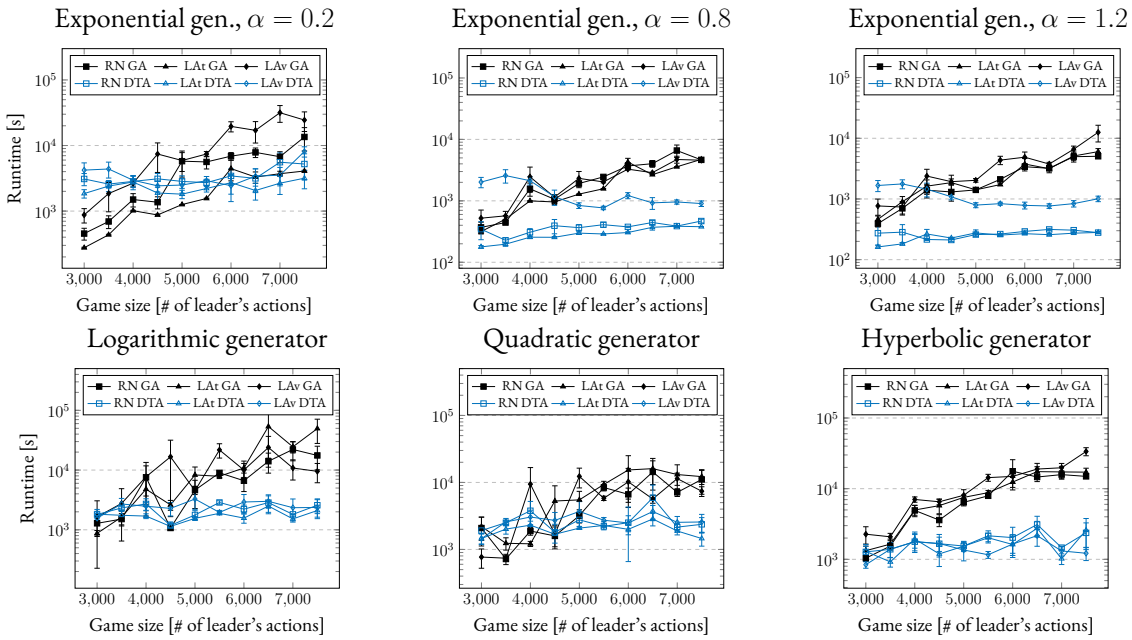


Figure 4.5: Mean runtimes of computing solutions using gradient ascent and Dinkelbach type algorithm in randomly generated games. Each graph differentiates risk-neutral (RN), loss-attentive (LAt) and loss-averse (LAv) evaluation functions. Every point shows also a standard error.

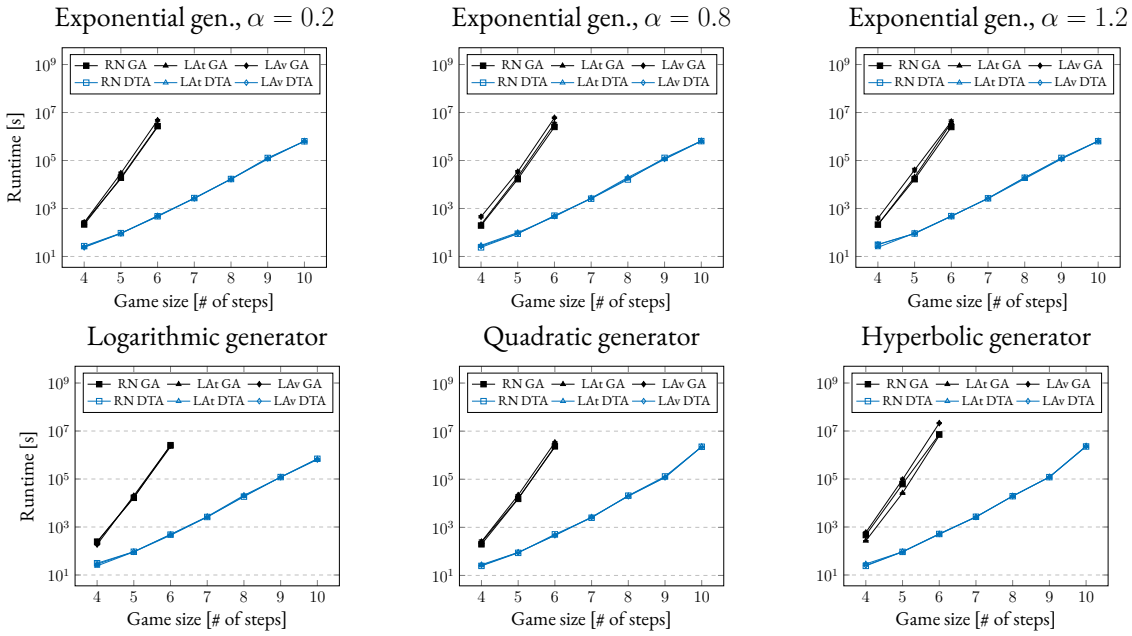


Figure 4.6: Mean runtimes of computing solutions using gradient ascent and Dinkelbach type algorithm in search games. Each graph differentiates risk-neutral (RN), loss-attentive (LAt) and loss-averse (LAv) evaluation functions. Every point shows also a standard error.

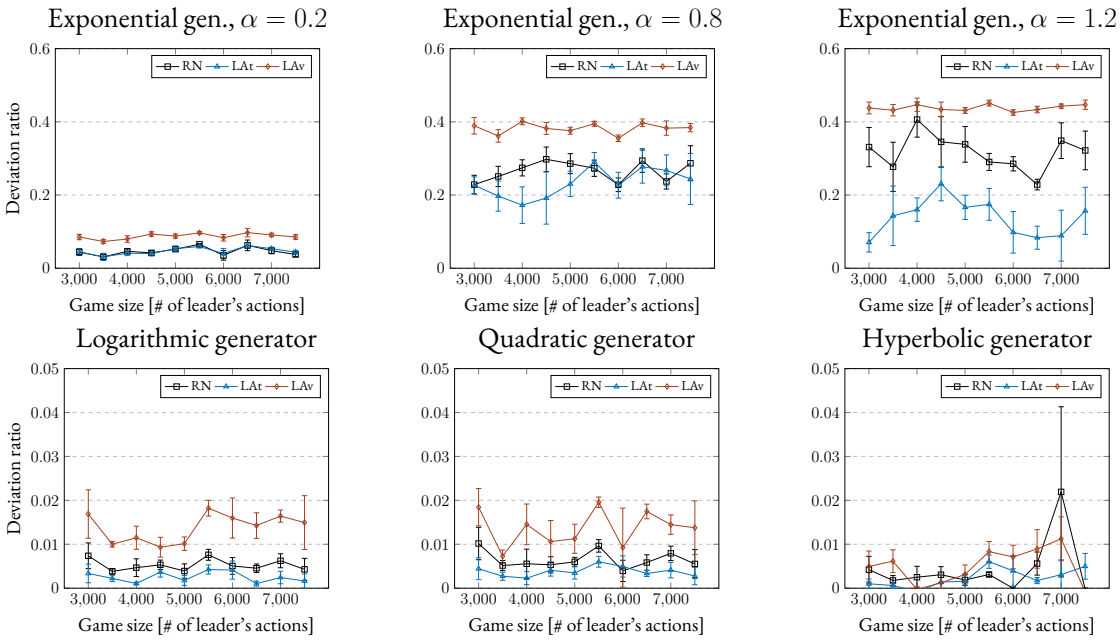


Figure 4.7: Mean deviations of solutions computed using gradient ascent and Dinkelbach type algorithm in randomly generated games. Each graph differentiates risk-neutral (RN), loss-attentive (LAt) and loss-averse (LAv) evaluation functions. Every point shows also a standard error.

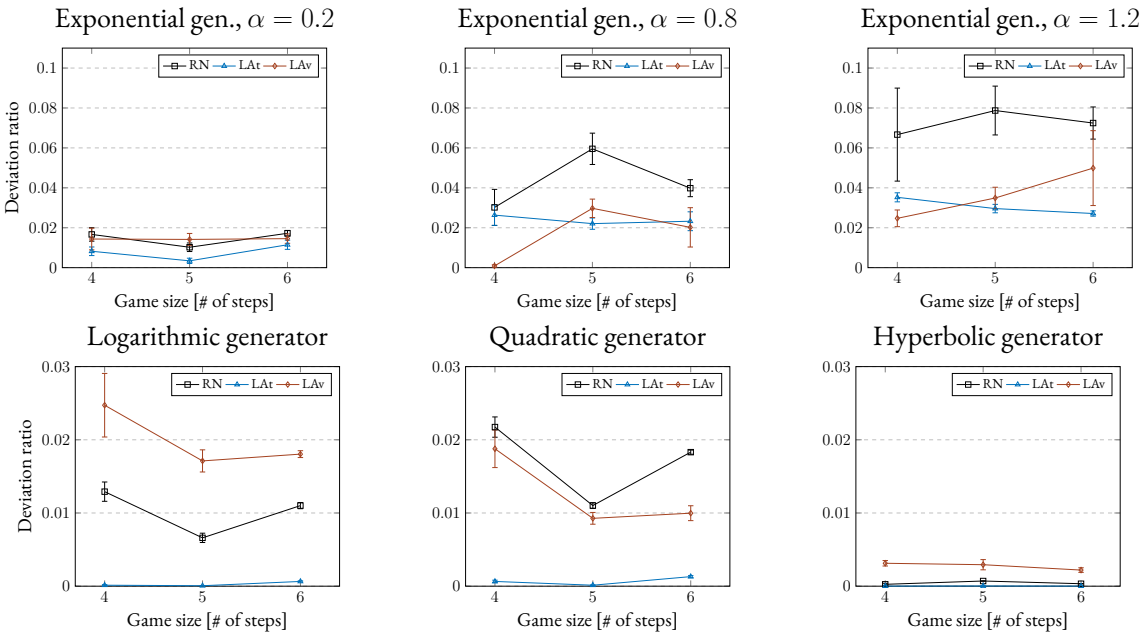


Figure 4.8: Mean deviations of solutions computed using gradient ascent and Dinkelbach type algorithm in search games. Each graph differentiates risk-neutral (RN), loss-attentive (LAt) and loss-averse (LAv) evaluation functions. Every point shows also a standard error.

	Logarithmic gen.	Quadratic gen.	Hyperbolic gen.
Risk-neutral restarts			
3000 leader's actions	7.60 ± 3.68	5.40 ± 3.30	1.80 ± 0.71
4000 leader's actions	1.00 ± 0.00	6.25 ± 4.60	1.40 ± 0.35
5000 leader's actions	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
Loss-attentive restarts			
3000 leader's actions	1.00 ± 0.00	4.80 ± 3.39	1.00 ± 0.00
4000 leader's actions	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
5000 leader's actions	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
Loss-averse restarts			
3000 leader's actions	19.00 ± 1.15	6.25 ± 4.58	1.60 ± 0.35
4000 leader's actions	5.75 ± 4.75	13.60 ± 7.31	1.00 ± 0.00
5000 leader's actions	4.00 ± 2.38	2.30 ± 1.53	1.20 ± 0.17

Table 4.2: The mean number of gradient ascent restarts needed to reach 1% deviation (or better) from the solution computed by the Dinkelbach type algorithm. All exponential generators required more than 20 restarts.

exponential generators, where the gradient ascent perform better than in random games[116]. Still, the algorithm on the Dinkelbach reformulation returns a solution of higher quality on average. Note that we could also compare only solutions for the smallest games because of the gradient ascent's poor scalability.

The number of restarts of the GA required to reach a deviation from the DTA's solution less than 1% is shown in Table 4.2. We restarted the gradient ascent 20 times and counted the first restart when the deviation was sufficiently small. With neither exponential generator was the algorithm able to find a good enough solution within the 20 restarts. This suggests that to obtain comparable results with the most commonly used quantal function in the literature – the exponential logit generator with a higher coefficient – in games with at least 7000 actions of the leader, the ascent has to run at least 153× longer than the DTA. In contrast, the concave generators and the hyperbole computed nearly optimal solutions within the first few restarts. This behavior is consistent with the deviations reported in Figure 4.7.

[116] This could be possibly attributed to the game's inner structure.

❖

In linearly independent randomly generated games the DTA achieved even faster speedups, as it employs the McCormick's envelopes: up to 3030-times faster than one restart of the GA for exponential generators and games with at least 7000 actions. However, due to the relaxation, the quality of the solutions was often worse than with gradient ascent, but no more than 12% for the exponential generators and 6% for other quantal functions. The gradient ascent running on the Dinkelbach subproblem has the theoretical guarantees, as proved in Proposition 4.5, but its runtimes are comparable to the gradient ascent on the direct formulation.

4.5 Summary of contributions

We introduced a Dinkelbach type formulation of a problem of computing a boundedly rational quantal Stackelberg equilibrium in normal form games. In contrast to the direct formulation, the Dinkelbach formulation has both the theoretical advantages (i.e., we can explicitly check how closely we approximate the global optimum) as well as positive computational consequences – the formulation offers up to 25.5-times speedup when compared with the original formulation.

5 Quantal Stackelberg Equilibrium in Extensive Form Games

Previous chapter introduced methods for computing optimal strategies to commit to against a quantal opponent in one-shot scenarios. These strategies are efficient in situations when we face myopic agents yet many other interactions in the real world are sequential, dynamically changing as per the agents' strategies. In this chapter, we hence develop methods for finding quantal Stackelberg equilibrium in extensive form games that model such dynamic interactions. There are two main challenges that prevent us from using techniques from security games or normal form games directly. First, while there is only one decision point for the boundedly rational player to act at in normal form games, any non-trivial extensive form game contains many causally linked decision points where the same player acts. The psychological studies show that humans prefer short-term, delayed heuristic decisions rather than long-term, premeditated decisions.[117] This behavior arises especially in conflicts[118] or when facing information overload caused by large decision space.[119] Therefore, a natural assumption is that the boundedly rational player acts according to their quantal response model in each decision point separately. This behavior cannot be modeled by an equivalent normal form game. Second, contrary to one-shot scenarios, the number of rational player's pure strategies in extensive form games is exponential in the game size. Therefore, even if the equilibria would coincide in normal and extensive form games, applying the algorithms for normal form games directly would result in exponentially worse scalability.

We begin our analysis by showing that finding quantal Stackelberg equilibrium is NP-hard, and the straightforward formulation of the equilibrium in extensive form games corresponds to a non-concave fractional program that is difficult to optimize. Therefore, we derive an equivalent Dinkelbach-type formulation of the equilibrium that does not contain any fraction and represents the rational player's strategy as an expression linear in the size of the game. We use the Dinkelbach formulation to identify sufficient condition for solving the problem in polynomial time. If the conditions are satisfied, the optimal solution can be found by gradient ascent. For other cases, we formulate a mixed integer linear program approximating the equilibrium through a linear relaxation of the quantal response model. We provide theoretical guarantees on the solution quality depending on the number of segments used to linearize the quantal response function and the arising bilinear terms.

The results in this chapter were published as J. Černý, V. Lisý, B. Bošanský, and B. An. "Computing Quantal Stackelberg equilibrium in extensive-form games". *Proceedings of the 35th AAAI Conference on Artificial Intelligence* 35:6, 2021, pp. 5260–5268.

[117] G. Gigerenzer and D. G. Goldstein. "Reasoning the fast and frugal way: Models of bounded rationality." *Psychological Review* 103:4, 1996, p. 650.

[118] J. R. Gray. "A bias toward short-term thinking in threat-related negative emotional states". *Personality and Social Psychology Bulletin* 25:1, 1999, pp. 65–75.

[119] N. K. Malhotra. "Information load and consumer decision making". *Journal of Consumer Research* 8:4, 1982, pp. 419–430.

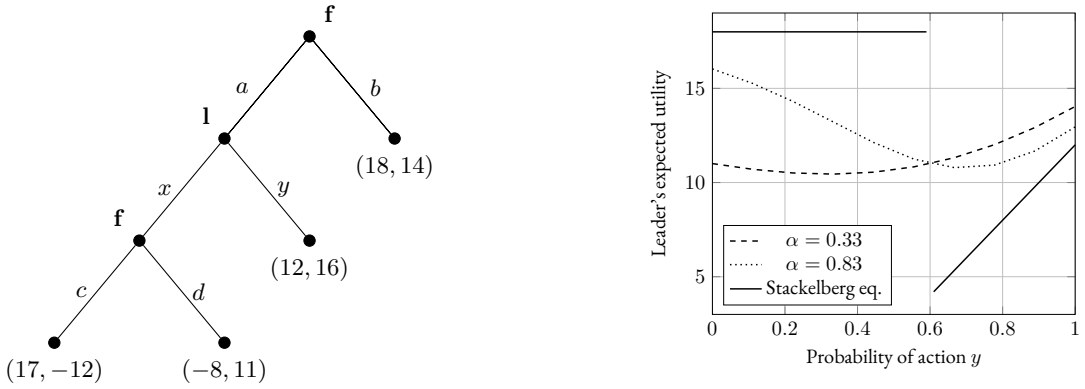


Figure 5.1: (Left) An example of a general sum extensive form game with utilities in form (u_l, u_f), and (Right) the objective functions of three equilibria: quantal Stackelberg equilibria with a generator $q = exp(\alpha x)$, where $\alpha \in \{0.33, 0.83\}$, and strong Stackelberg equilibrium.

In the empirical evaluation we compare the direct formulation solved by an algorithm for general non-linear optimization to our mixed integer linear program. We show that in 3.5 hours, our algorithm computes solutions that the baseline cannot reach within three days. Moreover, for solvable instances the solutions of our algorithm outperform the baseline's solutions.

5.1 Problem definition and properties

Identically to the previous chapter, even in extensive form games we consider two causes of emergence of boundedly rational behavior that combine into a formal definition of quantal response function QR: (i) a subjective perception of expected utilities and (ii) a proneness to making mistakes when choosing an action to play. The equilibrium then has the following form.

Definition 5.1. *Given an extensive form game G, a behavioral strategy $\beta_l^{QSE} \in B_l$ and a quantal response function QR of the follower form a Quantal Stackelberg Equilibrium if and only if*

$$\beta_l^{QSE} = \arg\max_{\beta_l \in B_l} u_l(\beta_l, QR(\beta_l)). \qquad \text{(QSE-EFG)}$$

Quantal Stackelberg equilibrium of an extensive form game is not equivalent to the same equilibrium of a normal form representation of the said game because instead of picking a pure strategy in the whole game according to a given model of bounded rationality, the follower acts quantally in their information sets separately[120]. In other words, a quantal response over pure strategies can not be in general decomposed into valid quantal responses in information sets, and vice versa.

[120] We consider this construction, in a manner of agent quantal response equilibrium, more suitable for sequential games.

Example 5.1. *Consider an extensive form game depicted in Figure 5.1. We construct quantal response functions for the follower in this game from an evaluation function*

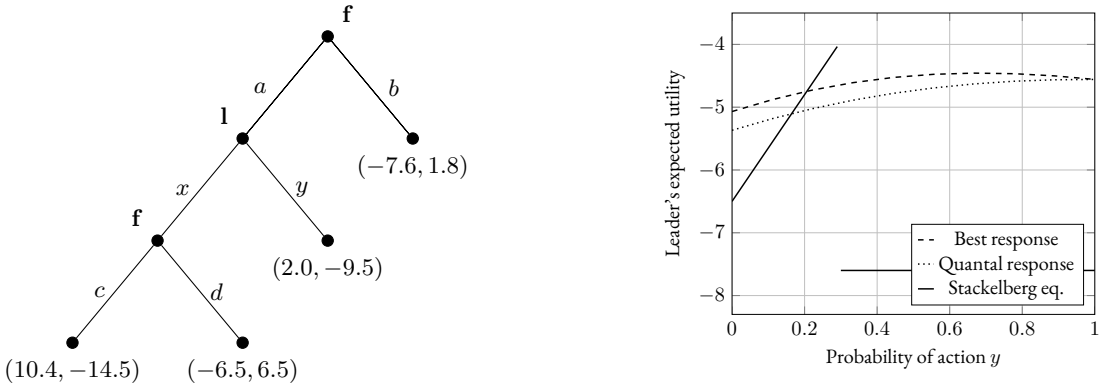


Figure 5.2: (Left) An example of a general sum extensive form game, and (Right) the objective functions of three equilibria: quantal Stackelberg equilibria with evaluation functions using best response and quantal response, respectively, and strong Stackelberg equilibrium.

and two generalized Luce models. Let the follower be unaware that they make systematic mistakes and let them evaluate actions $a \in \chi(I)$ in any information set I via best responses in the lower decision points as

$$e(\beta_l, a) = \frac{\sum_{z \in Z: a \in seq_i(z)} u_i(z) C(z) \beta(seq(z))}{\beta_f(seq_f(I)a) \sum_{h \in I} \beta_l(seq_l(h))},$$

where $\beta_f \in BR(\beta_l)$ is an arbitrary best response[121]*. In reality, they will act quantally according to a generator $q(x) = exp(\alpha x)$. In the first quantal response, we set $\alpha = 0.33$, while in the second one, let α be equal to 0.83. On the right of the Figure we show the non-concave objective functions of both resulting quantal Stackelberg equilibria, as well as strong Stackelberg equilibrium. The choice of the behavioral model significantly affects the solution. While with $\alpha = 0.33$ the leader commits to playing action y, with $\alpha = 0.83$ their strategy is completely opposite: to play action x. An optimal solution in traditional Stackelberg equilibrium is to play any strategy with probability of action y lower than 0.6. However, if the leader deploys a strategy close to this threshold against either of the two behavioral models, their utility will be, in fact, close to the global minima of the corresponding quantal equilibria criterion functions. For $\alpha = 0.33$, the utility is low for all Stackelberg strategies.*

We observe a similar behavior even when varying the evaluation functions, instead of the generator. For the extensive form game depicted on the right in Figure 5.2 we construct two quantal Stackelberg equilibria with a fixed generator $q(x) = exp(\alpha x)$, $\alpha = 0.068$ and two different evaluation functions. The first evaluation function is the same as in the previous example, i.e., the follower assumes they act entirely rationally and uses the expected utility of an arbitrary best response as their evaluation function in both information sets. The second evaluation function represents a follower who recognizes their own bounded rationality and computes an expected utility of an action under the assumption they act quantally in the subsequent decision points.

[121] Note that all best responses yield the same utility.

On the left in Figure 5.2 we show the objective functions of the corresponding quantal Stackelberg equilibria and the fully rational strong Stackelberg equilibrium. An optimal strategy for the leader in the model with quantal response evaluation function is to always play action y. In contrast, when the follower uses the best response, the leader should commit to playing action y with probability only ≈ 0.71. Note that both strategies are relatively far from the tradition Stackelberg strategy and the leader would lose a large portion of the utility in case they decide to deploy a quantal Stackelberg strategy against an entirely rational follower. The objective function of both boundedly rational equilibria are fairly similar because the game is small and contains only two information sets of the follower. Our experiments confirmed that as the sizes of games reach lower tens of information sets of the follower the role of the evaluation function becomes substantially more important.

These examples replicate the observations made also in security games: playing Stackelberg strategies against boundedly rational opponents may inflict huge losses in utility for the leader.[122] In fact, it is not difficult to design an extensive form game in which a unique Stackelberg equilibrium is a global minimum of a quantal Stackelberg criterion with an arbitrarily low utility for the leader.

[122] R. Yang, C. Kiekintveld, F. Ordonez, M. Tambe, and R. John. "Improving resource allocation strategy against human adversaries in security games". In: *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*. Barcelona, Catalonia, Spain, 2011, pp. 458–464.

❖

Let us denote the unique predecessor of a node $h \in H \setminus \{h_0\}$ in an extensive form game G as $pr(h)$. The equilibrium then has the following formulation.

Observation 5.1. *Let G be an extensive form game and q be a generator of a generalized Luce model with an evaluation function e. Finding the leader's strategy β_l^{QSE} in quantal Stackelberg equilibrium in G can be formulated as the following non-concave problem:*

$$\begin{aligned}
\beta_l^{QSE} &= \arg\max_{\beta_l \in B_l} v(\emptyset, \emptyset) \\
v(pr(h)) &= \sum_{a \in \chi(h)} v(h,a) C(a) && \forall h \in H_c \\
v(pr(h)) &= \sum_{a \in \chi(h)} v(h,a) \beta_l(a) && \forall h \in H_l \\
v(pr(h)) &= \frac{\sum_{a \in \chi(h)} v(h,a) q(e(\beta_l, a))}{\sum_{a \in \chi(h)} q(e(\beta_l, a))} && \forall h \in H_f \\
v(pr(z)) &= u_l(z) && \forall z \in Z.
\end{aligned} \tag{5.1}$$

The variable v is defined for every action interconnecting two consecutive nodes in the game tree (i.e., an edge) and it serves to propagate the leader's utility from the leafs up to the root through both the chance nodes H_c and nodes of the players H_l and H_f. As Example 5.1 shows, this formulation using behavioral strategies is non-concave, it may have multiple local optima, and it contains fractional terms.

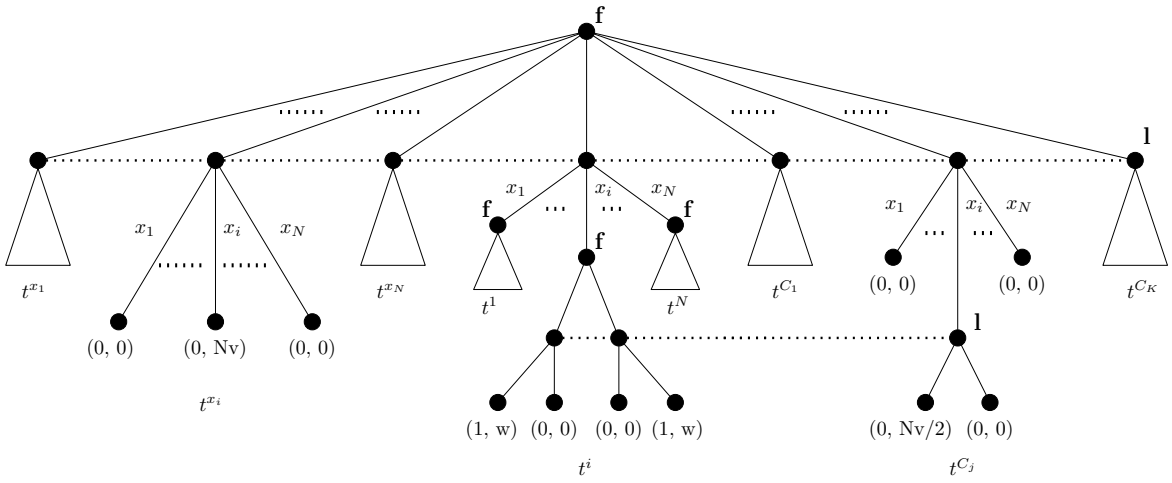


Figure 5.3: The game $G_{3SAT}(I)$ used by Letchford and Conitzer[123] in a reduction from the 3SAT problem with modified utilities to accommodate for a quantal follower. The subtrees are visualized as triangles, otherwise the figure follows a standard denotation of extensive form games. In the depicted clause subtree t^{C_j}, the corresponding clause C_j contains a literal $+x_i$.

Similarly to Stackelberg equilibrium, computing quantal Stackelberg equilibrium in extensive form games is an NP-hard problem, as we show via two reductions.

Theorem 5.1. *The problem of finding a quantal Stackelberg equilibrium in a two-player extensive form game where the follower acts according to a generalized Luce model with a logit generator is NP-hard.*

Proof. We adapt the argument from the proof of NP-hardness of strong Stackelberg equilibrium in extensive form games[123]. We reduce any instance I of a 3SAT problem to a game $G_{3SAT}(I)$. We use the same tree structure as in the original result, depicted in Figure 5.3, but modify the utilities to accommodate for the bounded rationality of the follower. One of the main complications is that the original proof relies on the follower playing the leader's preferred action in case of indifference, while the quantal response chooses all actions with the same utility uniformly.

Let the instance I consist of N variables $x_1, \ldots, x_N$ and K clauses $C_1, \ldots, C_K$. The structure of the game $G_{3SAT}(I)$ is as follows. In the root, the follower picks one subtree of three possible types. First, for each variable x_i there is a subtree t^{x_i}. Second, for each clause C_j of I the game contains a subtree t^{C_j}. Finally, there is the target subtree t, which is the only subtree in which the leader can reach a non-zero utility. In the roots of all subtrees the leader acts, and they all belong to the same information set, let us call it S. There is one action per variable in S. In the variable subtree t^{x_i}, the immediate leafs have utility $(0, 0)$ everywhere except for the action corresponding to variable x_i. In the clause subtree t^{C_j}, the actions lead to immediate leafs $(0, 0)$ in case the action does not appear in C_j. If C_j contains $+x_i$, there is a small subtree (as shown in the Figure) with leafs $(0, N/2)$ and $(0, 0)$. Vice versa, if there is $-x_i$ in C_j, the leafs have utilities $(0, 0)$ and $(0, N/2)$. In the

[123] J. Letchford and V. Conitzer. "Computing optimal strategies to commit to in extensive-form games". In: *Proceedings of the 11th ACM Conference on Electronic Commerce*. 2010, pp. 83–92.

target subtree, the players play a coordination game. Besides S, the leader has one information set S_i for each variable x_i. There are two possible actions in each S_i that correspond to setting the variable to $true$ or $false$. The original result shows that I is satisfiable if and only if the leader is able to incentivize the follower to reach subtree t. With a quantal follower, reaching t with probability 1 is not possible. However, we will show that there is a probability threshold for reaching t separating solvable and unsolvable instances of 3SAT.

The key observation is that if q is strictly increasing, then the boundedly rational player plays actions with strictly greater utilities with strictly greater probability. The leader hence tries to maximize the follower's utility in the target tree t in G_{3SAT} because as the players' utilities in t are correlated, the follower playing into t with higher probability will increase the leader's expected utility[124]. We modify the utilities in the original G_{3SAT} in the following way:

[124] Note that the leader's utilities everywhere else in the tree are zero.

1. in subtrees t^{x_i}, the utility corresponding to action x_i is $(0, Nv)$;
2. in subtree t, the coordinated utilities are $(1, w)$; and
3. in subtrees t^{C_j} the utilities are $(0, Nv/2)$.

We set $v = wq(w)/(q(0)+q(w))$. As an evaluation function e of the follower, we use a modified entirely rational expected utility function with some given constant $\gamma > 0$. Let $u_f(\beta, a)$ be an expected-utility of a follower's action a against the leader's strategy β. We define e as

$$e(a,\beta) = \begin{cases} u_f(\beta,a) & \text{if } u_f(\beta,a) \in (\infty, v-\gamma) \cup (v+\gamma, \infty), \\ v & \text{if } u_f(\beta,a) = v, \\ v-\gamma & \text{if } u_f(\beta,a) \in [v-\gamma, v), \text{ and} \\ v+\gamma & \text{if } u_f(\beta,a) \in (v, v+\gamma]. \end{cases}$$

Now we bound the reaching probability of t in satisfiable and unsatisfiable instances.

Satisfiable: Assume the formula is satisfiable and the leader commits to playing the satisfying assignment in their lower information sets S_i and the uniform strategy in the upper set S, similarly as in the strong Stackelberg equilibrial strategy in the original G_{3SAT}[125]. Because the follower plays according to their quantal function, their expected utility in t is v and so is in each subtree t^{x_i}. In the worst case, the follower receives utility v also in t^{C_j}, so the target subtree t is reached with a probability at least $p_0 = 1/(N + K + 1)$.

[125] We highlight only the main differences, for details about the original proof see the paper of Letchford and Conitzer[123].

Unsatisfiable: Now assume that I is not satisfiable. We consider four cases:

1. The leader plays an arbitrary pure strategy corresponding to some assignment in the lower information sets S_i and a uniform strategy in their upper set S. There must exist C_u which is not satisfied. Even in case the leader is able to guarantee the

minimal utility 0 for the follower in all $t^{C_j}, j \neq u$, the subtree t is reached with probability at most $p_1 = q(v)/(q(v)(N + 1) + q(3/2v) + q(0)(K - 1))$.

2. The leader plays an arbitrary pure strategy corresponding to some assignment in information sets S_i, but changes their strategy in S. Then there exists a subtree t^{x_s}, such that the follower's expected utility in this subtree is $(1+\epsilon)v, \epsilon > 0$. Even if the leader guarantees the minimal utility 0 for the follower in all $t^{x_i}, i \neq s$ and t^{C_j}, the subtree t is reached with probability at most $p_2 = q(v)/(q(v) + q(v + \gamma) + q(0)(N - 1)K)$.
3. The leader plays a mixed strategy in some S_i and a uniform strategy in S. Then their expected utility in the target tree t is $(1 - \epsilon)v, \epsilon > 0$. Even in case the leader is able to guarantee the minimal utility 0 for the follower in all t^{C_j}, the subtree t is reached with probability at most $p_3 = q(v - \gamma)/(q(v)N + q(v - \gamma) + q(0)K)$.
4. The leader plays a mixed strategy in some S_i and also changes their strategy in S. Then there exists a subtree t^{x_s}, such that the follower's expected utility in this subtree is $(1+\epsilon)v, \epsilon > 0$. Even if the leader guarantees themselves utility v in t and the minimal utility 0 for the follower in all $t^{x_i}, i \neq s$ and t^{C_j}, the subtree t is reached with probability at most $p_4 = p_2 = q(v)/(q(v) + q(v + \gamma) + q(0)(N - 1)K)$.

To separate the satisfiable and unsatisfiable instances, the inequalities $p_1, p_2, p_3, p_4 < p_0$ must hold. They can be expressed as

$$q(v) < \max\left(\frac{q(\frac{3v}{2})}{K}, \frac{q(v+\gamma)}{N+K}\right)$$
$$q(v - \gamma) < \frac{q(v)}{K+1}.$$

Because q is an exponential function, it has an inverse q^{-1}. Moreover, for a non-negative w it always holds that $w > v$. The set of inequalities is thus satisfied by

$$w^* = \max(\phi, -2q^{-1}(\frac{1}{K}))$$
$$\gamma = \max(\phi, -q^{-1}(\frac{1}{N+K}), -q^{-1}(\frac{1}{K+1})),$$

for an arbitrary $\phi > 0$. Finding the probability of reaching t in quantal Stackelberg equilibrium of the modified $G_{3SAT}(I)$ with $w \leftarrow w^*$ and $\gamma \leftarrow \gamma^*$ hence answers the question of satisfiability of I. □

		Follower		
		A	B	C
Leader	o	0,0	10,-10	0,0
	r	0,0	0,0	10,-10

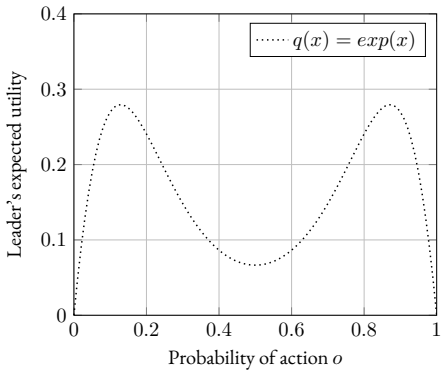


Figure 5.4: (Left) An example of a zero sum normal form game and (Right) a criterion function of its quantal Stackelberg equilibrium with a generator $q(x) = exp(x)$ and expected utility as an evaluation function. The figure follows a standard denotation of normal form games.

This proof shows a reduction from a strongly NP-complete problem, but requires the follower to use a specific generator function – the logit quantal response. Now we give an alternative proof that works for arbitrary quantal response functions, but the reduction is from a weakly NP-complete problem.

Theorem 5.2. *The problem of finding a quantal Stackelberg equilibrium in a two-player extensive form game that is*

- *zero sum and the follower acts according to a generalized Luce model with a logit generator and an expected-utility evaluation function, or*
- *general sum and the follower acts according to an arbitrary quantal response*

is NP-hard.

Proof. We reduce the problem of solving an instance I of the partition problem to finding quantal Stackelberg equilibrium in a specific extensive form game $G_P(I)$. We assume the instance of the partition problem to be a multiset of positive integers $(x_i)_{i\in[K]}$. The question is whether there is a set of indices $J \subset [K]$, such that

$$\sum_{i\in J} x_i = \sum_{i\in[K]\setminus J} x_i. \tag{5.2}$$

A key role in the reduction plays a special normal form game that in itself has two distinct quantal Stackelberg equilibria, and in neither of them the leader commits to a uniform strategy[126]. We show examples of such zero and general sum games.

[126] The fundamental concept of this reduction comes from Viliam Lisý, which I formalized and extended to determine the equilibria and their corresponding expected utilities, effectively distinguishing solvable and unsolvable instances.

Reduction to a zero sum game

We begin with zero sum games. In Figure 5.4, we depict a zero sum normal form game that has two such quantal Stackelberg equilibria. We assume the follower acts according to a logit generator and uses expected utilities as action values. In the first

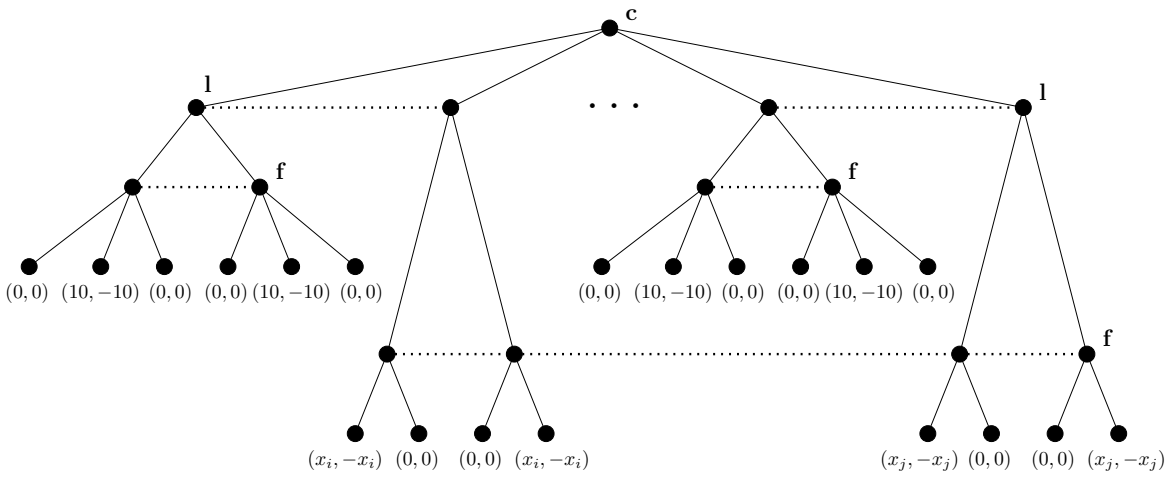


Figure 5.5: The game $G_P(I)$ used in a reduction from the partition problem. Only the subtrees for items x_i and x_j are explicitly included.

equilibrium, the leader commits to their action o with probability s. The second equilibrium they play the other action r, also with probability s. The expected utility of the leader when playing either of these strategies is m, while any other strategy, and particularly the uniform strategy, achieves a strictly lower utility.

Now we construct the game $G_P(I)$ that makes the leader commit to a strategy that solves the partition problem. The game starts with a uniform chance node. Each item is then associated with a subgame under the chance node. The structure of the tree is depicted in Figure 5.5 and shows two subgames, for items x_i and x_j. Each subgame has two components. We call the first component the normal form subtree and we depict it on the left. It is the extensive form representation of the normal form game from Figure 5.4. To maximize their utility, the leader is motivated to play the first action with probability either s or $1 - s$, but not a uniform strategy. The second component, on the right, is the partition subtree. One large information set of the follower connects all the partition subtrees. The purpose of these subtrees is to enforce the solution of the partition problem. Consider the quantal Stackelberg equilibria in G_P for solvable and unsolvable instances.

Solvable: First, we construct the equilibrium of this game in case the partition problem has a solution J[127]. To maximize the utility in the normal form subtrees, in each of their information sets, the leader chooses only from the two equilibrial strategies s and $1 - s$. If an item belongs to the set J, the leader plays the strategy s. Choosing the strategy $1 - s$ means the item is from the complementary set. The expected utility of the follower of the first action a_1 in their bottom large two-action information set is then

$$u_f(a_1) = -\sum_{i \in J} s\frac{x_i}{2K} - \sum_{i \in [K] \setminus J} (1 - s)\frac{x_i}{2K},$$

[127] In other words, the index set J solves Equation (5.2).

while the expected utility of the second action a_2 is

$$u_f(a_2) = -\sum_{i \in J}(1-s)\frac{x_i}{2K} - \sum_{i \in [K] \setminus J} s\frac{x_i}{2K}.$$

Because J is the solution, we have $u_f(a_1) = u_f(a_2)$ and the follower is incentivized to play uniformly. The leader's utility in the partition subtrees is hence

$$u_l^U = \sum_{i \in [K]} \frac{x_i}{4K} = \frac{1}{4K}\sum_{i \in [K]} x_i.$$

Next, we show that utility u_l^U is in fact optimal in the partition subtrees; the leader can never achieve a higher utility. Let x be a vector of the multiset integers of the partition problem and σ be a vector of arbitrary probabilities of playing the first action in the leader's partition subtrees. We aim to prove that for any σ and the corresponding vector of complementary probabilities of playing the second action $\overline{1} - \sigma$, where $\overline{1}$ is a vector of 1's, it holds that

$$\frac{1}{2K}\frac{x^T\sigma q\left(-\frac{x^T\sigma}{2K}\right) + x^T(\overline{1}-\sigma)q\left(-\frac{x^T(\overline{1}-\sigma)}{2K}\right)}{q\left(-\frac{x^T\sigma}{2K}\right) + q\left(-\frac{x^T(\overline{1}-\sigma)}{2K}\right)} \leq u_l^U.$$

Simple algebra shows this is equivalent to

$$\frac{x^T(2\sigma - \overline{1})}{2}\left(q\left(-\frac{x^T\sigma}{2K}\right) - q\left(-\frac{x^T(\overline{1}-\sigma)}{2K}\right)\right) \leq 0. \quad (5.3)$$

Because we have

$$q\left(-\frac{x^T\sigma}{2K}\right) - q\left(-\frac{x^T(\overline{1}-\sigma)}{2K}\right) \leq 0 \iff \frac{x^T(2\sigma-\overline{1})}{2} \geq 0,$$

the Equation (5.3) always holds and u_l^U is indeed an upper bound. Because the leader's utility is maximized in both the normal form and the partition subtrees, it is a quantal Stackelberg equilibrium and the leader's utility if the partition problem is solvable is therefore

$$u_l^* = \frac{K}{2} + \frac{1}{2K}\sum_{i \in [K]} x_i.$$

Unsolvable: Second, assume that the partition problem does not have a solution. We show that in this case, the utility of the leader in quantal Stackelberg equilibrium will be always strictly lower than u_l^*. Observe that

because the equilibrium with solvable instances achieves a maximum possible utility in the partition subtrees, in order to attempt to reach the same overall utility with unsolvable instances, the leader has to commit to the solution of the normal form game. Therefore, in each partition subtree, their only viable strategy is to play the first action with probability either s or $1-s$. First, we analyze the utility of the leader in case the strategy of the follower is not uniform. From Equation (5.3), it follows that in case a vector σ maximizes a utility of the leader, it holds that

$$\frac{x^T(2\sigma - \overline{1})}{2}\left(q\left(-\frac{x^T\sigma}{2K}\right) - q\left(-\frac{x^T(\overline{1}-\sigma)}{2K}\right)\right) = 0.$$

Consequently, if the strategy is not uniform, the difference in the generator functions is nonzero and it is easy to show that the scalar product $x^T(2\sigma - \overline{1})/2$ never reaches zero, thus, making it impossible for a non-uniform strategy to be optimal. Therefore, to achieve utility u_l^*,the leader has to enforce a uniform strategy of the follower. Given that the leader has to commit to either s or $1-s$ in their upper information sets, we analyze the conditions when the follower is incentivized to play a uniform strategy. Let the set J be defined similarly as earlier: an item belongs to J if the first action in the leader's partition subtree is played with probability s. We have

$$u_f(a_1) = u_f(a_2) \iff (1-2s)\sum_{i\in[K]} x_i + (2s-1)\sum_{i\in[K]\setminus J} x_i = 0.$$

Because there is no J such that the sums are equal and because by the setting of the normal form game $s \neq 1-s$, the leader never simultaneously enforces optimal utility in the normal form game and the partition subtrees. Their utility is hence strictly smaller than u_l^*.

Together, it means that by analyzing the leader's utility in the equilibrium of $G_P(I)$ we are able to separate solvable and unsolvable instances of the partition problem.

Reduction to a general sum game

The situation in general sum games is even simpler. The structure of the proof is exactly as the proof for zero sum games above, but the role of the normal form subtree can be played by any cooperative coordination game, an example of which we depict in Figure 5.6. For any quantal response function, the follower plays the action with higher expected utility with a higher probability. The uniform strategy for the leader thus corresponds to the strict minimum of their utility achievable against any quantal opponent. Any other strategy will make the two actions of the follower have different expected utilities and hence the better will be played with probability

		Leader	
		o	r
Follower	A	1,1	0,0
	B	0,0	1,1

Figure 5.6: An example of a coordination normal form game we may use in the reduction.

more than a half, giving the leader better utility than the uniform strategy. Since the game is completely symmetric, it has two distinct quantal Stackelberg equilibria.

A similar argument holds also for the partition subtree, which stays unchanged from the zero sum game. In solvable instances, the leader's commitment makes any quantal player be indifferent and play uniformly. In case of unsolvable instance, one of their actions will be better and played with a strictly higher probability. This will give the follower more utility than the uniform strategy and hence it would be suboptimal for the leader. □

5.2 Dinkelbach-type equilibrium formulation

The non-concavity of formulation (5.1) makes it difficult to optimize over and guarantee global optimality. Therefore, we search for an alternative representation of quantal Stackelberg equilibrium that would express the problem as a single fractional criterion, instead of a set of equations. Such representation would allow us to leverage reformulation methods from fractional programming to eliminate the fraction. For this purpose we use the realization plans. Note that in formulation (5.1), the utility from each leaf is propagated up through the variables v by multiplying it by the values of a behavioral strategy and chance of all actions on the way to the root. Because the product of behavioral probabilities of a sequence is, by definition, equivalent to the realization of the sequence, the criterion of the formulation can be expressed as

$$\max_{r_l} \sum_{z \in \mathcal{Z}} u_l(z) C(z) r_l(z) QR(r_l, z), \tag{5.4}$$

where

$$QR(r_l, z) = \prod_{a \in seq_f(z)} \frac{q(e(r_l, a))}{\sum_{a' \in \chi(I(a))} q(e(r_l, a'))}.$$

Because realization plans are equivalent to behavioral strategies, and to facilitate exposition, instead of $e(\beta_l, a)$, we write the evaluation function as $e(r_l, a)$. The purpose of this reformulation is to express the criterion QSE-EFG as a single fraction, similarly as in security games[128] or in normal form games, explained in the previous chapter. Problems in this form can be then solved using the Dinkelbach's method for nonlinear fractional programming[129].

Theorem 5.3. *Maximizing the leader's expected utility in quantal Stackelberg equilibrium as in Formulation* (QSE-EFG) *is equivalent to finding a unique root of the following function* $\mathcal{D}$*:*

$$\mathcal{D}(p) = \max_{r_l} F(r_l, p), \tag{5.5}$$

where

$$F(r_l, p) = \sum_{\pi \in \Pi_f} \prod_{a \in \pi} q(e(r_l, a)) \left(\sum_{z \in Z(\pi)} u_l(z) C(z) r_l(z) - p \right).$$

[128] R. Yang, F. Ordonez, and M. Tambe. "Computing optimal strategy against quantal response in security games". In: *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems*. 2012, pp. 847–854.

[129] W. Dinkelbach. "On nonlinear fractional programming". *Management Science* 13:7, 1967, pp. 492–498.

Proof. We derive this formulation from quantal Stackelberg expressed as in Equation (5.4). Because the follower acts quantally in every information set, the smallest common multiple over all leafs is

$$\prod_{I \in I_f} \sum_{a \in \chi(I)} q(e(r_l, a)).$$

The fractional representation of the equilibrium is hence

$$\max_{r_l} \frac{\sum_{z \in Z} u_l(z) C(z) r_l(z) Q(z, r_l)}{\prod_{I \in I_f} \sum_{a \in \chi(I)} q(e(r_l, a))},$$

where

$$Q(z, r_l) = \prod_{\substack{a \in seq_f(z), I \in I_f \\ seq(I) \not\sqsubseteq seq(z)}} q(e(r_l, a)) \sum_{a \in \chi(I)} q(e(r_l, a)).$$

Because $Q(z, r_l)$ is a sum of products of functions q applied to a fixed path from root to z and one action in each information set outside this path, it iterates over pure strategies enabling to reach z. $Q(z, r_l)$ is therefore equivalent to

$$\sum_{\pi \in \Pi_f : z \in Z(\pi)} \prod_{a \in \pi} q(e(r_l, a)).$$

Applying the same idea also for the denominator and swapping the sum over leafs with the sum over pure strategies in the nominator we obtain

$$\max_{r_l} \frac{\sum_{\pi \in \Pi_f} \prod_{a \in \pi} q(e(r_l, a)) \sum_{z \in Z(\pi)} u_l(z) C(z) r_l(seq_l(z))}{\sum_{\pi \in \Pi_f} \prod_{a \in \pi} q(e(r_l, a))}. \tag{5.6}$$

By the Dinkelbach reformulation, maximizing this equation is equivalent to finding a root of

$$\max_{r_l} \sum_{\pi \in \Pi_f} \prod_{a \in \pi} q(e(r_l, a)) \sum_{z \in Z(\pi)} u_l(z) C(z) r_l(z) - p \sum_{\pi \in \Pi_f} \prod_{a \in \pi} q(e(r_l, a)),$$

which is the desired equation. □

Due to the leader's strategy being represented using a realization plan, the formulation (5.5) has $|\Sigma_l|$ variables. The expression $e(r_l, a)$ is evaluated for every follower's action in the game tree, the number of evaluations is hence also linear in $|I_f|$. The outer sum, however, enumerates the follower's pure strategies and is thus exponential in $|I_f|$. In many real world applications this fact might not be critical, as the follower's strategy space is often much smaller than a combinatorial strategy space of the leader[130].

[130] For instance, in sequential variants of some security games the follower may be choosing a single target to attack, while the leader has to deploy or move multiple units over a large strategic map.

Algorithm 2: Dinkelbach-type algorithm for approximating the optimal leader's strategy of quantal Stackelberg equilibrium in extensive form games

$UB \leftarrow \max_{z \in \mathcal{Z}} u_l(z), LB \leftarrow \min_{z \in \mathcal{Z}} u_l(z)$
$r_l^* \leftarrow \arg\max_{r_l} F(r_l, LB)$
repeat
  $p \leftarrow (UB - LB)/2$
  $v \leftarrow \max_{r_l} F(r_l, p)$
  $r_l^p \leftarrow \arg\max_{r_l} F(r_l, p)$
  if $v < 0$ **then** $LB \leftarrow p, r_l^* \leftarrow r_l^p$ **else** $UB \leftarrow p$
until $UB - LB < \epsilon$
return r_l^*

Because function D is convex, its root can be found using a binary search method, as described in Algorithm 2. We refer to formulation (5.5) as to the Dinkelbach subproblem of the Dinkelbach formulation of quantal Stackelberg equilibrium in extesnive form games. Algorithm 2 iteratively updates the upper bound (UB) and lower bound (LB) on the value of QSE according to a binary search method for finding a root of a function[131]. For running the binary search it is essential to solve the Dinkelbach subproblems[132]. The following proposition presents conditions under which the subproblem can be efficiently approximated.

[131] Note that this binary search approach generalizes a similar technique used for computing quantal Stackelberg equilibrium in restriction to logit generators in security games or its more general formulation in normal form games, described in the previous chapter.

[132] More specifically, we need to evaluate $\mathcal{D}(p)$ for any p.

Proposition 5.2. *Let q be a twice differentiable generator of a generalized Luce model and e be twice differentiable evaluation function of the follower. The Dinkelbach subproblem for $p \in [\min_{z \in \mathcal{Z}} u_l(z), \max_{z \in \mathcal{Z}} u_l(z)]$ is concave if for any $\pi \in \Pi_f, a \in \pi$ and realization plan r_l, the matrix*

$$\delta(\pi, a) = q'(e_a)(u_{r_l} e'^T_a + e'_a u^T_{r_l}) + (u^T_{r_l} r_l - p)\Big(e''_a q'(e_a) + \\ + {e'_a}^2 q''(e_a) + \sum_{a' \neq a \in \pi} e'_a e'_{a'} q'(e_a) q'(e_{a'}) \prod_{a'' \neq a' \neq a \in \pi} q(e_{a''})\Big)$$

is negative semidefinite, where $e_a = e(r_l, a)$, $e'_a = e'(a, r_l)$, $e''_a = e''(a, r_l)$, and $u^T_{r_l} r_l = \sum_{z \in \mathcal{Z}(\pi)} u_l(z) C(z) r_l(z)$.

Proof. The formulation of the subproblem from function (5.5) is concave when its Hessian matrix is negative semidefinite. The Hessian matrix is of a form

$$\sum_{\pi \in \Pi_f} \sum_{a \in \pi} \delta(\pi, a) \prod_{a' \neq a \in \pi} q(e_{a'}).$$

The result follows from the fact that negative semidefiniteness is preserved under summation and multiplication by a positive number, and the generator of the Luce model is always positive. □

In case the conditions are met[133], local-optimization algorithms akin to the projected gradient ascent, given function (5.5) is L-smooth, are guaranteed to reach the optimum in polynomial time.[134] However, we consider the differentiability of the evaluation to be the most limiting factor for using this theoretical result in practical computations. Intuitively, many evaluation functions should be formulated as non-smooth maximization problems[135], which do not satisfy the assumptions of the characterization. This is why a significant portion of the later text focuses instead on a linear approximation of the Dinkelbach formulation, which applies to any two-player EFG and is generalizable to any linearizable evaluation function.

5.3 Approximating the Dinkelbach subproblem

In case a game does not satisfy the conditions in Proposition 5.2, the guarantee of convergence is lost. A solution commonly suggested in the literature is then to linearize the criterion (5.5) and transform the problem into a mixed integer linear program that can be solved using standard methods. For linearizing the criterion we need to approximate both the quantal generator q and the evaluation function e.

We begin by approximating the quantal generator q. We focus on logit quantal response, which is the most commonly studied quantal response in the literature. In this case the function q is defined as $q(x) = exp(\alpha x), \alpha \in \mathbb{R}^+$. The player becomes more rational as α approaches infinity. We can express the product of generator functions from formulation (5.5) through a substitutional variable x_π as

$$\prod_{a\in\pi} exp(\alpha e(r_l, a)) = exp\left(\alpha \sum_{a\in\pi} e(r_l, a)\right) \to x_\pi.$$

The exp function is linearizable into a piece-wise function $\overline{exp}$ with K segments as

$$\overline{exp}\left(\alpha \sum_{a\in\pi} e(r_l, a)\right) = \sum_{k=0}^{K} \alpha^k t_\pi^k + exp(\underline{e}) \to \overline{x}_\pi, \tag{5.7}$$

where $(\alpha^k)_{k\in[K]}, \alpha^k \in \mathbb{R}$ is a slope of the k-th segment, subjected to constraints

$$\begin{aligned} \sum_{k=0}^{K} t_\pi^k &= \sum_{a\in\pi} e(r_l, a) - \underline{e} \\ t_\pi^k &\le z_\pi^k \frac{(\overline{e} - \underline{e})}{K} \le t_\pi^{k+1} \\ 0 \le t_\pi^k &\le \frac{\overline{e} - \underline{e}}{K}, \quad z_\pi^k \in \{0, 1\}, \end{aligned} \tag{5.8}$$

where the binary variables z indicate whether the linear segment is used and the real variables t define what portion of the segment is active. Moreover, the constants $\underline{e}$ and $\overline{e}$ are defined as

[133] In case the generator and the evaluation function are differentiable, the condition can be checked via semidefinite programming in polynomial time. A simple setup satisfying the condition is also when q and e are linear and close to zero, and the utilities are negatively correlated, as in zero sum games.

[134] Y. Nesterov. *Introductory lectures on convex optimization: A basic course*. 1st ed. Applied Optimization 87. Springer US, 2004.

[135] For example, the utility of a best response is one such problem.

$$\underline{e} = \alpha|I_f| \min_{a \in A_f, r_l} e(r_l, a), \text{and}$$
$$\overline{e} = \alpha|I_f| \max_{a \in A_f, r_l} e(r_l, a).$$

The maximum difference in values of exp and its linearization $\overline{exp}$ using K segments on the interval $[\underline{e}, \overline{e}]$ can be bounded using standard methods as[136]:

$$|exp(x) - \overline{exp}(x)| \leq exp(\overline{e}) \frac{(\overline{e} - \underline{e})^2}{8K^2}, \quad x \in [\underline{e}, \overline{e}]. \tag{5.9}$$

With a linearized logit generator, the Dinkelbach subproblem (5.5) is expressed as

$$\mathcal{D}(p) = \max_{r_l} \left(\sum_{z \in Z(\pi)} u_l(z) C(z) r_l(z) - p \right) \overline{x}_\pi.$$

Clearly, the criterion contains multiple bilinear terms $\overline{x}_\pi r_l(z)$. For linearizing the bilinear terms, we use the MDT technique[137]. MDT is a parametrizable method which enables controlling the error in exchange of introducing binary variables. The product $c(\pi, z) = \overline{x}_\pi r_l(z)$ is expressed using linear equations

$$\begin{aligned} c(\pi, z) &= \sum_{i=0}^{b-1} \sum_{j \in \mathcal{E}} i b^j r_{i,j}, \quad \overline{x}_\pi^{\mathcal{E}} = \sum_{i=0}^{b-1} \sum_{j \in \mathcal{E}} i b^j s_{i,j}, \\ 1 &= \sum_{i=0}^{b-1} s_{i,j}, \qquad r_l(z) = \sum_{i=0}^{b-1} r_{i,j}, \qquad \forall j \in \mathcal{E} \\ s_{i,j} &\in \{0, 1\}, \qquad 0 \leq r_{i,j} \leq s_{i,j}, \qquad \forall i \in [b-1], \forall j \in \mathcal{E} \end{aligned} \tag{5.10}$$

where $\mathcal{E} \subset \mathbb{Z}$ is a finite subset controlling the error of the approximation with basis b. $\overline{x}_\pi^{\mathcal{E}}$ is a representation of $\overline{x}_\pi$ in MDT over $\mathcal{E}$. The following lemma identifies the approximation error introduced by the selection of K and $\mathcal{E}$.

Lemma 5.1. *Let $|\mathcal{E}| = L$ and x_π be linearized with K segments. Then the linearization error is bounded as*

$$|x_\pi r_l(z) - \overline{x}_\pi^{\mathcal{E}} r_l(z)| \leq \epsilon_K + \epsilon_{\mathcal{E}}, \tag{5.11}$$

where the epsilons are set as

$$\epsilon_K = exp(\overline{e}) \frac{(\overline{e} - \underline{e})^2}{8K^2}, \text{ and}$$
$$\epsilon_{\mathcal{E}} = \max(N, exp(\overline{e}) - b^{M+1} + N, N - exp(\underline{e})),$$

and the constants M and N as $M = \lfloor \log_b(exp(\overline{e})) \rfloor$ and $N = b^{M-L+1}$.

[136] M. Yano, J. D. Penn, G. Konidaris, and A. T. Patera. *Math, numerics & programming (for mechanical engineers)*. MIT Press, 2013.

[137] S. Kolodziej, P. M. Castro, and I. E. Grossmann. "Global optimization of bilinear programs with a multiparametric disaggregation technique". *Journal of Global Optimization* 57:4, 2013, pp. 1039–1063.

Proof. Let $\mathcal{E} = \{M{-}L{+}1, M{-}L{+}2, \ldots, M\}$. $\mathcal{E}$ hence defines a discretization of variable $\overline{x}_\pi$ on interval $[N, b^{M+1} - N]$ with a step of size N. Because $\overline{x}_\pi$ is defined on interval $[exp(\underline{e}), exp(\overline{e})]$, the maximum difference between $\overline{x}_\pi$ and $\overline{x}^{\mathcal{E}}_\pi$ is $\epsilon_{\mathcal{E}}$. As the realization variables r_l are always at most 1, we have

$$|x_\pi r_l(z) - \overline{x}^{\mathcal{E}}_\pi r_l(z)| \leq |x_\pi - \overline{x}_\pi| + |\overline{x}_\pi - \overline{x}^{\mathcal{E}}_\pi|.$$

The Inequality (5.9) then implies $|x_\pi - \overline{x}_\pi| \leq \epsilon_K$, concluding the proof. □

Now we move to the linearization of the evaluation function e. We present a domain independent formulation of a common situation when the follower is not aware of their subrationality and evaluates the actions in the current information set on the basis of acting rationally in the subsequent information sets, weighted by the probability of reaching the current set. In that case, the evaluation function e can be expressed as

$$\begin{aligned} e(r, a) &= v(I(a)) - s(seq_f(I(a))a), \\ v(inf_f(\sigma_f)) &= s(\sigma_f) + \sum_{\substack{I \in I_f \\ seq_f(I) = \sigma_f}} v(I) + \sum_{\substack{z \in Z \\ seq_f(z) = \sigma_f}} u_l(z) C(z) r_l(z), \quad \forall \sigma_f \in \Sigma_f \\ 0 &\leq s(\sigma) \leq M(1 - r_f(\sigma)), \qquad \forall \sigma \in \Sigma_f \end{aligned} \tag{5.12}$$

where r_f is the binary best response realization plan of the follower, v is the optimal expected utility contribution in an information set and s is a slack variable compensating the deficiency in action's suboptimal utility. Now, we can finally state the approximation error for computing the quantal Stackelberg equilibrium.

Proposition 5.3. *Consider a linearization of the Dinkelbach subproblem*

$$\mathcal{D}(p) = \max_{r_l} \sum_{z \in Z(\pi)} u_l(z) C(z) c(\pi, z) - p\overline{x}^{\mathcal{E}}_\pi,$$

with constraints (5.8), (5.10), (5.12) *with K segments,* $|\mathcal{E}| = L$, *substitution* (5.7) *and the realization plan constraints from Definition 3.5. Let* r^*_l *be a realization plan computed by Algorithm 2 with precision* ϵ_B, *and* r^{QSE}_l *be a realization plan of the leader in quantal Stackelberg equilibrium. Then for the utility difference*

$$d = |u_l(r^*_l, QR(r^*_l)) - u_l(r^{QS}_l, QR(r^{QS}_l))|$$

it holds that

$$d \leq \epsilon_B + \frac{\overline{u}_l |\Pi_f| \epsilon_K + |\Pi_f| \max_{z \in \mathcal{Z}} |u_l(z)| (\epsilon_K + \epsilon_{\mathcal{E}})}{exp(\underline{e})},$$

where ϵ_K *and* $\epsilon_{\mathcal{E}}$ *are defined as in Lemma 5.1.*

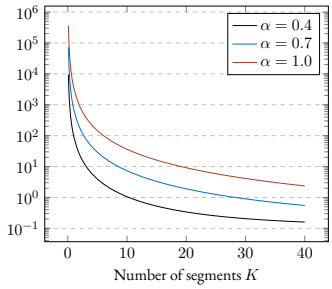


Figure 5.7: An example of the upper bound derived in Proposition 5.3 for a game with a logit generator with α in $\{0.4, 0.7, 1.0\}$, utility ranges $[-10, 10]$ and $[0, 1]$ for the leader and the follower, respectively, 3 follower's information sets with 2 actions each, $b = 3$, $L = 15$, and $\epsilon_B = 0.1$.

Proof. We proceed similarly as in Proposition 4.6. However, we derive specific bounds for the Dinkelbach representation of the equilibrium in extensive form games that are different from their counterparts in normal form or security games. First, we express the fractional criterion function (5.6) of the direct formulation of the equilibrium as $QSE(r_l) = u_l(r_l, QR(r_l)) = N(r_l)/D(r_l)$, where

$$\begin{aligned} N(r_l) &= \sum_{\pi \in \Pi_f} \sum_{z \in Z} u_l(z) exp(y_\pi) r_l(seq_l(z)), \\ D(r_l) &= \sum_{\pi \in \Pi_f} exp(e_\pi), \text{ and} \\ e_\pi &= \alpha \sum_{a \in \pi} e(r_l, a) \quad \forall \pi \in \Pi_f. \end{aligned}$$

The linearized criterion function is then $\overline{QSE}(r_l) = \overline{N}(r_l)/\overline{D}(r_l)$, where

$$\begin{aligned} \overline{N}(r_l) &= \sum_{\pi \in \Pi_f} \sum_{z \in \mathcal{Z}} u_l(z) \overline{x_\pi r_l(seq_l(z))}, \\ \overline{D}(r_l) &= \sum_{\pi \in \Pi_f} \overline{x}_\pi^{\mathcal{E}}, \text{ and} \\ \overline{x}_\pi^{\mathcal{E}} &= \overline{exp}(e_\pi), e_\pi = \alpha \sum_{a \in \pi} e(r_l, a) \quad \forall \pi \in \Pi_f. \end{aligned}$$

Denote $\delta^{\overline{QSE}}$ the optimal solution of $\max_{\delta \in \Delta_l} \overline{QSE}(\delta)$. Then we have

$$\begin{aligned} d = \left| \frac{N(\delta^{QSE})}{D(\delta^{QSE})} - \frac{N(\delta^*)}{D(\delta^*)} \right| \leq \left| \frac{N(\delta^{QSE})}{D(\delta^{QSE})} - \frac{\overline{N}(\delta^{\overline{QSE}})}{\overline{D}(\delta^{\overline{QSE}})} \right| \\ + \left| \frac{\overline{N}(\delta^{\overline{QSE}})}{\overline{D}(\delta^{\overline{QSE}})} - \frac{\overline{N}(\delta^*)}{\overline{D}(\delta^*)} \right| + \left| \frac{\overline{N}(\delta^*)}{\overline{D}(\delta^*)} - \frac{N(\delta^*)}{D(\delta^*)} \right|. \end{aligned} \tag{5.13}$$

We bound each term separately. For the first and the last term we use Lemma 7 of Yang, Ordonez, and Tambe[128]. By this lemma, for each $\delta \in \Delta_l$, we have

$$\left| \frac{\overline{N}(\delta)}{\overline{D}(\delta)} - \frac{N(\delta)}{D(\delta)} \right| \leq \frac{1}{\overline{D}(\delta)} \left(\frac{N(\delta)}{D(\delta)} d_D + d_N \right), \tag{5.14}$$

where $d_D = \left| D(\delta) - \overline{D}(\delta) \right|$ and $d_N = \left| N(\delta) - \overline{N}(\delta) \right|$. Using Lemma 5.1, we bound the differences as

$$\begin{aligned} d_D &\leq |\Pi_f| \epsilon_K \to C_1 \\ d_N &\leq |\Pi_f| \max_{z \in \mathcal{Z}} |u_l(z)| (\epsilon_K + \epsilon_{\mathcal{E}}) \to C_2. \end{aligned}$$

Because $\overline{D}(\delta) \geq exp(\underline{e})$, and $QSE(\delta) \leq \overline{u}_l$, together with the Inequality (5.14) we conclude that

$$\left| \frac{\overline{N}(\delta)}{\overline{D}(\delta)} - \frac{N(\delta)}{D(\delta)} \right| \leq \frac{\overline{u}_l C_1 + C_2}{exp(\underline{e})}. \tag{5.15}$$

By the same argument as in Proposition 4.6 the first term of Inequality (5.13) shares the bound with the last term, and the second term is bounded by ϵ_B. The result follows from combining the bounds. □

5.4 Empirical evaluation

We compare the Dinkelbach-type algorithm (DTA) to the standard benchmark for nonlinear optimization: the COBYLA algorithm,[138] implemented in the open-source NLOPT library. COBYLA is a gradient-free algorithm capable of handling linear equality constraints induced by realization plans. We opted for COBYLA because the follower's evaluation function is non-differentiable, possibly in infinitely many points. We apply COBYLA directly to formulation (5.4). For evaluating the algorithms, we used two domains: a variant of search game commonly used to evaluate algorithms for SE[139] and a network game, handcrafted to be difficult for quantal Stackelberg equilibrium.

[138] M. J. Powell. "A view of algorithms for optimization without derivatives". *Mathematics Today-Bulletin of the Institute of Mathematics and its Applications* 43:5, 2007, pp. 170–174.

[139] J. Čermák, B. Bošanský, K. Durkota, V. Lisý, and C. Kiekintveld. "Using correlated strategies for computing Stackelberg equilibria in extensive-form games". In: *Proceedings of the 30th AAAI Conference on Artificial Intelligence*. 2016, pp. 439–445; C. Kroer, G. Farina, and T. Sandholm. "Robust Stackelberg equilibria in extensive-form games and extension to limited lookahead". In: *Proceedings of 32nd AAAI Conference on Artificial Intelligence*. 2018.

5.4.1 Experimental domains and their instance generation

We assume the defender acts as a leader, while the attacker assumes the follower's role. We consider three exponential generators of generalized Luce models, $\alpha \in \{0.4, 0.7, 1.0\}$. The tolerance parameter for the COBYLA algorithm in NLOPT was set to 10^{-2} and $\epsilon_B = 1\%$ of the leader's utility range for the DTA's binary search. The linearization uses $K = 3$, the basis of MDT is set to $b = 3$ and the size of the precision interval $\mathcal{E}$ is $L = 4$. For each combination of game size $\times$ generator function, we constructed 20 instances. All implementations were done in C++17. We used NLOPT 2.6.1, and a single-threaded IBM CPLEX 12.8 carried out all computations of solutions of mixed integer linear programs. The experiments were performed on a 3.2GHz CPU with 16GB RAM.

Search game

The game is played on a directed graph, depicted on the left in Figure 5.8. The attacker's goal is to reach one of the destination nodes (D0 – D7) from the starting node (S), while the defender aims to catch the attacker with one of the two units operating in the marked areas of the graph (P0 and P1). The attacker receives a different reward for reaching a different destination node (the reward is selected randomly from interval $[0, 2]$). While the defender can move freely with unit P1, unit P0 is static – placed by the defender at the beginning of the game. If the attacker evades unit P0, the defender is given N steps to set unit P1. The defender receives a signal if P1 is within 1 step from the attacker. In case the defender captures the

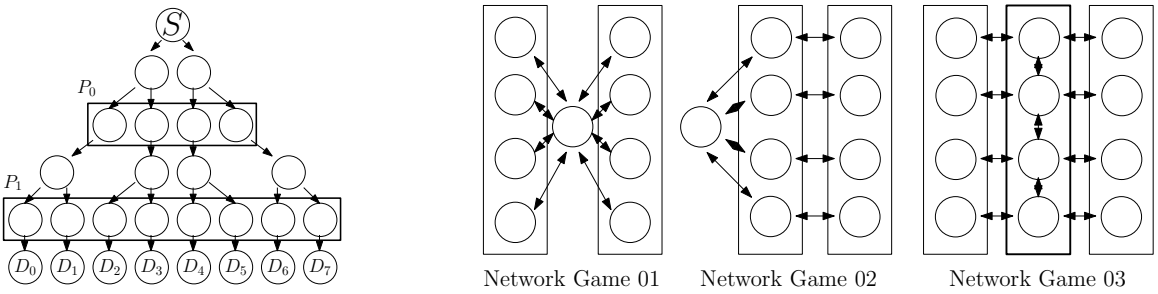


Figure 5.8: Graphs for the experimental domains. (Left) Search game, and (Right) network games.

attacker, the defender gets a positive reward of $1 - n/(N+1)$, where $n \leq N$ is a number of taken steps, and the attacker receives 0. We consider a version of the game in which the attacker perceives no information about the whereabouts of the defender's units, and unit P1 starts above D_0.

Network game

The network game is played on a directed graph too. Some nodes in the graph are grouped together into mutually disjunctive areas. In the beginning, the attacker observes their areas of possible infiltration and the area the defender uses to enter the network, and selects one area for further probing. The defender then picks a node from the entering area, while the attacker chooses a node from their area to compromise. The defender is given a maximum number of steps N to survey the network and discover the attacker. There is binary information: if the attacker is located within γ steps from the defender, the defender observes it. If she captures the attacker in $n \leq N$ steps she receives a utility $1 - n/(N+1)$, 0 otherwise. The attacker is given a reward associated with the compromised node if not found (the reward is chosen randomly from interval $[-2.5, 2.5]$–the negative utility represents nodes with compromising costs higher than the data's price or a possible honeypot), -3 otherwise. We designed three networks, shown on the right in Figure 5.8. Attacker's areas are depicted in thin-lined rectangles. The defender's entering area is depicted in a thick-lined rectangle in game 03 and selected randomly (4 nodes per seed) in other games. We set $\gamma = 0$ for game 01, 1 otherwise. It is a type of coordination game; hence, the vast majority of the attacker's strategies can be best responses to the defender's strategy, which makes computing Stackelberg strategies particularly difficult.

5.4.2 Experimental results

In Figures 5.9 and 5.10, the x-axis varies the game size, while the y-axis shows the runtimes of the algorithms. Every point in the graphs corresponds to the mean over the sampled instances and shows the achieved standard error. We terminated all running seeds after 24h and depict them in the graphs with this lower bound on

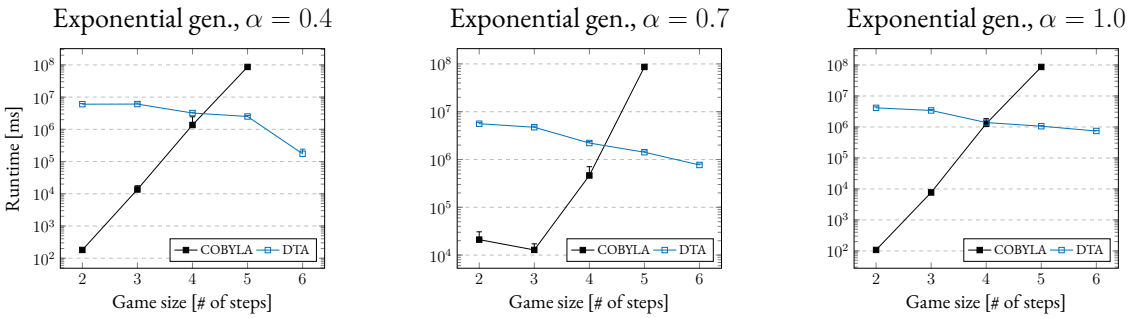


Figure 5.9: Mean runtimes of computing the approximations of quantal Stackelberg equilibria using COBYLA and Dinkelbach type algorithm in the search game in extensive form. Every point shows also a standard error.

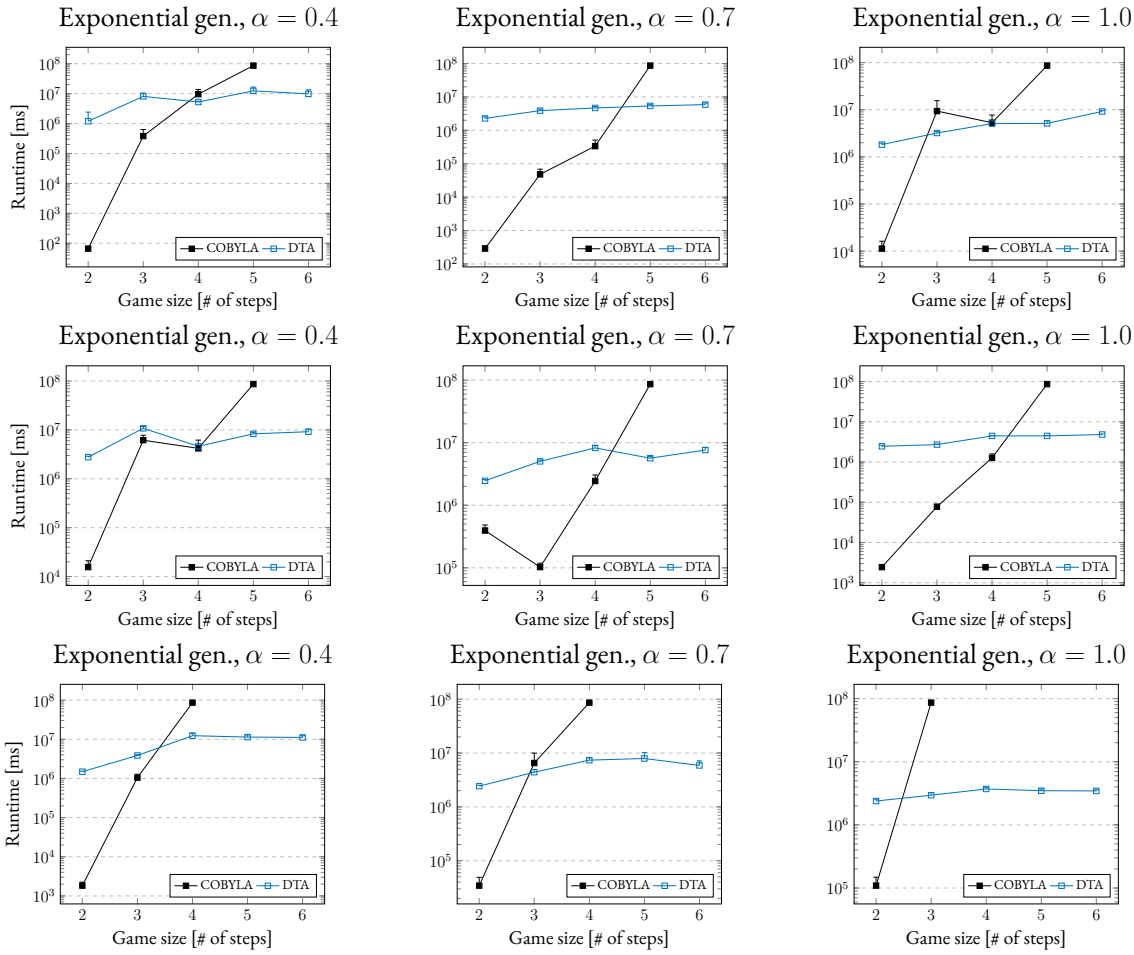


Figure 5.10: Mean runtimes of computing the approximations of quantal Stackelberg equilibria using COBYLA and Dinkelbach type algorithm in the three network games: (Top) game 01, (Middle) game 02, and (Bottom) game 03. Every point shows also a standard error.

	2 steps	3 steps	4 steps
Search game			
$\alpha = 0.4$	-2.13% ± 0.29%	-0.20% ± 0.42%	16.23% ± 3.77%
$\alpha = 0.7$	-7.24% ± 1.00%	-4.75% ± 0.79%	18.24% ± 3.14%
$\alpha = 1.0$	-21.24% ± 1.81%	-10.99% ± 1.64%	-1.98% ± 4.91%
Network game 01			
$\alpha = 0.4$	0.23% ± 0.03%	1.27% ± 0.20%	2.83% ± 0.46%
$\alpha = 0.7$	-1.99% ± 0.38%	0.42% ± 0.38%	3.84% ± 0.74%
$\alpha = 1.0$	-3.82% ± 0.51%	0.50% ± 0.50%	2.32% ± 0.92%
Network game 02			
$\alpha = 0.4$	1.17% ± 0.27%	1.16% ± 0.44%	13.32% ± 7.35%
$\alpha = 0.7$	0.20% ± 0.21%	1.80% ± 0.42%	8.87% ± 3.47%
$\alpha = 1.0$	-2.35% ± 0.35%	-0.22% ± 0.49%	3.74% ± 4.05%
Network game 03			
$\alpha = 0.4$	1.09% ± 0.27%	1.30% ± 0.42%	-
$\alpha = 0.7$	1.02% ± 0.45%	5.73% ± 1.85%	-
$\alpha = 1.0$	0.25% ± 0.68%	8.33% ± 4.94%	-

Table 5.1: Comparison of solution quality. The table shows mean deviations of solutions computed using COBYLA and Dinkelbach type algorithm in both search and network games. A positive value indicates that Dinkelbach type algorithm returned better solution than COBYLA. Every value shows also a standard error.

runtime if the computation was still ongoing. As the figures show, despite the overhead of the Dinkelbach type algorithm on smaller instances, it scales significantly better than COBYLA. For 6 steps, we ran longer jobs, and COBYLA computed no game within 3 days. Interestingly, as in some other NP-hard problems,[140] increasing the search game's strategy space enables finding better strategies faster, and binary search terminates earlier.

The relative errors of computed solutions are presented in Table 5.1. The values correspond to the mean ratio of the difference in the defender's expected utility computed using COBYLA and the DTA to the length of the defender's utility range in the game. Due to linear approximations used by COBYLA, it can find close-to-optimal solutions in smaller instances. In some cases, its solution is even better than that of DTA because of the algorithm's approximation parameters. However, as the table reveals, the quality of COBYLA's solutions degrades with increasing game size, reaching error of 18.24% for 4 steps in the search game.

[140] P. Cheeseman, B. Kanefsky, and W. M. Taylor. "Where the really hard problems are". In: *Proceedings of the 12th International Joint Conference on Artificial Intelligence*. Morgan Kaufmann Publishers Inc., Sydney, New South Wales, Australia, 1991, pp. 331–337.

5.5 Summary of contributions

In this chapter we studied quantal Stackelberg equilibrium – a strategy the rational player should commit to against a boundedly rational player – in extensive form

games. We show that computing the equilibrium is NP-hard; still, quantal Stackelberg equilibrium is useful for evaluating scalable heuristics or improving the understanding of human decision-making in experiments with human participants. We introduce the first practical algorithm for computing the equilibrium in extensive form games and show that contrary to direct formulation, our algorithm solves larger games with smaller errors.

Part III

Coordination

6 Quantal Correlated Equilibrium in Normal Form Games

Robert Aumann introduced the concept of correlated equilibrium in 1974 as a generalization of Nash equilibrium. Throughout the years, correlated equilibrium became one of the most prominent concepts in game theory, because of its suitability for studying coordinated multiagent systems,[141] as well as appealing computational complexity that is provably polynomial. Signaling has also applications in physical or cyber security, for example, fighting misinformation.[142] Besides, many of the contemporary breakthroughs in Nash equilibrium computation in two-player zero sum games or even multiplayer games were achieved through uncoupled no-regret methods originally designed to approximate correlated equilibrium in general-sum games. Recent years have witnessed an increasing attention to correlated strategies also in sequential games, either in the form of (coarse) correlated equilibrium or team-maxmin equilibrium with coordination device. With the new approaches, even large scenarios are solved in a matter of days.

Despite the favorable scalability of contemporary state-of-the-art approaches, one of the fundamental limitations that hinders applications of game-theoretic models in real world remains their assumption of perfect rationality of players. Numerous deployments of concepts related to leader-follower equilibria proved that accounting for the imperfect decision-making of human players helps to avoid unnecessary utility losses and leads to substantially improved performance, as we explained in the introductory chapters. Perhaps the first attempt to introduce "boundedly rational" behavior into correlated equilibrium was through trembling-hand perfection, and it was studied in both normal form[143] and extensive form[144] games. Trembling-hand perfection amends the inherent issues associated with perfectly rational concepts akin to Nash equilibrium through prescribing optimal strategies even when the game play strays off the equilibrium path. Yet, this does not translate into correct predictions of human behavior. To this end, quantal response is much more suitable. Despite its importance, to the best of our knowledge, no work that studies correlated quantal response strategies has ever been published.

In this chapter, we investigate the amalgamation of quantal response and correlated equilibrium using the generalized Luce models of quantal behavior. We provide two possible definitions of quantal correlated equilibrium, inspired by the standard way of constructing a correlated equilibrium: we replace a best-response condition with quantal response, enforcing quantal behavior either per each signal

The results in this chapter were published as J. Černý, B. An, and A. N. Zhang. "Quantal correlated equilibrium in normal form games". In: *Proceedings of the 23rd ACM Conference on Economics and Computation*. 2022, pp. 210–239.

[141] I. Ashlagi, D. Monderer, and M. Tennenholtz. "On the value of correlation". *Journal of Artificial Intelligence Research* 33, 2008, pp. 575–613.

[142] O. Candogan and K. Drakopoulos. "Optimal signaling of content accuracy". *Operations Research* 68:2, 2020, pp. 497–515.

[143] A. Dhillon and J. F. Mertens. "Perfect correlated equilibria". *Journal of Economic Theory* 68:2, 1996, pp. 279–302.

[144] A. Marchesi and N. Gatti. "Trembling-hand perfection and correlation in sequential games". In: *Proceedings of the 35th AAAI Conference on Artificial Intelligence*. 2021, pp. 5566–5574.

separately or over the whole set of strategies. Our formulation is general enough to model individual differences in quantal behavior between the players, which other works rarely consider. Our first motivation is to study if these quantal counterparts of correlated equilibrium satisfy the intuitive requirements of such an equilibrium. Indeed, we show that every quantal response equilibrium is quantal correlated, and the traditional correlated equilibrium is reached in the limit as quantal responses approach a best response. The space of all equilibria is compact, and in case the equilibrium is unique for all correlation devices, it is also connected. We conclude our initial analysis by showing that the concept remains PPAD-hard.

Next, we formulate a robust homotopic algorithm capable of traversing the principal branch of equilibrial correspondence. We employ carefully designed variable substitutions and model reformulations that ameliorate numerical issues caused by steep quantal response functions or wide ranges of utilities. As a consequence, we are able to simultaneously trace the equilibrium and gradiently optimize a probability distribution over the signals while maintaining the homotopy's convergence guarantees. Finally, in the last part of this chapter, we investigate the algorithm's scalability and quality of solutions using two experimental domains: random games and supply chain games. Supply chains constitute a prime application domain for quantal correlated equilibrium: a setting where a central authority aims to coordinate the retailers to streamline the economy, but is only capable of sending signals because of the retailers' autonomy[145]. The results indicate that the homotopy approach provides high-quality solutions while being several orders of magnitude faster than BARON, a state-of-the-art non-convex optimization solver.

[145] For instance, the authority might represent a local government and the signals may have a form of specific taxation policies.

6.1 Problem definition

We consider generalized Luce quantal response functions and formulate two representations of quantal correlated equilibrium. We focus on the standard construction of correlated equilibrium because the canonical representation does not make sense in case the players are assumed to always play quantally and are hence incapable of following a single recommended action. First, we assume the players behave according to their generalized Luce quantal response functions with generators $(q_i)_{i\in N}$ after they receive a signal. We refer to this form of quantal correlated equilibrium as the per-signal equilibrium. Inserting bounded rationality into the model is simple using the standard construction.

Definition 6.1. *Let $G = (N, A, S, u)$ be a signaling game. The behavioral strategies $(\beta_i)_{i\in N}, \beta_i \in B_i$ and a signaling scheme $\lambda \in \Lambda$ form a per-signal quantal correlated equilibrium if*

$$\begin{aligned} u_i(a_i|s_i) &= \sum_{a_{-i}\in A_{-i}} \sum_{s_{-i}\in S_{-i}} \lambda(s_i, s_{-i})\beta_{-i}(a_{-i}|s_{-i})u_i(a_i, a_{-i}), \\ \beta_i(a_i|s_i) &= \frac{q_i(u_i(a_i|s_i))}{\sum_{a'_i\in A_i} q_i(u_i(a'_i|s_i))}. \end{aligned} \tag{S-QCE-NFG}$$

Player 2 \ Player 1	A	B
o	2,11	0,1
r	0,0	10,2

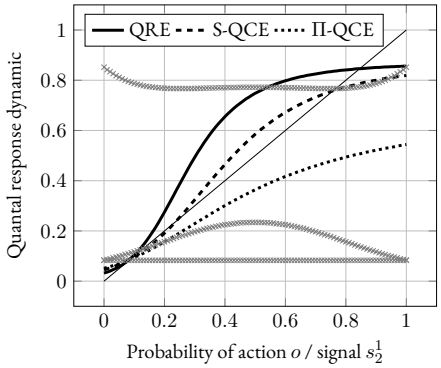


Figure 6.1: Equilibria in a variant of the Battle of Sexes game. (Left) The payoff matrix of the game. The figure follows a standard denotation of normal form games. (Right) Three quantal dynamics: quantal response, per-signal correlated, and over-pure-strategies correlated. The players share the same generator $q(x) = (x + (119 + \sqrt{401937})/332)^3$, and receive either one or two possible signals (each with probability 0.5), respectively. Equilibria are the points intersecting the quadrant's axis.

Example 6.1. *Consider the game in Figure 6.1, where the first player always receives a single signal while the second player conditions their strategy on one of two possible signals received with an equal probability. Both players act using a generator* $q(x) = (x + (119 + \sqrt{401937})/332)^3$. *The interaction dynamic of a per-signal quantal correlated equilibrium is depicted in the graph on the right using a thick dashed line. There are three per-signal equilibria, for* $\beta_1(o) \in \{0.0833, 0.233, 0.772\}$.

In our second formulation, instead of playing quantally for each signal independently, we assume the players act quantally over the whole set of pure strategies in the extended game. We call this formulation the over-pure-strategies equilibrium.

Definition 6.2. *Let* $G = (N, A, S, u)$ *be a signaling game. The mixed strategies* $(\delta_i)_{i\in N}, \delta_i \in \Delta_i$ *and a signaling scheme* $\lambda \in \Lambda$ *form an over-pure-strategies quantal correlated equilibrium if*

$$u_i(\pi_i) = \sum_{\pi_{-i}\in\Pi_{-i}} \sum_{(a,s)\in\times(\pi_i,\pi_{-i})} \lambda(s_i, s_{-i})\delta_{-i}(\pi_{-i})u_i(a_i, a_{-i}),$$
$$\delta_i(\pi_i) = \frac{q_i(u_i(\pi_i))}{\sum_{\pi_i'\in\Pi_i} q_i(u_i(\pi_i'))}. \qquad \text{(Π-QCE-NFG)}$$

Example 6.2. *Figure 6.1 shows an over-pure-strategies quantal correlated dynamic as well, using a thick dotted line. Similarly as in per-signal equilibrium, we consider a setting with one and two equally possible signals per player. The game contains a single over-pure-strategies equilibrium for* $\delta_1(o) = 0.0833$.

If we compare both formulations, we realize that while each player selects one distribution that describes their complete behavior after receiving any signal in formulation (Π-QCE-NFG), the formulation (S-QCE-NFG) assumes that players choose

the strategy distribution for each signal separately. Consequently, after observing a signal, they may act independently on other signals. The psychological studies show that humans prefer such short-term, delayed heuristic decisions[146] over long-term, premeditated decisions. This behavior arises especially in conflicts[147] or when facing information overload caused by large decision space[148]. In the light of that, formulation (S-QCE-NFG) seems more natural than its counterpart (Π-QCE-NFG). For this reason, we will prefer the formulation (S-QCE-NFG) in the analysis and when we formulate an algorithm solving it, but the same approach could be applied also to formulation (Π-QCE-NFG).

Still, in both of these formulations, the signaling scheme λ is assumed to be given and fixed. We may, however, intend to search for a scheme that is optimal in some sense, as if the signals constituted an action space of an additional player – a signaler. Possible criteria may include maximization of social welfare or signaler's own utility. In case the criterion has a form of function f, the formulation under the condition that the players attain quantal correlated equilibrium is:

Definition 6.3. *Let $G = (N, A, S, u)$ be a signaling game and $f : \Lambda \times \Sigma \to \mathbb{R}$ be a criterion function. The optimal quantal correlated equilibrium is*

$$\max_{\lambda \in \Lambda, \sigma \in \Sigma} f(\lambda, \sigma) \qquad s.t.\ \sigma \in QCE(\lambda), \tag{OPT}$$

where $QCE(\lambda)$ refers to the set of quantal correlated equilibria for a given signaling scheme λ, and Σ is either the set of behavioral strategies B, for the per-signal equilibrium, or the set of mixed strategies Δ, for the over-pure-strategies equilibrium.

The difficulty of this optimization is affected by properties of function f as well as the topology of quantal correlated equilibria. For this reason, we study the quantal correlated equilibria in more detail in the following section.

6.2 Properties of the equilibrium

In this section we investigate basic properties of quantal correlated equilibria. We begin our analysis by examining the relation to quantal response equilibrium.

Proposition 6.1. *Let $G = (N, A, S, u)$ be a signaling game. Then*

1. *the over-pure-strategies quantal correlated equilibrium is a normal form quantal response equilibrium in the extended game, and*
2. *the per-signal quantal correlated equilibrium is an extensive form (agent) quantal response equilibrium in the extended game.*

Consequently, both concepts exist for all continuous generators and any signaling scheme.

Proof. We construct two simple reductions from G to a normal form game for showing the relation to quantal response equilibrium; and to an extensive form game to show a relation to an agent quantal response equilibrium.

[146] G. Gigerenzer and D. G. Goldstein. "Reasoning the fast and frugal way: Models of bounded rationality." *Psychological Review* 103:4, 1996, p. 650.

[147] J. R. Gray. "A bias toward short-term thinking in threat-related negative emotional states". *Personality and Social Psychology Bulletin* 25:1, 1999, pp. 65–75.

[148] N. K. Malhotra. "Information load and consumer decision making". *Journal of Consumer Research* 8:4, 1982, pp. 419–430.

1. Let $G' = (N', A', u')$ be defined as follows:
   - the players are shared with G, i.e., $N' = N$,
   - the actions are $A' = \Pi_1 \times \cdots \times \Pi_n$[149], and
   - the utility u' is

   $$u'_i(a) = \sum_{(a_1|s_1,\ldots a_n|s_n)\in\times a} \lambda(s_1,\ldots,s_n)u(a_1,\ldots,a_n). \qquad \forall a \in A'$$

   The expected utilities hence correspond to the definition of utility in over-pure-strategies quantal correlated equilibrium. The quantal response equilibrium in G' is a fixed point of the dynamic and hence satisfies the definition of the desired equilibrium in G.

2. Let G' be an extensive form representation of the G's extended game, where the information sets are defined by the signal a player observes and the utility of their action a_i in a terminal node below the information set determined by signal s_i is given as

   $$u'_i(a_i, a_{-i}|s_i) = \sum_{s_{-i}\in S_{-i}} \lambda(s_i, s_{-i})u_i(a_i, a_{-i}).$$

   In extensive form quantal response equilibrium, each player acts according to their quantal response model in each information set separately. The expected utility in G' therefore corresponds to the definition of expected utility in per-signal quantal correlated equilibrium, and the fixed point of the agent quantal response dynamic is the per-signal equilibrium.

The existence follows from the existence of quantal response equilibria. □

[149] In other words, the actions of G' are the pure strategies of G.

Remark 6.1. When restricted to playing quantal responses, some mixed or behavioral strategies may become unavailable. In general, this leads to the failure of Kuhn's theorem in quantal strategies. In context of correlated equilibria, some per-signal quantal responses may not have an equivalent representation as over-pure-strategies quantal responses, or vice versa.

This fact becomes obvious when we examine the graph in Figure 6.1. Here, the action o is never played with a probability higher than 0.6 in the per-signal response against any viable strategy of the opponent. In contrast, the same action may be played as a quantal response with 0.6 or higher probability in over-pure-strategies formulation. As a consequence, the chances of both quantal correlated equilibria being equivalent (in some sort, e.g., as restrictions) is low. However, there exist special examples when both responses give rise to the same equilibrial strategy.

Proposition 6.2. *Let $G = (N, A, S, u)$ be a two-player signaling game where $|S_1| = 1$ and q_2 is exponential. Then the equilibrium strategies of player 1 in both quantal correlated equilibria coincide.*

Proof. Consider a quantal response of the second player against a fixed strategy δ_1 of the first player in a per-signal formulation, defined as

$$\beta_2(a_2|s_2) = q_2(u_2(a_2, s_2))/\sum_{a_2' \in A_2} q_2(u_2(a_2', s)).$$

When we multiply both the nominator and denominator by

$$\sum_{\pi \in \Pi_2 : (a_2,s_2) \in \pi} \prod_{(a_2',s_2') \in \pi \setminus (a_2,s_2)} q_2(u_2(a_2', s_2')),$$

we can express $\beta_2(a_2|s_2)$ as

$$\frac{\sum_{\pi \in \Pi_2 : (a_2,s_2) \in \pi} \prod_{(a_2',s_2') \in \pi \setminus (a_2,s_2)} q_2(u_2(a_2', s_2')) q_2(u_2(a_2, s_2))}{\left(\sum_{a_2' \in A_2} q_2(u_2(a_2', s_2))\right)\left(\sum_{\pi \in \Pi_2 : (a_2,s_2) \in \pi} \prod_{(a_2',s_2') \in \pi \setminus (a_2,s_2)} q_2(u_2(a_2', s_2'))\right)}.$$

Because q_2 is exponential, we may push the product inside the generator as a sum. At the same time, because the first player has only one trivial signal, we have

$$u_2(\pi) = \sum_{(a_2,s_2) \in \pi} \sum_{a_1 \in A_1} \lambda(s)\delta_1(a_1)u_2(a_1, a_2) = \sum_{(a_2,s_2) \in \pi} u_2(a_2|s_2).$$

Together, this enables us to relate the strategy in the per-signal formulation to the strategy in the over-pure-strategies formulation as

$$\beta_2(a_2|s_2) = \sum_{\pi \in \Pi_2 : (a_2,s_2) \in \pi} \delta_2(\pi).$$

Substituting for β_2 in the definition of expected utility of the first player, we get

$$\begin{aligned} u_1(a_1) &= \sum_{a_2 \in A_2} \sum_{s_2 \in S_2} \lambda(s_2)\beta(a_2|s_2)u(a_1, a_2) \\ &= \sum_{a_2 \in A_2} \sum_{s_2 \in S_2} \sum_{\pi \in \Pi_2 : (a_2,s_2) \in \pi} \lambda(s_2)\delta_2(\pi)u(a_1, a_2) = u_1(\pi = a), \end{aligned}$$

hence, both formulations lead to the same expected utilities of the first player. For arbitrary q_1, the equilibrium is thus reached at the same strategy. □

We clarified how quantal correlated equilibrium may be represented as a quantal response equilibrium, and outlined when both formulations of quantal correlation result in the same response. Now we examine the other direction: how quantal response equilibrium relates to quantal correlated equilibrium and when it may be extended into one.

Proposition 6.3. *Any quantal response equilibrium in a normal form game G*

1. *may be extended into a per-signal quantal correlated equilibrium lying on a corner of the signaling simplex; and*
2. *is a trivial over-pure-strategies quantal correlated equilibrium with a single signal per player.*

Proof. We show how to construct specific signal sets and signaling schemes such that the quantal response equilibrium may be represented as a quantal correlated equilibrium.

1. Consider λ that is a corner of the signaling simplex, i.e., there exists exactly one signal profile $s \in S$ such that $\lambda(s) = 1$. The signal profile identifies a specific subgame in the extended game, consisting of a copy of G. The expected utilities in the subgame are equivalent to expected utilities in G, and the solution is hence the quantal response equilibrium. If some signal s'_i is never observed, the corresponding strategy $\beta_i(\cdot|s'_i)$ is uniform because all expected utilities are zero. The per-signal quantal correlated equilibrium hence consists of uniform and quantal-response equilibrial strategies.

2. Assume that for each player i, $|S_i| = 1$. Then the extended game is equal to G and quantal response equilibrium is trivially correlated. In case at least one player i' has $|S_{i'}| > 1$, then $|\Pi_{i'}|$ in the extended game is strictly greater than the number of their actions in G. Because the generators are strictly positive, the quantal responses are interior points of the probabilistic simplex, and the quantal response equilibrium may never be extended into an over-pure-strategies quantal correlated equilibrium.

□

Remark 6.2. When studying the relations between correlated and uncorrelated quantal equilibria, we may ask if we may relate the number of equilibria for each concept. In some classes of games[150], the quantal response equilibrium is guaranteed to be unique. Because the extended game of a zero sum game is trivially zero sum as well, and quantal correlation may be represented as a standard quantal response interaction by Proposition 6.1, the quantal correlated equilibria are also unique. In other classes, the answer remains ambiguous: it is easy to construct examples with arbitrary ordering of the number of correlated and uncorrelated quantal equilibria.

[150] For example, in the zero sum games.

This fact is also clear when looking at Figure 6.1. In this game, there is one per-signal quantal correlated, two quantal response, and three over-pure-strategies quantal correlated equilibria. This motivates the attempt to characterize the topology of the space of quantal correlated equilibria.

Proposition 6.4. *Let $G = (N, A, S, u)$ be a signaling game and*

$$\mathcal{C} = \{(\lambda, QCE(\lambda)), \lambda \in \Lambda\},$$

where QCE is either S-QCE-NFG or Π-QCE-NFG. Then $\mathcal{C}$ is compact and the correspondence $\lambda \to QCE(\lambda)$ is upper hemicontinuous. Consequently, if $QCE(\lambda)$ is unique for all $\lambda \in \Lambda$, then $\mathcal{C}$ is connected.

Proof. We proceed similarly as in Theorem 3 in the original paper on quantal response equilibrium in normal form games[151]. We observe that as all generators are continuous, both formulations of quantal correlated equilibria may be written as zeros of continuous systems[152] of variables (λ, β) or (λ, δ). Because the systems are continuous, $\mathcal{C}$ is closed. As both λ and β, δ are bounded, $\mathcal{C}$ is compact. Therefore, the correspondence is upper hemicontinuous. When there exists only one quantal correlated equilibrium for any λ, then $\mathcal{C}$ is connected because it is an image of a connected space by a continuous correspondence. □

[151] R. D. McKelvey and T. R. Palfrey. "Quantal response equilibria for normal form games". *Games and Economic Behavior* 10:1, 1995, pp. 6–38.

[152] We use this fact in the next section to define the equilibrial homotopy.

Example 6.3. *Figure 6.1 shows an example of the correspondence $\mathcal{C}$ depicted in gray x-lines. Here, $\mathcal{C}$ is projected on the coordinates corresponding to the probability of observing signal s_2^1 and probability of playing action o in S-QCE-NFG(s_2^1). The graph varies $s_2^1 \in [0, 1]$ on the x-axis and depicts three correspondence branches which clearly delineates the projection's non-convexity and discontinuity. The full set $\mathcal{C}$ could hence never be convex as well.*

Searching for sufficient conditions of the concepts' convexity is difficult, as even obvious, natural choices of quantal generators, signals or utility functions (e.g., linear generators, $S = A$ or zero-sum games) lead to non-convex solutions. Optimizing over the set of signaling schemes is thus difficult and gradient methods may not reach the maximum even for concave criterion functions. Any system designer may still benefit from looking for an optimal signaling scheme, as for some criteria[153], the maximum is never reached in quantal response equilibrium.

[153] We omit trivial examples when the signaler's criterion depends on the entropy of the signaling scheme or the players' strategies, which naturally leads to optimal distributions unrelated to quantel response equilibrium.

Proposition 6.5. *Let G be a signaling game with positive utilities where each player behaves according to a quantal response with an exponential generator. Assume that the quantal response equilibrium in the underlying game is non-uniform and the signaler optimizes their fully rational expected utility that is always positive and negatively correlated with utilities of all players. Then the signaler's utility in quantal response equilibrium is smaller than in other quantal correlated equilibria.*

Proof. Assume that λ is a corner of the signaling simplex. According to Proposition 6.3, the equilibrial strategies in the signaling game form a quantal response equilibrium. We show that a simple change from the corner λ to a scheme with full support will result in a non-zero increase in the signaler's utility. Let λ' be a uniform distribution over S. The corresponding expected utilities then preserve the ordering of utilities with λ but their magnitude will be strictly smaller. Because quantal response equilibrium is continuous and the generators q are exponential, the resulting equilibrial strategies will be more flat, i.e., closer to uniform, decreasing the overall expected utility of all players because higher-utility actions are played with strictly lower probability. Because the signaler's utility is positive and negatively correlated with other players, their overall utility increases. □

The concepts' non-convexity brings into question their relation to correlated equilibrium, which, in contrast, is known to be convex. A classical result from quantal response equilibrium may be generalized for correlated strategies, describing correlated equilibrium as a limit of quantal correlated equilibrium with certain parametric generators.

Proposition 6.6. *Let q_p be a parametric generator continuous in $p \in \mathbb{R}$ with $q_{p'} \in O(q_{p''})$ for any $p' < p''$. Let $\{p^1, p^2, \dots\}$ be a sequence such that $\lim_{t\to\infty} p^t = \infty$, and $\{\beta^1, \beta^2, \dots\}, \beta^j \in B$ be a sequence of corresponding quantal correlated equilibria with generators q_{p^t} for a fixed signaling scheme $\lambda \in \Lambda$. Then $\beta^* = \lim_{t\to\infty} \beta^t$ is a correlated equilibrium for λ.*

Proof. According to Proposition 6.1, the quantal correlated equilibria are quantal response equilibria in the extended games. By the the same reasoning as in Theorem 2 in the original quantal response equilibrium paper by McKelvey and Palfrey[151], the limiting quantal response equilibria are Nash equilibria. Therefore, the limiting quantal correlated equilibria are Nash equilibria in the extended game, which are (by definition) correlated equilibria[154]. □

[154] Note that this result does not depend on the precise definition of the expected utility, i.e., subjective perceptions of utilities are viable as long as they preserve the total ordering of objective utilities.

The convergence of these quantal responses to a best response may be hence seen as a driver of "convexification" of the solution space. This suggests that optimizing over signaling schemes for quantal response functions closely resembling best response may yield better results than when searching for a maximizing scheme with less steep generators. Because an essential step in the optimization is computing a quantal correlated equilibrium for a given scheme, as a last result in this section we examine the concepts' computational complexity.

Proposition 6.7. *Let G be a signaling game of n players and $q_1, \dots, q_n$ be their respective generators. The problem of finding a quantal correlated correlated equilibrium in G is PPAD-hard.*

Proof. Let G be a two-player signaling game with strictly positive utilities and signal sets of arbitrary cardinality, in which both players have n actions. Computing an ϵ-Nash equilibrium in G is PPAD-complete, as per Theorem 3.3. We show that computing quantal correlated equilibrium is PPAD-hard by reducing the problem of finding ϵ-Nash equilibrium to a problem of computing a specific quantal correlated equilibrium. We proceed as follows: let both players share the same logit generator $q(x) = e^{Cx}$. By Lemma 3.1, for each ϵ there exists a polynomially computable C such that the induced quantal response is an ϵ-best-response. The quantal response equilibrium is hence an ϵ-Nash equilibrium. Let λ be a corner of the signaling simplex. By Proposition 6.3, a quantal response equilibrium is a restriction of a quantal correlated equilibrium in G. Therefore, a quantal correlated equilibrium in G is an ϵ-Nash equilibrium. □

Now we are ready to move to the introduction of a practical algorithm for computing a quantal correlated equilibrium.

6.3 Homotopy method for finding the equilibrium

In this section, we first review the literature on finding quantal response equilibrium. None of existing algorithms could be adapted for computing (even non-optimal) quantal correlated equilibrium directly because of two main problems: they focus on the logit generator $q(x) = e^{\alpha x}$, and use its properties nontransferable to other generators in the Luce model; and they are able to compute only the normal form representation of quantal response equilibrium which does not translate to the per-signal equilibrium as in S-QCE-NFG. We hence formulate a new homotopy method optimizing quantal correlated equilibrium by reformulating the generalized Luce model to improve robustness, employing general product-separating functions to alleviate steepness of quantal generators and simultaneously tracing the equilibrium and gradiently optimizing the signaling scheme. We derive the precise algorithm and analyze its convergence properties.

Several methods have been introduced for computing quantal response equilibrium in different classes of games. Most of them focus on the logit generator that enables leveraging the unique correspondence with the Gibbs entropy regularizer.[155] Out of them, only the Karush-Kuhn-Tucker reformulation of per-player optimization is capable of converging in some general sum games.[156] The main limitation of this approach is that the sufficient assumptions of convergence can not be efficiently verified in practice. Other techniques rely on the structure of (weighted) zero sum games and employ a specific Karush-Kuhn-Tucker reformulation of the equilibrial point[157] or smooth Q-learning.[158]

To the best of our knowledge, the only method in the literature that does not depend on the entropy regularization is based on a homotopic approach. The idea of homotopy methods is to introduce a single-parametric system of (nonlinear) equations with a trivial solution on one end of the parametric range and the desired, unknown solution on the other. By following a path from the trivial solution (i.e., by continuously deforming the system from the simple to the complex one) we approach the desired one. Advantages of homotopy methods include their numerical stability and potential to be globally convergent. The homotopy for quantal response equilibrium with no optimization considered was first derived for normal form games with the same logit generator for all players[159]. Another homotopy for quantal response equilibrium was introduced in context of sponsored search auctions, using specific properties thereof.[160] Neither could be used for quantal correlation because of their strict assumptions about game domains and quantal models.

6.3.1 Tracing the equilibrial correspondence path

We formulate a homotopy method for quantal correlated equilibrium. Contrary to previous methods for quantal response equilibrium, our method has multiple favorable properties. It applies to any general sum game and enables to find even per-signal equilibria, which the other methods are inadaptable for. The first-order description of tracing promises a better scalability over the second-order Karush-

[155] P. Mertikopoulos and W. H. Sandholm. "Learning in games via reinforcement and regularization". *Mathematics of Operations Research* 41:4, 2016, pp. 1297–1324.

[156] K. Wang, L. Xu, A. Perrault, M. K. Reiter, and M. Tambe. "Coordinating followers to reach better equilibria: End-to-end gradient descent for Stackelberg games". In: *Proceedings of the 36th AAAI Conference on Artificial Intelligence*. 2022, pp. 5219–5227.

[157] C. K. Ling, F. Fang, and J. Z. Kolter. "What game are we playing? End-to-end learning in normal and extensive form games". In: *Proceedings of the 27th International Joint Conference on Artificial Intelligence*. 2018, pp. 396–402.

[158] S. Leonardos, G. Piliouras, and K. Spendlove. "Exploration-exploitation in multi-agent competition: Convergence with bounded rationality". *Advances in Neural Information Processing Systems* 34, 2021.

[159] T. L. Turocy. "A dynamic homotopy interpretation of the logistic quantal response equilibrium correspondence". *Games and Economic Behavior* 51:2, 2005, pp. 243–263.

Kuhn-Tucker methods[161]. Moreover, each player may use a different generator, not necessarily a logit one[162], which enables tailoring the concept to groups of players of different behavioral profiles. To be able to trace the parametric path, we assume the generators have parametric representations.

Definition 6.4. *Let q be a generator of a generalized Luce quantal function. We call $\hat{q}(x,t)$ a parametric representation of q in case $\hat{q}$ is differentiable and*

$$\hat{q}(x,t=0) = c, \quad c \in \mathbb{R}$$
$$\hat{q}(x,t=1) = q(x).$$

Example 6.4. *Perhaps the simplest way of creating parametric representations is in terms of exponentiation. Consider a logit generator $exp(\alpha x)$. One of its possible parametric representations is $exp(\alpha t x)$, which is equal to 1 for $t = 0$ and to the generator for $t = 1$. Similarly for a Luce generator $(x + C)^\alpha$, we may choose a parametric representation $(x + C)^{\alpha t}$. For logarithmic generators, e.g., $log(\alpha x + C)$, a possible representation may be $log(\alpha x + C)^t$. In all three examples, α and C are suitable constants such that the resulting generators give rise to valid quantal functions in a given game.*

Example 6.5. *Another way to construct parametric representations comes from some of the ideas behind the work of Isaac Newton, and is therefore referred to as newtonian.*[163] *For any generator q, the newtonian representation is formulated as*

$$\hat{q}(x,t) = tq(x) + (1-t).$$

Note that newtonian representations are differentiable whenever the original generator function is. For example, for the logit generator $exp(\alpha x)$ the newtonian representation looks as $texp(\alpha x) + (1-t)$.

Homotopic function $H(x,t) : \mathbb{R}^{m+1} \to \mathbb{R}^m$ is then a function with a homotopic parameter t, such that the system $H(x,t) = 0$ has a trivial solution for $t = 0$ and the desired solution for $t = 1$. Motivated by the definition of system (Π-QCE-NFG), we define a homotopy function $\overline{H}$ for quantal correlated equilibrium with strategy profile β and homotopic parameter t as[164]

$$\overline{H}(\beta,t) = \left(H_i^{k,l}(\beta,t)\right)_{i\in N,\ a_i^k\in A_i,\ s_i^l\in S_i}$$
$$\overline{H}_i^{k,l}(\beta,t) = \hat{q}_i(u_i(a_i^k|s_i^l),t) - \beta_i(a_i^k|s_i^l)\sum_{a_i\in A_i}\hat{q}_i(u_i(a_i|s_i^l),t).$$

Here, $m = \sum_{i\in N}|A_i|\cdot|S_i|$. The solutions are points (β,t), such that $H(\beta,t) = 0$ – a set of one or more paths – and we aim to trace one of the paths from $t = 0$ to $t = 1$. Clearly, the homotopic system is equivalent to system (Π-QCE-NFG) for $t = 1$ by the definition of parametric representations.

[160] J. Rong, T. Qin, B. An, and T.-Y. Liu. "Modeling bounded rationality for sponsored search auctions". In: *Proceedings of the 22nd European Conference on Artificial Intelligence*. 2016, pp. 515–523.

[161] For example, the experiments in Ling, Fang, and Kolter[157] consider optimization over rock-paper-scissors, one-card poker, and a security game with sequence-form payoff matrices of per-player size at most 16.

[162] Traversing the homotopic curve may correspond to different exploration-exploitation policies too, e.g., the Explore-Then-Exploit rule, accompanied by an appropriate parametric representation.

[163] C. B. Garcia and W. I. Zangwill. *Pathways to solutions, fixed points, and equilibria*. English. Prentice-Hall Series in Computational Mathematics. Englewood Cliffs, New Jersey: Prentice-Hall, Inc. XV, 479 p. $ 43.95 (1981). 1981.

[164] Note that this is a reformulation of the second equation in Π-QCE-NFG with the generators substituted by their parametric representations.

Proposition 6.8. *For arbitrary λ, the solution of the associated homotopy system for $t = 0$ is a uniform strategy for each signal and each player.*

Proof. For t=0 we have $H_i^{k,l}(\beta, t = 0) = c_i - \beta(a_i^k|s_i^l)|A_i|c_i = 0$, hence $\beta_i^{t=0}(a_i|s_i) = 1/|A_i|$. □

However, the experiments proved that such a system may become numerically unstable with paths containing multiple bifurcation points, causing the tracing to significantly slow down or stall. For example, assume that two players who act according to logit generators $q(x) = exp(x)$ engage in a game depicted in Figure 6.2. In case we trace the quantal correlated equilibrium with trivial single signals using the original homotopy introduced at the beginning of Section 6.3.1, the algorithm becomes stuck in a bifurcation point when $t \approx 0.4029$. The consequent numerical issues that arise result in a failure to reach the equilibrium.

		Player 1	
		A	B
Player 2	o	-2,23	28,10
	r	-28,-13	-22,-9

Figure 6.2: An example of a normal form game where the original homotopy method fails to reach the equilibrium.

We hence derive a reformulated system that relies on generalization of two folk techniques that increase the robustness of the method. First, for each player we choose a single reference action, denoted as $a_i^0, i \in N$. For any other action $a_i^j \neq a_i^0$ and signal $s_i \in S_i$, we have

$$\beta_i(a_i^0|s_i) = \frac{q_i(u_i(a_i^0|s_i))}{\sum_{a_i \in A_i} q_i(u_i(a_i|s_i))},$$
$$\beta_i(a_i^j|s_i) = \frac{q_i(u_i(a_i^j|s_i))}{\sum_{a_i \in A_i} q_i(u_i(a_i|s_i))}.$$

Because both equalities share the same sum $\sum_{a_i \in A_i} q_i(u_i(a_i|s_i))$, we may write

$$\beta_i(a_i^0|s_i)q_i(u_i(a_i^j|s_i)) = \beta_i(a_i^j|s_i)q_i(u_i(a_i^0|s_i)). \tag{6.1}$$

This reformulation enables to eliminate numerical errors associated with the sum, which for some quantal generators may reach extremely high values. On the down side, the β values are no longer normalized by the sum, becoming unbounded. For each signal $s_i \in S_i$, we hence enforce that

$$\sum_{a_i \in A_i} \beta_i(a_i|s_i) = 1.$$

Still, the possible differences in magnitudes of β and the effective range of some q_i's result in further numerical instabilities. To alleviate it, we aim to apply some concave bijective univariate function f over Equation (6.1). To introduce an efficiently implementable change of variables, we require that f is a product-separating function, i.e.,

$$f(xy) = f_2(f_1(x), f_1(y)), x, y \in \mathbb{R},$$

and f_1 has an inverse f_1^{-1}. As an example of f, consider $f(x) = x^{1/c}, c > 1$ with $f_1 = f$ and f_2 being a product of its arguments. Similarly, we could have $f = log$,

with $f_1 = f$ and f_2 being a sum. Applying f to Equation (6.1) then motivates a substitution of variables $\gamma = f_1(\beta)$ and the resulting homotopy is formulated as

$$\begin{aligned} H(\gamma, t) =& \Big(H_i^{k,l}(\gamma, t)\Big)_{i \in N,\ a_i^k \in A_i,\ s_i^l \in S_i} \\ H_i^{k,l}(\gamma, t) =& f_2(f_1(\hat{q}_i(u_i(a_i^0|s_i^l), t)), \gamma_i(a_i^k|s_i^l)) \\ & - f_2(f_1(\hat{q}_i(u_i(a_i^k|s_i^l), t)), \gamma_i(a_i^0|s_i^l)) \\ H_i^{0,l}(\gamma, t) =& \sum_{a_i^k \in A_i} f_1^{-1}(\gamma_i(a_i^k|s_i^l)) - 1. \end{aligned}$$

Because H is a reformulation of system $\overline{H}$, a simple modification of Proposition 6.8 remains true and the solution for $t = 0$ is trivially $\gamma_i^{t=0}(a_i|s_i) = f_1(1/|A_i|)$. To efficiently trace the path from this initial solution, we have to account for the possibility that a branch we follow is not monotonic in t. The pairs (γ, t) are hence parameterized by p, i.e., the homotopy will compute a parametric path $c(p) = (\gamma(p), t(p))$, where p is interpreted as the arclength along the path. As the following theorem shows, such path exists and is unique[165].

[165] E. L. Allgower and K. Georg. *Numerical continuation methods: An introduction*. Vol. 13. Springer Science & Business Media, 2012.

Theorem 6.1. *Let $H : \mathbb{R}^{m+1} \to \mathbb{R}^m$ be a smooth homotopic map. Let $u_0 \in \mathbb{R}^{m+1}$ be a point such that $H(u_0) = 0$ and the Jacobian matrix $H'(u_0)$ has maximum rank. Then there exists a unique smooth curve $p \in J \to c(p) \in \mathbb{R}^{m+1}$ which satisfies $c(0) = u_0$ and $H(c(p)) = 0$ for p in some open interval J containing zero, such that for all $p \in J$, the tangent $c'(p)$ is smoothly induced by the Jacobian matrix $H'(c(p))$ and satisfies the following three conditions:*

1. $H'(c(s))c'(s) = 0$,
2. $\|c(s)\| = 1$, *and*
3. $det\begin{pmatrix} H'(c(p)) \\ c'(p) \end{pmatrix} > 0$.

As a consequence of Theorem 6.1, the curve c associated with the quantal correlated equilibrium homotopy may be regarded as a local solution of an initial value problem defined as

1. $(\gamma(p), t(p))' = c'(H'(\gamma(p), t(p)))$, and
2. $(\gamma(0), t(0)) = (\gamma^{t=0}, 0)$,

where we abuse the notation a little and write c' as the tangent vector depending on the Jacobian matrix for a given value of p. Therefore, we may use any method suitable for solving initial value problems to trace c. The book of Allgower and Georg[165] recommends to use predictor-corrector continuation methods that better exploit the contraceptive properties of c with respect to the Newton-type iterative methods than general initial value problem solvers, and we will hence focus on them.

The standard predictor-corrector works in iterations, starting from the initial point $(\gamma^{t=0}, 0)$. As the name suggests, in each iteration ι it is given a point $(\gamma, t)^\iota$

on (or close to) the curve c and it performs two steps: the prediction and the correction. Most commonly, the Euler predictor is used[166], and it estimates the next point $\overline{(\gamma,t)}^{\iota+1}$ on the path using the current point and the step-size h as

$$\overline{(\gamma,t)}^{\iota+1} \leftarrow (\gamma,t)^{\iota} + hc'(H'((\gamma,t)^{\iota})).$$

Because the prediction often lies further from the curve, the correction step serves to refine it. To this end, we employ the Gauss-Newton correction method because under mild assumptions it guarantees an existence of a neighborhood of $(\gamma,t)^{\iota+1}$ such that successively applying the method to $\overline{(\gamma,t)}^{\iota+1}$ converges to a point $(\gamma,t)^{\iota+1}$ lying on the curve.[167] The Gauss-Newton method is formally defined as

$$\overline{(\gamma,t)}^{\iota+1} \leftarrow \overline{(\gamma,t)}^{\iota+1} - H'(\overline{(\gamma,t)}^{\iota+1})^{+}H(\overline{(\gamma,t)}^{\iota+1}),$$

where $^{+}$ denotes the Moore-Penrose inverse. Once a distance to the curve becomes sufficiently small, we set $(\gamma,t)^{\iota+1} \leftarrow \overline{(\gamma,t)}^{\iota+1}$. It also pays off to update the steplength h accordingly during iterations when c becomes more linear[168] to speed up the convergence. For this purpose, we use a simple h-adaptation by asymptotic expansion that updates h according to a contraction rate of two consecutive corrector runs and we switch to Newton adaptation when reaching $t = 1$.[169] The predictor-corrector terminates when t^{ι} attains a value close enough to 1.

As evident from the description, the efficiency of running the algorithm relies on the ability to compute the curve tangent c' and the Moore-Penrose inverse of the Jacobian matrix. Fortunately, both may be computed from QR factorization of the transposed matrix $H'^{\top}$ using a standard methods described in the book of Allgower and Georg[165]. QR factorization represents $H'^{\top} \in \mathbb{R}^{m,m+1}$ as

$$H'^{\top} = Q\begin{pmatrix} R \\ 0 \end{pmatrix},$$

where $Q \in \mathbb{R}^{m+1,m+1}$ is an orthogonal matrix and $R \in \mathbb{R}^{m,m}$ is a non-singular upper triangular matrix. A notable advantage of QR factorization is its numerical stability. Let z denote the last column of Q, then the tangent and the Moore-Penrose inverse may be obtained as

$$c' = sgn(det(Q)det(R))z, \text{ and}$$
$$H'^{+} = Q\begin{pmatrix} R^{\top -1} \\ 0 \end{pmatrix}.$$

The matrix $R^{\top}$ is not inverted in practice as calculating $w = H'^{+}b$ is typically done by forward solving $R^{\top}y = b$. It remains to show how the Jacobian matrix of the homotopic system for quantal correlated equilibrium looks like. Let us first consider the derivatives of $(H_i^{k,l})_{i\in N,\ a_i^k\in A_i, k>0,\ s_i^l\in S_i}$ with respect to γ and t.

[166] Our experiments with lower-degree Runge-Kutta methods yielded similar results.

[167] A. Ben-Israel and T. N. Greville. *Generalized inverses: Theory and applications*. Vol. 15. Springer Science & Business Media, 2003.

[168] Or conversely, when c becomes more curvy.

[169] K. Georg. "A note on stepsize control for numerical curve following". In: *Homotopy Methods and Global Convergence*. Springer, 1983, pp. 145–154.

$$
\begin{aligned}
\frac{\partial H_i^{k,l}}{\partial \gamma_i(a_i^k|s_i^l)} =& \frac{\partial f_2(f_1(\hat{q}_i(u_i(a_i^0|s_i^l),t)),\gamma_i(a_i^k|s_i^l))}{\partial \gamma_i(a_i^k|s_i^l)} \\
\frac{\partial H_i^{k,l}}{\partial \gamma_i(a_i^0|s_i^l)} =& -\frac{\partial f_2(f_1(\hat{q}_i(u_i(a_i^k|s_i^l),t)),\gamma_i(a_i^0|s_i^l))}{\partial \gamma_i(a_i^0|s_i^l)} \\
\frac{\partial H_i^{k,l}}{\partial \gamma_{i'}(a_{i'}^{j'}|s_{i'}^{l'})} =& \frac{\partial f_2(f_1(\hat{q}_i(u_i(a_i^0|s_i^l),t)),\gamma_i(a_i^k|s_i^l))}{\partial f_1(\hat{q}_i(u_i(a_i^0|s_i^l),t))} \\
& \cdot \frac{\partial f_1(\hat{q}_i(u_i(a_i^0|s_i^l),t))}{\partial \hat{q}_i(u_i(a_i^0|s_i^l),t))} \frac{\partial \hat{q}_i(u_i(a_i^0|s_i^l),t)}{\partial u_i(a_i^0|s_i^l)} \frac{\partial u_i(a_i^0|s_i^l)}{\partial \gamma_{i'}(a_{i'}^{j'}|s_{i'}^{l'})} \\
& - \frac{\partial f_2(f_1(\hat{q}_i(u_i(a_i^k|s_i^l),t)),\gamma_i(a_i^0|s_i^l))}{\partial f_1(\hat{q}_i(u_i(a_i^k|s_i^l),t))} \frac{\partial f_1(\hat{q}_i(u_i(a_i^k|s_i^l),t))}{\partial \hat{q}_i(u_i(a_i^k|s_i^l),t))} \\
& \cdot \frac{\partial \hat{q}_i(u_i(a_i^k|s_i^l),t)}{\partial u_i(a_i^k|s_i^l)} \frac{\partial u_i(a_i^k|s_i^l)}{\partial \gamma_{i'}(a_{i'}^{j'}|s_{i'}^{l'})} \\
\frac{\partial H_i^{k,l}}{\partial t} =& \frac{\partial f_2(f_1(\hat{q}_i(u_i(a_i^0|s_i^l),t)),\gamma_i(a_i^k|s_i^l))}{\partial f_1(\hat{q}_i(u_i(a_i^0|s_i^l),t))} \frac{\partial f_1(\hat{q}_i(u_i(a_i^0|s_i^l),t))}{\partial \hat{q}_i(u_i(a_i^0|s_i^l),t))} \\
& \cdot \frac{\partial \hat{q}_i(u_i(a_i^0|s_i^l),t)}{\partial t} \\
& - \frac{\partial f_2(f_1(\hat{q}_i(u_i(a_i^k|s_i^l),t)),\gamma_i(a_i^0|s_i^l))}{\partial f_1(\hat{q}_i(u_i(a_i^k|s_i^l),t))} \frac{\partial f_1(\hat{q}_i(u_i(a_i^k|s_i^l),t))}{\partial \hat{q}_i(u_i(a_i^k|s_i^l),t))} \\
& \cdot \frac{\partial \hat{q}_i(u_i(a_i^k|s_i^l),t)}{\partial t},
\end{aligned}
$$

where $i' \neq i$ and for any k, including 0,

$$
\begin{aligned}
\frac{\partial u_i(a_i^k|s_i^l)}{\partial \gamma_{i'}(a_{i'}^{k'}|s_{i'}^{l'})} =& \sum_{\substack{a_{-i}\in A_{-i} \\ a_{i'}^{k'}\in a_{-i}}} \sum_{\substack{s_{-i}\in S_{-i} \\ s_{i'}^{l'}\in s_{-i}}} \lambda(s_i,s_{-i}) u_i(a_i^k,a_{-i}) \frac{\partial f_1^{-1}(\gamma_{i'}(a_{i'}^{k'}|s_{i'}^{l'}))}{\partial \gamma_{i'}(a_{i'}^{k'}|s_{i'}^{l'})} \\
& \cdot \prod_{\substack{j\in -i\setminus i' \\ a_j\in a_{-i}, s_j\in s_{-i}}} f_1^{-1}(\gamma_j(a_j|s_j)).
\end{aligned}
$$

All other derivatives are equal to zero. Now we turn to the description of derivatives of $(H_i^{0,l})_{i\in N,\ s_i^l\in S_i}$, which are non-zero only in the case of

$$
\frac{\partial H_i^{0,l}}{\partial \gamma_i(a_i^k|s_i^l)} = \frac{\partial f_1^{-1}(\gamma_i(a_i^k|s_i^l))}{\partial \gamma_i(a_i^k|s_i^l)}.
$$

The whole algorithm for computing quantal correlated equilibrium is depicted in Algorithm 3. Here, h is the initial steplength with $\underline{h}$ being its minimum value, and

Algorithm 3: Predictor-corrector method for tracing the normal form quantal correlated equilibrium correspondence along the homotopic path

Input: $H, (\gamma, t)$ such that $H((\gamma, t)) = 0, \bar{\iota}$
Parameters: $h, \underline{h}, \bar{\iota}_{gn}, \underline{\epsilon}, \bar{\epsilon}, \epsilon_c, \bar{\epsilon}_c, \epsilon_t, a_\kappa, a_f, a_\eta$
$\iota \leftarrow 0$
while $t < (1 + \epsilon_t)$ ***and*** $\iota < \bar{\iota}$ **do**
  $H' \leftarrow Jacobian(\gamma, t)$
  $Q, R \leftarrow QR(H')$
  $\overline{(\gamma, t)} \leftarrow Euler((\gamma, t), Q, R)$
  $accept \leftarrow True, \quad \iota_{gn} \leftarrow 0, \quad f \leftarrow 1/a_f, \quad \|c|(\gamma, t)\| \leftarrow \infty$
  while $\|c|(\gamma, t)\| > \epsilon_c$ **do** `// check distance from the curve`
    $\overline{(\gamma, t)} \leftarrow GaussNewton((\overline{\gamma, t}), Q, R), \qquad \iota_{gn} \leftarrow \iota_{gn} + 1$
    $h,\ \|c|(\gamma, t)\|,\ f,\ accept,\ newton \leftarrow$
     $UpdateStep(h, \iota_{gn}, f, (\gamma, t), \overline{(\gamma, t)}, \|c|(\gamma, t)\|, \underline{\epsilon}, \bar{\epsilon}, a_\kappa, a_f, a_\eta, newton)$
    if $\|c|(\gamma, t)\| > \bar{\epsilon}_c$ ***or*** $\iota_{gn} > \overline{\iota_{gn}}$ **then** $accept \leftarrow False$
    if ***not*** *accept* **then** break
  if $accept$ **then** $(\gamma, t) \leftarrow \overline{(\gamma, t)}, \ \iota \leftarrow \iota + 1$ **else** $h \leftarrow h/a_f$
return (γ, t)

$\bar{\iota}_{gn}$ is a maximum number of iterations of the Gauss-Newton method. $\underline{\epsilon}$ and $\bar{\epsilon}$ are minimum and maximum distances from the curve, respectively, and ϵ_t is the termination distance of t from 1. a_κ, a_f, a_η are the maximum contraction, maximum decelaration, and perturbation parameters of the step adaptation. In Algorithm 4 we present a simple method for updating the steplength h. The method is a variant of Algorithm (6.1.10) from the book of Allgower and Georg[165]. The algorithm computes the distance of the current estimate of (γ, t) from the homotopy curve c and calculates the contraction rate κ as a ratio of two consecutive distances in the Gauss-Newton method using the parameter a_η that serves as a perturbation to prevent cancellation. The deceleration factor f is then calculated from κ and divides the current step h to estimate the next step.

Remark 6.3. The predictor-corrector method may potentially diverge because of the Jacobian unboundedness. Contrary to general sequential games,[170] computing Jacobian of the homotopy of quantal correlated equilibrium does not require normalization of opponents' strategies because the probability of observing a signal depends only on the signaling scheme. As a consequence, whenever f_1, f_2 and q_i's have bounded derivatives on their respective domains in the signaling game, the Jacobian is bounded. This holds also for all f_1, f_2 and q_i's considered in this chapter.

[170] T. L. Turocy. "Computing sequential equilibria using agent quantal response equilibria". *Economic Theory* 42:1, 2010, pp. 255–269.

6.3.2 Finding locally optimal signaling scheme

There may be multiple curves spanning across the solution space of the homotopy system $H(\gamma, t) = 0$. They may start and end at various points, some may be short and defined only over a subdomain of t, or entirely disconnected from others. In

Algorithm 4: Method UpdateStep for adapting the steplength in the predictor-corrector for tracing the equilibrial path

Input: $h, \iota, f, (\gamma, t), \overline{(\gamma, t)}, \|c|(\gamma, t)\|, \underline{\epsilon}, \overline{\epsilon}, a_\kappa, a_f, a_\eta, newton$
$\|c|(\gamma, t)\|^\iota \leftarrow \left\| H'(\overline{(\gamma, t)})^+ H(\overline{(\gamma, t)}) \right\|$
if *not* $newton$ ***and*** $(t-1)(\bar{t}-1) < 0$ **then** $newton \leftarrow True$
$f \leftarrow \max(f, a_f \sqrt{\|c|(\gamma, t)\|/\bar{\epsilon}})$
if $\iota > 2$ **then**
  $\kappa \leftarrow \|c|(\gamma, t)\|^\iota / (\|c|(\gamma, t)\|^{\iota-1} + a_\eta \underline{\epsilon})$
  if $\kappa > a_\kappa$ **then return** $h,\ \|c|(\gamma, t)\|^\iota,\ f,\ False,\ newton$
$f \leftarrow \max(f, a_f \sqrt{\kappa / a_\kappa})$
if $f > a_f$ **then** $f = a_f$
$h \leftarrow |h/f|$
if $newton$ ***and*** $\|c|(\gamma, t)\|^\iota < \underline{\epsilon}$ **then** $h \leftarrow -h(\bar{t}-1)/(\bar{t}-t+\underline{\epsilon})$
return $h,\ \|c|(\gamma, t)\|^\iota,\ f,\ True,\ newton$

the previous section, we described how to traverse a specific, unique branch that starts with uniform behavioral strategies[171] and gradually approaches the quantal correlated equilibrium for a given signaling scheme and quantal generators, moving across the whole domain of t[172]. In the literature, this branch is commonly referred to as the principal branch.[173] When interpreted in terms of learning, traveling along this path corresponds to a process when independent agents continuously explore and exploit an environment that is unknown to them, as explained by Leonardos, Piliouras, and Spendlove[158]. The exact same path is taken also by the replicator dynamic, a standard algorithm of evolutionary game theory, which was proved by Turocy[159].

Since the principal branch is unique, for a given λ and p, we have a unique equilibrium $\beta_p(\lambda)$, i.e., the equilibrium may be seen as a function of the signaling scheme and the homotopic parameter (or just the homotopic parameter in case of the quantal response equilibrium). Because of this correspondence and the branch's significance, it is often chosen as a domain to optimize over when selecting an optimal quantal response equilibrium, e.g., in auction parameter estimation in sponsored search auctions[158] or subrationality estimation for general normal form games.[174] Such optimization considers a fixed homotopy function and a criterion that seeks an optimal point on the homotopy's principal branch. As such, it is well suited for descriptive applications such as maximal-likelihood estimations from real world data. For quantal correlated equilibrium, a better suited optimization is the earlier mentioned formulation OPT:

$$\max_{\lambda \in \Lambda, \beta \in B} f(\lambda, \beta) \qquad s.t.\ \beta \in \text{S-QCE}(\lambda).$$

This formulation may be interpreted as a search for an optimal signaling scheme and hence offers prescriptive applications rather than descriptive ones as in the case of parameter estimations. For this purpose, we do not include the signaling scheme

[171] In other words, a centroid of the strategy simplex.

[172] More specifically, for quantal functions that approach best response, the corresponding quantal response equilibrium approximates a unique limiting Nash equilibrium on the principal branch.

[173] J. K. Goeree, C. A. Holt, and T. R. Palfrey. "Quantal response equilibria". In: *Behavioural and Experimental Economics*. Springer, 2010, pp. 234–242.

[174] R. D. McKelvey, A. M. McLennan, and T. L. Turocy. "Gambit: Software tools for game theory", 2006.

Algorithm 5: Gradient optimization of the signaling scheme

Input: $\lambda, (\gamma, t), H$ such that $H((\gamma, t)) = 0$
Parameters: $\eta, \underline{\eta}, \Delta_a, \Delta_r, \bar{\iota}$ `// + parameters of the predictor-corrector`
while $t < (1 + \epsilon_t)$ **do**
  $(\gamma, t) \leftarrow Predictor - Corrector(H, (\gamma, t), \bar{\iota}), \quad f'(\lambda, f_1^{-1}(\gamma)) \leftarrow backwards(f(\lambda, f_1^{-1}(\gamma)))$
  $\overline{H}, \overline{\lambda} \leftarrow ProjectedGradientAscent(\lambda, \eta, f'(\lambda, f_1^{-1}(\gamma)))$
  $Q, R = QR(\overline{H}'^{\top})$
  $accept \leftarrow True, \quad \iota \leftarrow 0$
  while $\left\| \overline{H}'(\overline{(\gamma, t)})^{+} \overline{H}(\overline{(\gamma, t)}) \right\| > \epsilon_c$ **do**
    $\iota \leftarrow \iota + 1$
    $\overline{(\gamma, t)} \leftarrow GaussNewton((\overline{\gamma, t}), Q, R)$
    if $\left\| \overline{H}'(\overline{(\gamma, t)})^{+} \overline{H}(\overline{(\gamma, t)}) \right\| > \epsilon_c^{max}$ ***or*** $\iota > \bar{\iota}$ **then**
      $accept \leftarrow False,$ break
  if *accept* **then** $(\gamma, t) \leftarrow \overline{(\gamma, t)}, \quad \lambda \leftarrow \overline{\lambda}, \quad H \leftarrow \overline{H}$
  if *accept* **then** $\eta \leftarrow \max(\underline{\eta}, \Delta_a \cdot \eta)$ **else** $\eta \leftarrow \Delta_r \cdot \eta$

in the definition of the homotopy for quantal correlated equilibrium, even though it would be possible. Instead, our aim is to purposefully optimize over the space of signaling schemes[175]. Both approaches may also be conveniently combined, e.g., by finding the parameters of quantal generators first and consequently designing an optimal set of signals.

[175] I.e., changing the scheme is a part of the optimization process, not the homotopical traversal.

For optimizing formulation OPT we consider using gradient based techniques that rely on computing the gradient

$$f'(\lambda, \beta_p(\lambda)) = \frac{\partial f(\lambda, \beta_p(\lambda))}{\partial \lambda} + \beta_p'(\lambda)^{\top} \frac{\partial f(\lambda, \beta_p(\lambda))}{\partial \beta_p}.$$

While the derivatives of the criterion function are easy to compute, the most challenging part is to estimate how the equilibrium shifts when we change the signaling scheme λ[176]. To this end, we may use the homotopy method, because in case we remember the intermediate results in the Gauss-Newton method, each step of the predictor-corrector method is differentiable with respect to the signaling scheme. By differentiating through the homotopy we hence approximate the gradient $\beta_p'(\lambda)$. The gradient is then used to perform a gradient ascent step projected on Λ as

[176] This corresponds to computing the derivative $\beta_p'(\lambda)$.

$$\lambda^{\iota+1} \leftarrow P_{\Lambda}\big(\lambda^{\iota} + \eta f'(\lambda, \beta_p(\lambda))\big),$$

where P_{Λ} denotes the projection and η is the learning rate. In practice, we do not compute the gradient from the whole homotopy run, as this proved to be excessively slow. Instead, we compute the gradient $\beta'_{p_\iota \to p_{\iota+1}}(\lambda)$ and perform the gradient ascent steps simultaneously with the homotopy traversal. In doing so, we per-

form a process akin to simulated annealing, which increases our chances of converging to a global optimum. The downside of this approach is that our homotopy continuously changes midway, affecting the shape of the principal branch. As a consequence, after each update of the signaling scheme, we have to refine the current point (γ, t) with respect to the changed curve. To this end, we may use the Gauss-Newton method again. The entire optimization procedure is depicted in Algorithm 5. The algorithm is given an initial learning rate η and its minimum value $\underline{\eta}$. After every iteration of the λ update, we perform an η-adaptation step using parameters Δ_a, Δ_r, both strictly smaller than 1. The parameter $\bar{\iota}$ then controls the number of predictor-corrector iterations performed. Note that in this context, the learning rate may be regarded as serving a similar purpose as steplength h in the homotopic predictor. Fortunately, we are still able to guarantee the existence of a neighborhood such that the Gauss-Newton method converges to a point on the changed curve.

Proposition 6.9. *Let $\Lambda \to (H : \mathbb{R}^{m+1} \to \mathbb{R}^m)$ be a correspondence of signaling schemes Λ to smooth homotopic functions, where each H has zero as a regular value. Let $f : \Lambda \times \Delta \to \mathbb{R}$ be a smooth function with bounded derivatives. Then for each $\lambda \in \Lambda, p \in J$, defined as in Theorem 6.1, $(\beta(p), t(p)) : H_\lambda((\beta(p), t(p))) = 0$, there exists sufficiently small η such that a Gauss-Newton sequence $\{\mathcal{N}^i((\beta, t))\}_{i=0}^{\infty}$ converges to a point $(\beta', t') : H_{\lambda'}((\beta', t')) = 0$, $\lambda' = P_\Lambda(\lambda + \eta f'(\lambda, \beta(p)))$.*

Proof. The main difference between the application of the Gauss-Newton method in tracing a branch of a homotopy and in the optimization procedure of Algorithm 5 lies in what is static and what moves. In tracing, we move a point using Euler's method and aim to converge back on the curve, while in the optimization, we move the signaling scheme λ, hence altering the curve, while the point remains static. The convergence for tracing is guaranteed because of the continuity of the Euler's method. When moving the curve, we make use of the continuity of quantal correlated equilibrium. According to Theorem 3.4.1 in the book of Allgower and Georg[165], there exists an open neighborhood

$$U, \{x \in \mathbb{R}^{m+1} : H_{\lambda'}(x) = 0\} \subset U,$$

such that every Gauss-Newton sequence starting in U converges to some x',

$$H_{\lambda'}(x') = 0.$$

Because the space of quantal correlated equilibria lying on some principal branch is compact and connected, the correspondence $\Lambda \to H$ is continuous. Moreover, as f has bounded continuous derivatives, there exists a sufficiently small η such that $(\beta(p), t(p))$ lies in the neighborhood U of the induced $H_{\lambda'}$. The Gauss-Newton sequence starting at $(\beta(p), t(p))$ hence converges to a zero of $H_{\lambda'}$. □

Now we turn to the question of optimality of the found solution. Proposition 6.4 claims that the set of quantal correlated equilibria is compact and connected, but

according to our empirical observations, the concept is hardly ever convex. Consequently, even if the criterion is concave, reaching a global maximum can not be guaranteed. Despite this fact, the experimental results presented in the following section show that the algorithm is often able to reach close to optimal solutions[177].

[177] Moreover, note that because of compactness, convergence to local optima is still guaranteed.

6.4 Empirical evaluation

We turn to the demonstration of the performance of the homotopy algorithm for quantal correlated equilibrium. We evaluate it using two metrics: (i) the runtime of the algorithm, and (ii) the quality of the found solutions. For both, we employ the BARON solver as a baseline to compare to. BARON is a commercial optimization solver for solving non-convex problems to global optimality, and is consistently regarded as the fastest and most robust solver[178]. In contrast, our implementation of the homotopy algorithm serves merely as a proof of concept and is done in Python 3 using the PyTorch library for computing the necessary gradients.

[178] According to results published at http://plato.asu.edu/ftp/minlp.html.

The implementation of Algoritm 3 is based on a general homotopy scheme described in the book of Allgower and Georg[165] in Appendix P3 which is further used also in Gambit Library's quantal response equilibrium solver of McKelvey, McLennan, and Turocy[174]. We set the parameters as $h = 0.35, \underline{h} = 10^{-8}, \bar{\iota}_{gn} = 100, \underline{\epsilon} = 10^{-4}, \bar{\epsilon} = 0.8, \epsilon_t = 10^{-4}, a_\kappa = 0.8, a_f = 0.8, a_\eta = 0.1, \eta = 0.8, \underline{\eta} = 10^{-5}, \Delta_a = 0.99, \Delta_r = 0.9$, and $\bar{\iota} = 10$. The initial λ was sampled uniformly randomly from the set of distributions over signal profiles. All experiments were performed on a computer with processor Intel(R) Xeon(R) W-2235 running at 3.80GHz, and 32GB RAM.

6.4.1 Experimental domains and their instance generation

The algorithm is domain independent, and we use two domains to evaluate its performance. The first domain are randomly generated games which serve to capture the expected performance of the algorithm over various classes of games with arbitrary utility structures. Since larger randomly generated games may exhibit undesired properties[179], we evaluate the algorithm also on a more structured domain. For this purpose we employ games inspired by supply-chain decision making and suppliers-retailers interaction.

[179] For instance, in Stackelberg games it may be easier to solve a random game than a game with specific structure.

Randomly generated normal form games

We construct general sum games with action spaces of different sizes for each player. When searching for an optimal signaling scheme, we consider two criteria: one linear and one quadratic. As the linear objective, we opt for social welfare[180] – a maximization of a sum of players' utilities – formally defined as

[180] We considered also another linear objective, a "Stackelberg-like" setting in which a selected player's utility is optimized, and we obtained comparable results in term of both scalability and quality of solutions.

$$welfare(\beta) = \sum_{i \in N} \sum_{a_i \in A_i} \sum_{s_i \in S_i} \beta_i(a_i|s_i) u_i(a_i|s_i).$$

The quadratic objective is a variant of Gini index; we aim to minimize absolute differences in players' utilities, formally:

$$gini(\beta) = \sum_{i \in N} \left(\frac{welfare(\beta)}{|N|} - \sum_{a_i \in A_i} \sum_{s_i \in S_i} \beta_i(a_i|s_i) u_i(a_i|s_i) \right)^2 .$$

We consider four different generators of generalized Luce models of quantal response functions: linear generator $q(x) = x + C$, quadratic generator $q(x) = (x + C)^2$, logarithmic generator $q(x) = \log(x + C)$ and exponential generator $q(x) = exp(x)$. We set C appropriately to ensure the induced quantal response functions are valid. The algorithm is capable of handling settings when each player has a different generator, and we verified that computing a quantal correlated equilibrium with various combinations of generators does not pose any unforeseen computational challenges. For the simplicity of presenting the results of the experiments, we focus on the setting when all players share the generator of the same kind. For each generator we employ its newtonian representation, as it proved to be the most robust in our initial exploration of the algorithm's settings. For the chosen utility range, setting $f_1(x) = x$ and $f_2(x, y) = xy$ is sufficient.

Supply chain games

In this game, the suppliers choose a warehouse to store a raw material, while the retailers choose a place to manufacture a good to sell at a market. Formally, each storage place is capable of storing one unit of a fixed raw material, and each manufacture produces one unit of a good from a fixed set of materials. Warehouses and manufactures are divided into mutually exclusive territories, and manufactures placed in a given territory are assumed to buy raw materials from the warehouses situated in the same territory exclusively. There are costs associated with running a supplier business: obtaining the raw material, shipping it to a warehouse and using the warehouse; and the profit stems from selling it to the nearby retailer. Similarly, the retailers have to pay for obtaining the raw materials and running the manufacture; and they profit from selling the good. The costs and prices are driven by the local market in the territory: we assume the warehouses are rented and the more suppliers decide to use the storage, the higher the price for usage. Similarly, the price of a raw material fluctuates depending on the supply and demand. For determining the prices we use a simple allocation algorithm, assuming closer manufactures are preferred over more distant when delivering a raw material. We define the game as a tuple $SC = (P, R, T, H, F, M, \tau_{pr}, \tau_{hf}, \delta, \mu_h, \mu_f, \zeta_b, \zeta_s, \zeta_h, \zeta_m, \upsilon, \rho)$, where P is a set of suppliers and R is a set of retailers. The set T then consists of different territories, H is a set of warehouses and F is a set of manufactures. M is a set of available raw materials. The function $\tau_{pr} : P \cup R \to 2^T$ assigns a supplier or a retailer a set of territories where they may legally operate. The function $\tau_{hf} : H \cup F \to T$ then identifies a territory where a given warehouse or manufacture is located. Func-

Algorithm 6: Material allocation and pricing

Input: action profile π
Output: facility utilization $\mathcal{A}$, material prices $\mathcal{Z}$
for $t \in T$ **do**
  $A_h \leftarrow [\pi_p \in \pi \mid p \in P,\ \tau_{hf}(\pi_p) = t]$
  $A_f \leftarrow [\pi_r \in \pi \mid r \in R,\ \tau_{hf}(\pi_r) = t]$
  while $|A_h| > 0$ *and* $|A_f| > 0$ **do**
    for $f \in A_f$ **do**
      $\overline{\delta}[f] \leftarrow 0, \quad \overline{P}[f] = [], \quad \overline{\mu} \leftarrow \mu_f \restriction f$
      $\overline{A}_h \leftarrow sort\_ascending(A_h, \delta \restriction f)$
      for $h \in \overline{A}_h$ **do**
        $\overline{\delta}[f] \leftarrow \overline{\delta}[f] + \delta(h, f), \quad \overline{P}[f] \leftarrow \overline{P}[f] \cup h$
        $\overline{\mu}(\mu_h(h)) := \overline{\mu}(\mu_h(h)) - 1$
        if $\sum_{m \in M} \overline{\mu}(m) = 0$ **then** break
      if $\sum_{m \in M} \overline{\mu}(m) > 0$ **then** $\overline{\delta}[f] \leftarrow \infty$
    $f^* \leftarrow arg\min_{f \in A_f} \overline{\delta}[f]$
    if $\overline{\delta}[f^*] = \infty$ **then** break
    $A_f \leftarrow A_f \backslash f, \qquad A_h \leftarrow A_h \backslash \overline{P}[f^*], \qquad \overline{\delta}[f^*] \leftarrow \infty$
  for $h \in H \mid \tau_{hf}(h) = t$ **do**
    $\mathcal{A}(h) \leftarrow \frac{|\{p \in P \mid \pi_p = h\}| - |\left[h' \in A_h \mid h' = h\right]|}{|\{p \in P \mid \pi_p = h\}|}$
  for $f \in F \mid \tau_{hf}(f) = t$ **do**
    $\mathcal{A}(f) \leftarrow \frac{|\{r \in R \mid \pi_r = f\}| - |\left[f' \in A_f \mid f' = f\right]|}{|\{r \in R \mid \pi_r = f\}|}$
  for $m \in M$ **do**
    $\mathcal{Z}(t, m) \leftarrow$
    $\zeta(t, m) - \frac{|[h \in A_h \mid \mu_h(h) = m]|}{|\{p \in P \mid \mu_h(\pi_p) = m, \tau_{hf}(\pi_p) = t\}|} - \frac{\sum_{f \in A_f} \mu_f(f, m)}{\sum r \in R,\, \tau_{hf}(\pi_r) = t \mu_f(\pi_r, m)}$

tion $\delta : H \times F \to \mathbb{R}$ serves to identify a distance between a manufacture and a warehouse. Raw materials are assigned to warehouses and manufactures using functions μ. First, function $\mu_h : H \to M$ specifies which material may be stored in a warehouse. Second, function $\mu_f : F \times M \to \mathbb{N}$ determines an amount of raw material to operate a manufacture. Functions ζ are associated with material costs. To obtain a raw material, the supplier pays a price $\zeta_b : T \times M \to \mathbb{R}$. A baseline selling price of a material to a retailer is determined by $\zeta_s : T \times M \to \mathbb{R}$, which is later amended by the allocation algorithm. A baseline storing cost is $\zeta_h : H \to \mathbb{R}$, and for moving the raw material to a warehouse the supplier pays a price given by function $\zeta_m : P \times H \to \mathbb{R}$. Function $\upsilon : F \to \mathbb{R}$ then identifies prices for manufacture usage, and function $\rho : F \to \mathbb{R}$ gives an expected profit of a retailer from selling a good. Given an action profile π of warehouses and manufactures chosen by the suppliers and retailers, the utility of the suppliers is defined as

$$u_p(\pi) = \mathcal{A}(\pi_p)\mathcal{Z}\Big(\tau_{hf}^{-1}(\pi_p), \mu_h(\pi_p)\Big) - \zeta_b\Big(\tau_{hf}^{-1}(\pi_p), \mu_h(\pi_p)\Big) - \zeta_m(p, \pi_p) - \zeta_h(\pi_p)|\{i \in P \mid \pi_i = \pi_p\}|,$$

and the utility of the retailers as

$$u_r(\pi) = \mathcal{A}(\pi_r)\rho(\pi_r) - \upsilon(\pi_r) - \sum_{m \in M} \mu_f(\pi_r, m)\mathcal{A}(\pi_r)\mathcal{Z}\left(\tau_{hf}^{-1}(\pi_r), m\right).$$

Here, $\mathcal{A}$ is the relative utilization of warehouses and manufactures, and $\mathcal{Z}$ is the material pricing computed by Algorithm 6. Furthermore, we assume there exists a central governing authority that aims to maintain a certain number of manufactures operational in each territory, specified by a function $t : F \to \mathbb{N}$, such that $\sum_{f \in F} t(f) = |R|$. An optimal signaling scheme should hence minimize the deviation from this assignment, which is formally expressed as

$$\sum_{f \in F} \left(t(f) - \sum_{\substack{r \in R \\ \tau_{hf}(f) \in \tau_{pr}(r)}} \sum_{s_r \in S_r} \beta_r(f|s_r) \sum_{s_{-r} \in S_{-r}} \lambda(s_r, s_{-r}) \right)^2.$$

In the experiments, we construct random supply chain games to access the algorithms' performance. Each supplier and retailer is assigned a non-empty subset of territories randomly from 2^T. In each territory, the number of warehouses and manufactures is random-generated from intervals $[2, 4]$ and $[1, 3]$, respectively. Each warehouse is assigned a random material it may store, and each manufacture requires a random amount of material to operate, each amount drawn from interval $[0, 2]$. The distance between a warehouse and a manufacture is an integer selected randomly from interval $[1, 4]$. The values of material cost functions ζ consist of integers drawn from interval $[1, 3]$ for ζ_b, $[4, 6]$ for ζ_s, $[1, 3]$ for ζ_h, and $[1, 4]$ for ζ_m. The manufacture use prices determined by function υ are random integers from interval $[1, 3]$, and the expected profit given by function ρ is always in interval $[6, 9]$. Similarly as in randomly generated normal-form games, we assume that all players share the generator of the same kind and we use its corresponding exponential representation as it outperformed the newtonian in this domain. We set the constant of the generators to $C = 15$ and $f_1(x) = x$, $f_2(x, y) = xy$. The supply-chain game may be conveniently modeled as an action-graph game, and as such, the expected utilities of its strategy profiles can be computed in polynomial time using a trie-based algorithm.[181]

[181] A. X. Jiang, K. Leyton-Brown, and N. A. Bhat. "Action-graph games". *Games and Economic Behavior* 71:1, 2011, pp. 141–173.

6.4.2 Experimental results

For each domain, we tested three settings defined by a number of players, number of signals, and in case of the supply chain game also by a number of materials. First setting is used for principle branch tracing only, the second for comparing the quality of optimization on smaller games, and the third for optimization beyond BARON's capabilities. Each domain has one parameter controlling a game's action space used to illustrate scalability. For randomly generated games it is a number of

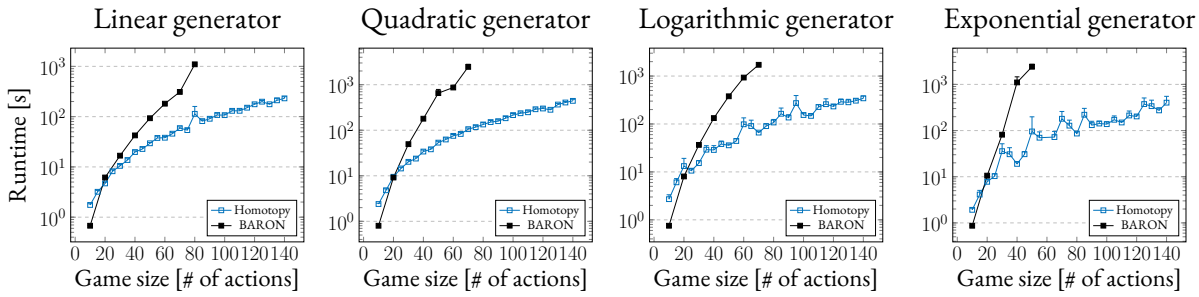


Figure 6.3: Mean runtimes of computing (S-QCE-NFG) with fixed λ using BARON and the homotopy algorithm in randomly generated games. Every point shows also a standard error.

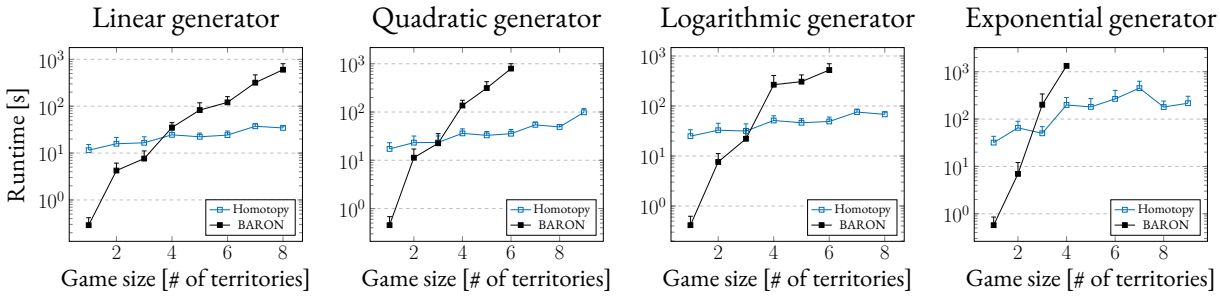


Figure 6.4: Mean runtimes of computing (S-QCE-NFG) with fixed λ using BARON and the homotopy algorithm in supply chain games. Every point shows also a standard error.

actions per player, while in supply chain games it is a number of territories in the game. Moreover, we assessed also scalability with respect to the number of signals. For each combination of setting $\times$ game size $\times$ generator function, we constructed 10 game instances per domain.

First, we access the algorithms' capability to reach quantal correlated equilibrium with fixed signaling scheme. For the homotopy this corresponds to how fast it is able to move along the principal branch. In the second part, we present results of searching for a signaling scheme optimizing a given criterion over the set of quantal correlated equilibria.

Scalability in actions for computing the equilibrium

The line of graphs in Figure 6.3 shows the results achieved in randomly generated games. We consider three-player games with 2 signals per player. The x-axis varies the number of actions of the first two players, the third player makes only binary decisions, while the y-axis shows the runtimes of the algorithms. Every point in the graphs corresponds to the mean over the sampled instances and shows the achieved standard error. We terminated all still running seeds after 45m and we depict them in the graphs with this lower bound on runtime if the computation was still on-

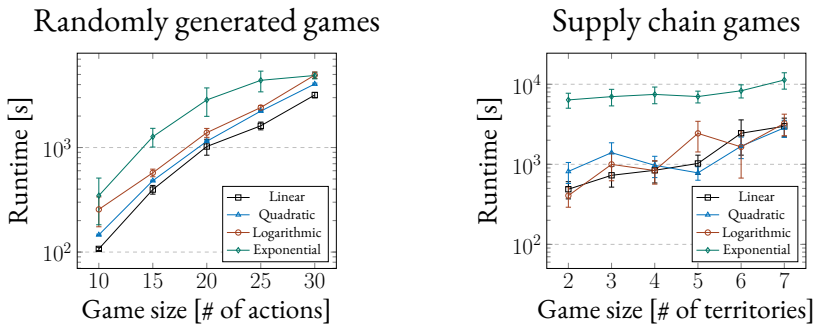


Figure 6.5: Scalability of the homotopy method across different generators in (Left) randomly generated games and (Right) supply chain games. Every point shows also a standard error.

going. As the graphs show, despite the overhead of the homotopic algorithm on smaller instances, it scales better than BARON as the game size increases.

The results on supply chain games are depicted in the line of graphs in Figure 6.4. We assume there are 3 suppliers, 2 retailers and 2 materials. Each retailer receives 1 of 2 possible signals, while the suppliers observe only a single trivial signal. The graphs follow the same format as in randomly generated games, and we observe a similar behavior. However, the difference in scalability is even more profound. Formally, the algorithm minimizes a function $|t - 1|$ for $((\beta, t), \lambda)$ along the principal branch, using a Newton-type steplength adaptation that guarantees superlinear convergence.[182] The same convergence holds also for other special points of interest on the branch, i.e., zero or extremal points of a smooth functional $c(p) \to \mathbb{R}$[183].

Finally, in Figure 6.5 we present the scalability results of different generators on larger game instances. Similarly as before, the results were averaged over 10 games. The algorithms were given 4 hours to compute the results. We examined three-player randomly generated games with 5 signals per player and supply chain games 4 suppliers, 3 manufacturers, 2 materials, and 5 signals per manufacturer. Other parameters remained as in the main text. BARON ran out of the 32GB of memory already on the smallest instances.

[182] M. Avriel. *Nonlinear programming: Analysis and methods*. Courier Corporation, 2003.

[183] We may hence expect similar scalability as presented here also for other criteria, e.g., maximum likelihood estimation along the path.

Scalability in signals for computing the equilibrium

In randomly generated normal form games, we assumed a setting with 3 players choosing from 20, 20, and 2 actions, respectively. The number of signals per each player was $\#s$, $\#s$, and 2, respectively, where $\#s$ denotes the number of signals and it is the parameter affecting the game's size. The results are depicted in Figure 6.6. Similarly as when increasing the number of available actions, the homotopy method performs notably better than BARON, especially with the logit generator.

In supply chain games, we assumed there are 3 suppliers, 2 manufacturers, and a single material. We fixed the number of territories to 3. Similarly as in the experiments with increasing size of action spaces, we assume only the manufacturers receive one of the $\#s$ signals each. The results could be found in Figure 6.7. Contrary

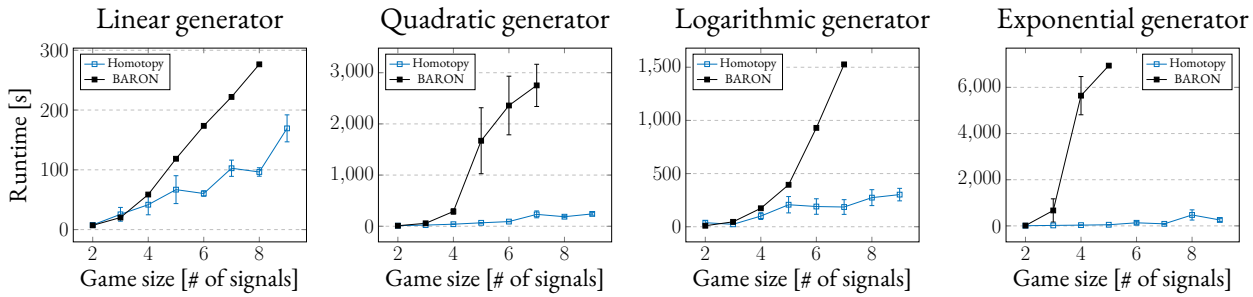


Figure 6.6: Mean runtimes of computing (S-QCE-NFG) with fixed λ using BARON and the homotopy algorithm in randomly generated games. Every point shows also a standard error.

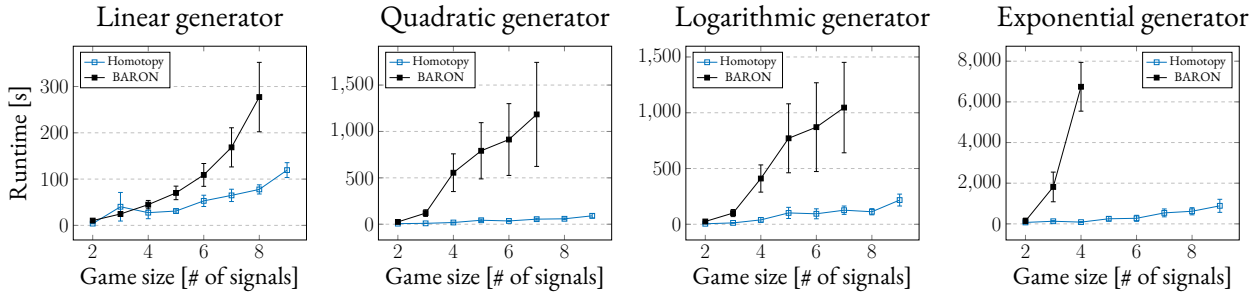


Figure 6.7: Mean runtimes of computing (S-QCE-NFG) with fixed λ using BARON and the homotopy algorithm in supply chain games. Every point shows also a standard error.

to experiments varying the action spaces, in these experiments, even the smallest games do not give BARON any significant advantage over the homotopy method. The homotopy continues to outperform BARON as the game size increases. Interestingly, BARON's runtimes seem to exhibit much larger deviations than when the number of signals was fixed.

Finding optimal signaling scheme

The runtimes and relative errors of solutions computed while optimizing a signaling scheme in randomly generated games are presented in Tables 6.1 (maximizing social welfare) and 6.2 (minimizing Gini index). We consider small two-player square games where the first player has 2 signals while the second player receives only a single trivial signal, otherwise BARON would not scale beyond the smallest games. Each table shows mean runtimes of both algorithms and deviations Δ that correspond to the mean difference in the solutions' criteria values computed using the homotopy and BARON. The almost non-existent deviations suggest that homotopy reached (close-to) optimal solutions, and we omit the standard errors as they are negligible. We do not observe any obvious trend with increasing game size with the exception of the exponential generator where the solution's quality

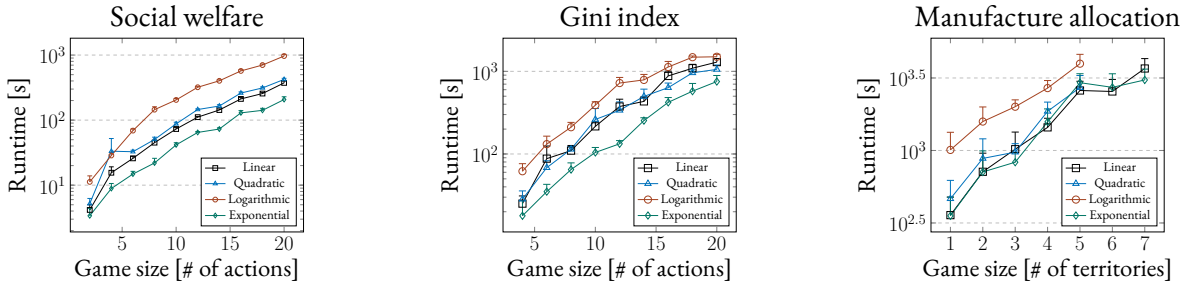


Figure 6.8: The runtimes of different generators in larger games. With the exception of the deviations, each value shows also a standard error.

	Homotopy [s]	BARON [s]	deviation
Quadratic generator			
2 actions	2±0	0±0	$-6 \cdot 10^6$
3 actions	3±0	5±1	$-8 \cdot 10^6$
4 actions	6±1	23±3	$-6 \cdot 10^4$
5 actions	8±1	345±240	$-1 \cdot 10^5$
6 actions	11±0	481±191	$-6 \cdot 10^6$
7 actions	18±1	1052±212	$-8 \cdot 10^6$
8 actions	23±3	1619±371	$-8 \cdot 10^6$
Logarithmic generator			
2 actions	4±0	0±0	$-7 \cdot 10^6$
3 actions	12±3	1±0	$-2 \cdot 10^5$
4 actions	12±2	1±0	$-5 \cdot 10^6$
5 actions	19±2	3±0	$-6 \cdot 10^6$
6 actions	25±0	10±3	$2 \cdot 10^2$
7 actions	35±1	11±2	$-3 \cdot 10^6$
8 actions	45±1	61±43	$-3 \cdot 10^6$
Exponential generator			
2 actions	1±0	1±0	$1 \cdot 10^1$
3 actions	2±0	15±3	$-6 \cdot 10^2$
4 actions	2±0	208±62	$-7 \cdot 10^3$
5 actions	3±0	898±191	$-8 \cdot 10^3$

Table 6.1: The results of searching for a signaling scheme optimizing social welfare in randomly generated games. Comparison of the runtimes of the homotopy algorithm and BARON, and deviations from global solutions in smaller games.

	Homotopy [s]	BARON [s]	deviation
Linear generator			
2 actions	2±0	0±0	$2 \cdot 10^2$
3 actions	7±1	0±0	$1 \cdot 10^2$
4 actions	15±4	0±0	$6 \cdot 10^4$
5 actions	16±3	7±2	$2 \cdot 10^3$
6 actions	35±9	169±139	$2 \cdot 10^3$
7 actions	126±85	267±204	$2 \cdot 10^3$
Quadratic generator			
2 actions	3±0	0±0	$-3 \cdot 10^7$
3 actions	7±0	9±1	$-2 \cdot 10^7$
4 actions	12±2	1266±472	$4 \cdot 10^7$
Logarithmic generator			
2 actions	6±1	0±0	$-2 \cdot 10^6$
3 actions	14±4	2±0	$-1 \cdot 10^6$
4 actions	27±6	13±6	$-9 \cdot 10^7$
5 actions	44±11	45±15	$-2 \cdot 10^3$
6 actions	61±12	159±50	$-8 \cdot 10^7$
7 actions	75±15	1156±422	$-4 \cdot 10^6$
Exponential generator			
2 actions	1±0	1±0	$-3 \cdot 10^3$
3 actions	3±0	51±13	$-1 \cdot 10^3$
4 actions	17±13	1191±466	$-7 \cdot 10^5$

Table 6.2: The results of searching for a signaling scheme optimizing Gini index in randomly generated games: comparison of runtimes, deviations, and generators' scalability. The figure follows the same format as Table 6.1.

seems to improve. This may be a consequence of mitigation of exponential steepness due to quantal-response averaging with games getting larger. For social welfare, BARON fails to compute solutions with linear generator even for the smallest games, which seems to be a consequence of numerical instabilities. In Figure 6.8 we present scalability comparison of different generators in two-player square games with 2 signals per player. The results indicate that quantal correlated equilibria with logarithmic generators consistently take the longest to compute, while the the most common logistic exponential generator is among the fastest.

The results in supply chain games can be found in Table 6.3. The data in the table describe the runtimes and deviations in games with 2 suppliers, 1 retailer and 1 material. We observe similar patterns as in randomly generated games: homotopy scales significantly better than BARON, while maintaining comparable quality of solutions. Again, in Figure 6.8 we compare scalability of individual generators in games with 3 suppliers, 2 retailers and 2 materials. As expected, the logarithmic one per-

	Homotopy [s]	BARON [s]	deviation
Linear generator			
1 territory	15±4	5±3	$-2 \cdot 10^7$
2 territories	23±4	525±411	$-3 \cdot 10^7$
3 territories	74±33	2746±995	$-8 \cdot 10^8$
4 territories	209±154	2726±1084	$-4 \cdot 10^8$
Quadratic generator			
1 territory	20±5	799±745	$-6 \cdot 10^7$
2 territories	32±8	3734±1193	$7 \cdot 10^7$
3 territories	62±14	4866±938	$-4 \cdot 10^5$
Logarithmic generator			
1 territory	58±23	2±1	$-6 \cdot 10^7$
2 territories	67±14	72±52	$-5 \cdot 10^7$
3 territories	179±47	756±694	$-1 \cdot 10^6$
4 territories	261±74	1084±758	$-1 \cdot 10^4$
Exponential generator			
1 territory	5±1	1167±730	$-1 \cdot 10^2$
2 territories	21±8	2990±1097	$-5 \cdot 10^3$
3 territories	60±17	5665±890	$-1 \cdot 10^4$

Table 6.3: The results of searching for a signaling scheme optimizing manufacture allocation in supply chain games: comparison of runtimes, deviations, and scalability. The figure follows the same format as Table 6.1.

forms the worst, while the other three generators remain almost indistinguishable, which is consistent with the results in the table.

❖

Overall, the results suggest that the homotopy method is a viable option for computing the equilibrium in terms of both scalability and quality of solutions. However, finding an equilibrium with fixed signaling scheme is significantly easier than optimizing a scheme. This indicates that computing the gradient more efficiently may significantly improve scalability.

6.5 Summary of contributions

We initiated an investigation of quantal response in correlated equilibrium. We consider generalized Luce models of quantal response that enables us to induce different quantal behavior for each player, tailored to specific behavioral profiles. We introduced two ways of including quantality while conditioning players' strategies on signals received from a correlation device – either per each signal separately, or over

the whole set of pure strategies in the extended game. We argued that psychological studies favor the first interpretation and therefore we focused predominantly on it.

In the theoretical part, we verified the equilibrium meets the expectations in terms of its relation to quantal response and correlated equilibria. We examined the solution's complexity and proved it remains PPAD-hard; and showed that coordinating the players using signals may be beneficial for the signaler as their utility becomes strictly greater than in quantal response equilibrium.

In the algorithmic part, we developed a homotopy approach increasing robustness of computation using multiple techniques: we eliminated the normalization sum in Luce models, we reformulated the product of strategies using product separating functions, and we simultaneously trace the equilibrium and optimize the signaling scheme while maintaining the convergence guarantees of the Gauss-Newton method. Empirical results show the homotopy is consistently faster[184] than the state-of-the-art solver BARON and provides competitive solutions.

[184] Up to 300-times.

7 Quantal Correlated Equilibrium in Extensive Form Games

Extending techniques and results from one-shot to sequential scenarios is non-trivial work, frequently associated with increased technicalities and computational complexity. Sequential solution concepts are also more nuanced, requiring careful treatment of the underlying information structure and players' hindsight rationality. For this reason, players' behavior in extensive form games can not often be easily translated to their normal form representation, especially in interactions among boundedly rational players.

In Chapter 5, we witnessed an example of this phenomenon. Boundedly rational strategies in normal form quantal Stackelberg equilibrium hinge on the players' myopic perspectives and introducing sequentiality into their decision making processes results in a vastly different form of behavior. Consequently, computing the extensive form analog of the equilibrium necessitates new and significantly more complex methods. Here, we aim to do likewise for the quantal correlated equilibrium we investigated in the previous chapter. Instead of sending signals only once at the beginning of the game, we wish to coordinate the players throughout a prolonged period of time during which they interact repeatedly.

Because correlated equilibrium is conceptually similar to Nash equilibrium, we hoped to build upon techniques for computing quantal response equilibrium in extensive form games. Methods therein are based on Karush-Kuhn-Tucker conditions[185] or regret minimization,[186] and promise favorable scalability up to games of moderate sizes. Unfortunately, the approach of these works relies heavily on a logit quantal response reformulation in two-player zero sum games and collapses beyond this class. For this reason, we formulate another homotopy method instead, with a substantially redesigned corrector emanating from a conjugate gradient method. The new corrector allows us to sidestep the significant computational drawback of the Gauss-Newton method- the Jacobian matrix's repeated inversion.

We begin this chapter by constructing the quantal correlated equilibrium in extensive form games from its two precursor equilibria, the extensive form correlated equilibrium and the agent quantal response equilibrium, drawing from the ideas of both concepts. Due to the construction, finding the equilibrium continues to be PPAD-hard. To compute it, we derive a homotopic formulation of this new equilibrium and formally introduce a non-linear conjugate gradient method to act as a corrector in the tracing algorithm. A vital component of this method is a line search

The results in this chapter are a part of a preprint available under J. Černý, B. An, and A. N. Zhang. *Quantal correlated equilibrium in extensive form games*. Technical report. 2023.

[185] C. K. Ling, F. Fang, and J. Z. Kolter. "What game are we playing? End-to-end learning in normal and extensive form games". In: *Proceedings of the 27th International Joint Conference on Artificial Intelligence*. 2018, pp. 396–402.

[186] G. Farina, C. Kroer, and T. Sandholm. "Online convex optimization for sequential decision processes and extensive-form games". In: *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*. 2019, pp. 1917–1925.

function. In the last part of the chapter, we show that by employing a specific class of hyperbolic quantal responses, the problem of solving the line search problem reduces to maximizing a polynomial function.

7.1 Problem definition

Similarly, as in the previous chapter, we formulate the quantal correlated equilibrium in extensive form games using the standard construction. We assume there is a lottery mechanism, which distribution we will again call a signaling scheme, that gives rise to an equilibrium in the extended game. The mechanism reveals signals to the players privately, sampling them from some public distribution over signals. Formally, for an extensive form game G, assume that for each information set I in G we have possible signals S_I. A complete signal of player i is a tuple of one signal per each information set in which i acts, i.e.,

$$(s \in S_I)_{I \in I_i} \in \prod_{I \in I_i} S_I.$$

We denote the Cartesian set $\prod_{I \in I_i} S_I$ as S_i. Signal profile in the whole game is then a tuple of one signal per each player, i.e.,

$$(s_i \in S_i)_{i \in N} \in \prod_{i \in N} S_i.$$

Again, we denote the Cartesian set $\prod_{i \in N} S_i$ as S. A signal distribution characterizing a signaling scheme is then a probability distribution $\lambda \in \Lambda$ over S.

For how the signals are revealed, we take inspiration from the established definition of correlated equilibrium in extensive form games.[187] We presume the lottery mechanism sends a signal to a player just at the moment the player reaches an information set associated with the signal. The player then makes strategic decisions in the information set as in the agent quantal response equilibrium.[188] In other words, when rationalizing their local behavioral strategy, we assume they treat their other selves acting in other information sets as independent players with a known behavior[189]. The extended signaling game's behavioral strategies are conditioned on signals that may be received in the information sets as $\beta_i(a_i|s_i)$ for $a_i \in \chi(I), s_i \in S_I$, for some $I \in I_i$. From the perspective of player i who received a signal s_i in one of their information sets, the probability of player j taking action a_j in some information set $I_{j,k} \in I_j$ is then

$$p(a_j|s_i) = \sum_{s_j \in S_{I_{j,k}}} \lambda(s_j|s_i)\beta_j(a_j|s_j),$$

where $\lambda(s_j|s_i)$ is the probability that the correlation device samples signal s_j given that player i received signal s_i in I. The probability of being located in some game

[187] B. von Stengel and F. Forges. "Extensive-form correlated equilibrium: Definition and computational complexity". *Mathematics of Operations Research* 33:4, 2008, pp. 1002–1022.

[188] R. D. McKelvey and T. R. Palfrey. "Quantal response equilibria for extensive form games". *Experimental Economics* 1:1, 1998, pp. 9–41.

[189] See the original papers for motivation for such behavior.

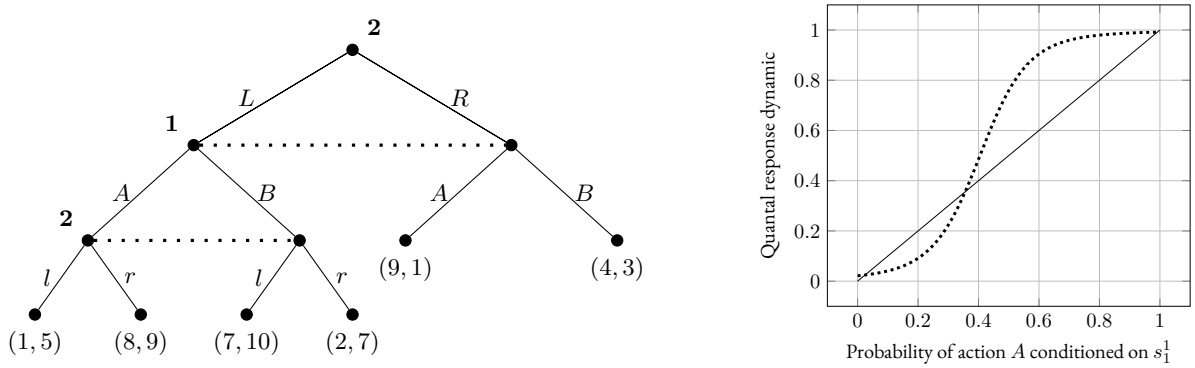


Figure 7.1: Quantal correlated equilibrium in an extensive form game. (Left) A game tree representation of a sequential game. The figure follows a standard denotation of extensive form games. (Right) An extensive form quantal response dynamic with correlated strategies. Each player acts with a generator $exp(x)$. The equilibrium is the point $\beta_1(A|s_1^1) \approx 0.36$, intersecting the quadrant's axis.

state $h \in I$ may be then further evaluated as

$$p(h|s_i) = \frac{\prod_{a_j \in seq_{-i}(h)} p(a_j|s_i)C(h)}{\sum_{h' \in I} \prod_{a_j \in seq_{-i}(h')} p(a_j|s_i)C(h')}.$$

The utility player i expects after playing action a_i in I is hence

$$u_i(a_i|s_i) = \sum_{\substack{h \in I \\ z \in Z(h,a_i)}} \prod_{a_j \in seq(ha_i \to z)} p(a_j|s_i)p(h|s_i)C(h \to z)u_i(z), \quad (7.1)$$

where $Z(h, a_i)$ is a set of leaves in a subtree rooted in state h after taking action a_i, $seq(ha_i \to z)$ is a sequence of actions from state h to z after taking action a_i, and $C(h \to z)$ is a probability of reaching leaf z from h due to Nature.

As in normal form games, we consider generalized Luce quantal response functions and assume the players behave according to their generators $(q_i)_{i \in N}$ after they receive a signal. In extensive form games, they received their signals in each information set separately.

Definition 7.1. *Let G be an extensive form signaling game with signals S. The behavioral strategies $(\beta_i)_{i \in N}$, $\beta_i \in B_i$ and a signaling scheme $\lambda \in \Lambda$ form a quantal correlated equilibrium if for every player i and their each information set $I \in I_i$, action $a_i \in \chi(I)$ and signal $s_i \in S_I$,*

$$\beta_i(a_i|s_i) = \frac{q_i(u_i(a_i|s_i))}{\sum_{a_i' \in \chi(I)} q_i(u_i(a_i'|s_i))}, \quad \text{(QCE-EFG)}$$

where $u_i(a_i|s_i)$ is defined as in Equation 7.1.

		Player 1	
		s_1^1	s_1^2
Player 2	s_2^1	0.2	0.4
	s_2^2	0.3	0.1

Figure 7.2: A correlation device λ for the game in Figure 7.1.

Example 7.1. *Consider the extensive form game depicted on the left in Figure 7.1. Assume that in this game, both players behave according to a logit generator* $q(x) = exp(x)$*. Player 1 may receive two possible signals* s_1^1 *and* s_1^2*. Player 2 receives only a single trivial signal in the root and either of the signals* s_2^1 *and* s_2^2 *in their bottom information set. Let the signals be correlated according to a signaling scheme depicted in Figure 7.2. Moreover, let the conditional strategy of player 1 taking action A after seeing signal* s_1^2 *be* $\beta_1(A|s_1^2) = 0.46$*. The graph on the right in Figure 7.1 shows the quantal response dynamic, beginning from the strategy* $\beta_1(A|s_1^1)$*. The only equilibrium here is reached for* $\beta_1(A|s_1^1) \approx 0.36$*. In the other intersecting points, the corresponding* $\beta_1(A|s_1^2)$ *is not equal to* 0.46*.*

Since Equation 7.1 naturally extends the definition of utility in extensive form games into a setting with a signaling lottery mechanism, similarly as in normal form games, the sequential analog of quantal correlation inherits many of its properties from this class of games. Most importantly, computing it remains PPAD-hard in general. Further on, we hence focus on designing a method for finding the equilibrium more efficiently.

7.2 Homotopy method for finding the equilibrium

Our construction of a homotopy method for quantal correlated equilibrium in extensive form games may follow the same path as in normal form games. Given parametric representations of generators of Luce models $(\hat{q}_i)_{i\in N}$, the equilibrial formulation (QCE-EFG) translates to a homotopy function $H(x,t) : \mathbb{R}^{m+1} \to \mathbb{R}^m$ with a parameter t, defined as follows:

$$H(\beta,t) = \left(H_i^{j,k,l}(\beta,t)\right)_{i\in N,\ I_{i,j}\in I_i,\ a_i^k\in\chi(I_{i,j}),\ s_i^l\in S_I}$$
$$H_i^{j,k,l}(\beta,t) = \hat{q}_i(u_i(a_i^k|s_i^l),t) - \beta_i(a_i^k|s_i^l)\sum_{a_i\in A_i}\hat{q}_i(u_i(a_i|s_i^l),t),$$

where the utility has the form of Equation 7.1. Solving the system for $H(x,t) = 0$ continues to be simple even with many non-trivial information sets, again, yielding uniformly random behavioral strategies in every information set. The system could be further reformulated as in Section 6.3, and the solutions may be traced along the unique homotopic path via a predictor-corrector method in the same manner as with more general systems[190]. The issue arising in extensive form games is that for computing the derivative c' and the Moore-Penrose inverse, we need to perform the computationally demanding QR decomposition of an enormous Jacobian matrix that becomes less and less sparse as we move in the information set hierarchy closer to the tree root[191]. Instead, we use a derivative-free quantal response as a predictor and design a conjugate gradient method to serve as a tailored corrector.

[190] E. L. Allgower and K. Georg. *Numerical continuation methods: An introduction*. Vol. 13. Springer Science & Business Media, 2012.

[191] Our exploratory evaluation confirmed this approach becomes unfeasible fairly quickly, even for smaller game trees with lower tens of information sets.

7.2.1 Tracing the equilibrial correspondence path

Nonlinear conjugate gradient methods are iterative algorithms for solving unconstrained optimization problems in a form

$$\min_{x \in \mathbb{R}^m} f(x),$$

where $f : \mathbb{R}^m \to \mathbb{R}$ is a continuously differentiable real-valued function that is bounded from below.[192] Conjugate gradient methods then generate a sequence of candidate solutions $x_k, k \geq 1$ from an initial guess x_0 via a recurrent equation formulated as follows:

$$x_{k+1} \leftarrow x_k + \tau_k d_k,$$

where d_k is also computed recursively as

$$d_{k+1} \leftarrow -\frac{\partial f(x_{k+1})}{\partial x} + \zeta_k d_k$$

from an initial $d_0 = -\partial f(x_0)/\partial x$, ζ_k is a conjugate gradient update parameter, and τ_k is a solution of a line search problem defined as

$$\tau_k = \arg\min_{\tau \geq 0} f(x_k + \tau d_k).$$

Many choices exist for specifying the update parameter ζ, leading to different guarantees of convergence conditioned on the exact solvability of the line search problem or the quality of its approximate solutions. For our needs, it suffices to know that global convergence can be achieved.[193]

[192] W. W. Hager and H. Zhang. "A survey of nonlinear conjugate gradient methods". *Pacific journal of Optimization* 2:1, 2006, pp. 35–58.

[193] G. Zoutendijk. "Nonlinear programming, computational methods". *Integer and Nonlinear Programming*, 1970, pp. 37–86.

Theorem 7.1. *Let the conjugate gradient update parameter be defined as*

$$\zeta_k = \frac{\|\partial f(x_{k+1})/\partial x\|}{\|\partial f(x_k)/\partial x\|},$$

where $\|\cdot\|$ is the Euclidean norm, and let the line search be solvable exactly. Then the sequence of candidate solutions x_k generated by the corresponding conjugate gradient method converges to a global minimum of function f.

❖

Now let us consider function f_t in a form

$$f_t(\beta) = \frac{1}{2}\|H(\beta, t)\|^2.$$

The global minimum of f_t is a solution of the system $H(x, t) = 0$, lying on the homotopic path for some value of t. Any globally convergent conjugate gradient method may hence act as a corrector in a homotopic algorithm. For a predictor,

Algorithm 7: Predictor-corrector method for tracing the extensive form quantal correlated equilibrium correspondence along the homotopic path

Input: $H, (\beta, t)$ such that $H(\beta, t) = 0$
Parameters: $\epsilon_t, \epsilon_c, \bar{\epsilon}_c, \bar{\iota}, \bar{k}$
$accept \leftarrow True, \quad \iota \leftarrow 0, \quad t, \bar{t} \leftarrow 0$
while $t < (1 + \epsilon_t)$ ***and*** $\iota < \bar{\iota}$ **do**
  $\bar{t} \leftarrow UpdateStep(t, \bar{t}, accept)$
  $\overline{\beta}_0 \leftarrow QuantalResponse(\beta, t, \bar{t})$ // prediction
  $d \leftarrow -H(\overline{\beta}, \bar{t})$
  $accept \leftarrow True, \quad k \leftarrow 0$
  while $\|H(\overline{\beta}, \bar{t})\| > \epsilon_c$ **do** // correction
    $\tau \leftarrow LineSearch(\overline{\beta}_k, \bar{t}, d)$
    $\overline{\beta}_{k+1} \leftarrow \overline{\beta}_k + \tau d$
    $\zeta \leftarrow GetZeta(\overline{\beta}_{k+1}, \overline{\beta}_k, d)$
    $d \leftarrow -H(\overline{\beta}_{k+1}, \bar{t}) + \zeta d$
    $k \leftarrow k + 1$
    if $\|H(\overline{\beta}, \bar{t})\| > \bar{\epsilon}_c$ ***or*** $k > \bar{k}$ **then** $accept \leftarrow False,$ break
  if $accept$ **then** $\beta \leftarrow \overline{\beta}_k, \quad \iota \leftarrow \iota + 1, \quad t \leftarrow \bar{t}$
return (β, t)

consider a partial solution $(\beta, t)^\iota$ that lies on the curve in iteration ι. For a parametric representation of a generator $\hat{q}_i$ of player i and some $t^{\iota+1} > t^\iota$, let us compute the prediction from $(\beta, t)^\iota$ as

$$\overline{\beta}_i^{\iota+1}(a_i|s_i) \leftarrow \frac{\hat{q}_i(u_i(a_i|s_i), t^{\iota+1})}{\sum_{a'_i \in \chi(I)} \hat{q}_i(u_i(a'_i|s_i), t^{\iota+1})},$$

[194] In other words, each player quantal responds to the strategies of their opponents from the previous iteration.

for every $I \in I_i, a_i \in \chi(I)$, and $s_i \in S_I$[194]. The entire homotopy algorithm for tracing the extensive form quantal correlated equilibrium correspondence is depicted in Algorithm 7. Here, the $QuantalResponse$ function calculates the prediction as described, while the three other functions can be instantiated with standard or domain-dependent predictor-corrector and conjugate gradient methods. More specifically, the $LineSearch$ function finds the optimal τ, $GetZeta$ identifies the update parameter, and $UpdateStep$ selects the new value of the homotopic parameter. The algorithm's constants – ϵ's, $\bar{\iota}$, and $\bar{k}$ – then serve to detect numerical convergence issues and enable to act accordingly.

We provide a simple example of Algorithm 7's function instantiation. As the $UpdateStep$ method we may employ a straightforward increase in t by $\epsilon > 0$ as

$$\bar{t} = \begin{cases} t + \epsilon & \text{if } accept, \text{ and} \\ t + \frac{\bar{t} - t}{2} & \text{otherwise.} \end{cases}$$

The $GetZeta$ method could be then calculated as in Theorem 7.1. Consider the following class of Luce models for the definition of the $LineSearch$ method.

Definition 7.2. *A generator q of a generalized Luce model is called hyperbolic if*

$$q(x) = \frac{C_1}{C_2 - x},$$

where x belongs to some open interval M of real numbers, and C_1 and C_2 are some real constants such that C_1 is positive and C_2 upper bounds M.

Hyperbolic generators have a representation that approximates a parametric one. Let the constant C_2 be defined for a homotopic parameter t as

$$C_2 = \frac{sup(M)}{t}.$$

Then the corresponding quantal response is uniform as t approaches zero and converges to the best response if M covers the expected utilities tightly as t tends to one. Even more importantly, the line search problem then has a favorable formulation.

Proposition 7.1. *Let G be an extensive form signaling game and let all players behave according to hyperbolic generators of generalized Luce models. Finding the optimal τ in the line search problem then reduces to finding a root of a polynomial.*

Proof. Let us reformulate the homotopy function for quantal correlated equilibrium in extensive form games using a reference action $a_0 \in \chi(I)$ in each information set I, similarly as in normal form games, as

$$H_i^{j,0,l}(\beta, t) = 1 - \sum_{a_i^k \in \chi(I)} \beta_i(a_i^k | s_i^l). \tag{7.2}$$

$$H_i^{j,k,l}(\beta, t) = \hat{q}_i(u_i(a_i^k | s_i^l), t)\beta_i(a_i^k | s_i^l) - \hat{q}_i(u_i(a_i^0 | s_i^l), t)\beta_i(a_i^0 | s_i^l) \tag{7.3}$$

When generators are hyperbolic, the expression (7.3) has a form

$$\frac{1}{C_1}\Big(C_2(\beta_i(a_i^0 | s_i^l) - \beta_i(a_i^k | s_i^l)) + u_i(a_i^k | s_i^l)\beta_i(a_i^k | s_i^l) - u_i(a_i^0 | s_i^l)\beta_i(a_i^0 | s_i^l)\Big).$$

Then, when solving the line search problem, the variables β are substituted with

$$\beta = \overline{\beta} + \tau d, \tag{7.4}$$

where $\overline{\beta}$ and d are constants, and the only variable is τ. The function f_t in the line search is a sum of squares, where each term is a square of either expression (7.2) or the hyperbolic reformulation of expression (7.3) with substitution (7.4). Since expression (7.2) is linear in τ, its square is quadratic. Now consider the definition of expected utility in Equation 7.1. We distinguish between two special cases.

1. The game is either an extensive form game with full information or an extensive form representation of a normal form game. Then the expected utility need not be normalized, and each term in the function describing the line search problem has a degree at most twice the depth of the tree.

2. If the game contains a non-trivial information set, the expected utility needs to be normalized[195]. In that case, the line search problem has a form of a sum of polynomial fractions. Since the denominators are squares and hence always non-negative, we may add the fractions and use the Dinkelbach reformulation again to arrive at a polynomial expression.

We obtain the result after applying the first-order condition for optimality. Moreover, note that in the first case, the degree of the polynomial is totally independent of the number of actions and signals in the game. □

[195] An alternative would be not to normalize the expectation as a part of the players' alleged boundedly rational behavior and tendency to weight the outcome by the probability of reaching the information set. Such a modeling decision would not impact the asymptotic behavior of the equilibrium as quantal responses approach the best response.

7.3 Summary of contributions

We generalized our investigation of quantal correlation from normal form to extensive form games. Similarly, as in previous chapters, we consider generalized Luce models of quantal response. We provide a definition of the equilibrium in sequential games that enables us to tailor the quantal behavior not just to individual players but also to separate information sets.

The major drawback of the homotopy method introduced in the context of normal form games is its dependence on the computationally demanding inversion of the Jacobian matrix. With many non-trivial information sets, this issue becomes even more profound in extensive form games. Our main theoretical result in this chapter is the reformulation of the homotopy into a setting with hyperbolic quantal response functions. This allows us to leverage the iterative conjugate gradient methods and thus sidestep the inefficient inversion, despite the fact that the equilibrium remains PPAD-hard in general.

Part IV

Epilogue

8 Conclusion

This dissertation studies solution concepts for strategic decision making and their algorithmic complexity and attainability, particularly in the context of adversarial cognitively restricted agents acting in informationally complex environments. The presented research spans formal analyses of foundational models, designs of computational methods, as well as empirical demonstrations of their performance on selected domains.

The driving motivation of this work is that our choices, as humans, are made in dynamic environments where outcomes are affected by a great number of actors and depend on their mutual rationality and personal knowledge. For instance, the performance of democracies is substantially influenced by voters' (lack of) information awareness and rationality. Yet more than simply accounting for information supplied by various media and service outlets is needed; the single agent machine learning techniques alone are unequipped to safely reason in the ever-changing adversarial landscape of continual strategic deception in human society. An overarching goal of research in this direction is to provide foundations for future artificial intelligence systems built on solution concepts that are

- robust with respect to different degrees of rationality and informativeness;
- capable of exploiting the information structure of the environment; and
- enable steering the interactions towards more socially desirable outcomes.

In scenarios with intendedly yet only boundedly rational agents and limited access to information, many properties of traditional solution concepts from game-theoretic literature are lost. Already in Chapter 3, we show that a (market) leader may no longer benefit from their commitment power[196] and solutions may have more intricate formulations than with full rationality[197]. The question of how to design efficient systems where imperfect agents interact then leads to the field of behavioral game theory and the integration of its postulates into information design theory. Indeed, accommodating the boundedly rational nature of human decision making proved crucial in applications akin to deployed patrol planning systems for the US Coast Guards or wildlife security. It was also shown to explain systematic deviations from predictions, e.g., in combinatorial auctions. As the adoption of complex artificial intelligence systems relying on idealized assumptions about strategic agents is still commonplace in the contemporary world, foundational studies of these problems and their solutions are to become yet more vital.

[196] When the agents act subrationally, there may not be an incentive to break ties in favor of the leader.

[197] Optimal strategies in boundedly rational interactions frequently have a form of solutions of non-convex mathematical programs, as many results in this dissertation demonstrated.

8.1 Thesis contributions

The main contribution of this dissertation is the study of several aforesaid foundational solution concepts modeling commitment and coordination in interactions among agents that act according to the quantal response model of behavior. In Chapter 4, we lay down the foundations of optimal commitment against a single cognitively restricted agent in one-shot games. We show that the problem is highly non-linear, and the number of local optima may grow linearly with the number of actions, even in the most uncomplicated games. The critical result of this chapter is a derivation of a so-called Dinkelbach-type formulation of the equilibrium that enables the identification of sufficient conditions for its concavity and formulation of a mixed integer linear program to find an approximate solution. The solutions also have a guaranteed quality that we further verify through numerical experiments.

Generalizing this approach to sequential games directly proves difficult because of the inherent exponential blowup in strategies associated with sequential decision making. This issue can not be sidestepped entirely, as we show in Chapter 5 by proving the concept's NP-hardness. Yet, we are able to ameliorate it on the side of the committing player by representing their strategy using a realization plan linear in the size of the game. We reformulate the equilibrium in a way to use Dinkelbach's approach again, characterizing polynomially solvable games as a by-product and designing a mixed integer linear program for approximating the optimal strategy. Again, we provide strict theoretical guarantees on the quality of the solution and carry out a successful empirical evaluation of the algorithm.

The second half of the dissertation is then dedicated to the coordination of quantal players. We aim to study strategic signaling in the context of bounded rationality by coupling correlated equilibrium with quantal response equilibrium. Chapter 6 offers two potential definitions of the said cross-breed in one-shot games and shows they naturally extend the precursor equilibria. Indeed, every quantal response equilibrium is quantal correlated, and the traditional correlated equilibrium is reached in the limit as quantal responses approach the best response. The space of all equilibria is compact, and in case the equilibrium is always unique, it is also connected. The principal contribution of this part of the dissertation is the introduction of an efficient homotopy method capable of simultaneously tracing the equilibrium and gradiently optimizing the signaling structure despite the concept being PPAD-hard. The algorithm's performance is then evaluated in the experiments.

The extension of the homotopic approach to sequential games is again not straightforward because of the many decision points to behave quantally at therein. This makes inverting the homotopic Jacobian matrix – which is the key component of the tracing procedure in one-shot games – computationally unfeasible. To solve this issue, we reformulate the homotopy by employing a specific class of hyperbolic quantal responses in Chapter 7. The hyperbolic formulation allows us to adopt a non-linear conjugate gradient method as an efficient homotopic corrector for tracing the path of the quantal correlated equilibrial correspondence.

8.2 Future work

My long-term hope is that the results presented in this dissertation may facilitate and provide guidance in designing real world artificial intelligence systems capable of accounting for the flawed reasoning of human actors and effectively leveraging complex information structures hidden in the environments. Previous chapters provided numerous examples of optimal signaling and coordination of boundedly rational agents, yet mostly in fundamental controlled models. To succeed in deploying well-performing systems, it is my conviction the results need to be extended in three major directions.

- *Investigating effects of bounded rationality in complex multi-agent models.*

  More specifically, I believe it to be necessary to study the robustness of signaling strategies with respect to miss-classified behavioral models and implications on the leader's outcomes to provide performance guarantees in dynamically changing environments.

- *Estimating behavioral models and information structures from interaction data.*

  Simultaneous acting and robust estimation of agents' motives and reasoning capabilities are particularly crucial as correctly calibrated models are rarely available in practice. Investigating safe model exploration across repeated interactions is hence to become essential.

- *Steering the interactions towards socially desirable outcomes.*

  Promoting social justice and the responsible use of artificial intelligence is one of the primary goals of my work. I frequently evaluate the performance of my algorithms with respect to social welfare, the Gini coefficient, or equity in resource availability. I believe a conscious focus on this aspect of research may have far-reaching impacts on human society.

I hope this thesis may serve as a stepping stone to more widespread acceptance of systematic behavioral models in modern artificial intelligence systems. Yet indeed, a tremendous amount of open questions and promising concepts still lie ahead.

Bibliography

Allgower, E. L. and K. Georg. *Numerical continuation methods: An introduction*. Vol. 13. Springer Science & Business Media, 2012.

Ashlagi, I., D. Monderer, and M. Tennenholtz. "On the value of correlation". *Journal of Artificial Intelligence Research* 33, 2008, pp. 575–613.

Aumann, R. J. "Correlated equilibrium as an expression of Bayesian rationality". *Econometrica: Journal of the Econometric Society*, 1987, pp. 1–18.

Avriel, M. *Nonlinear programming: Analysis and methods*. Courier Corporation, 2003.

Ben-Israel, A. and T. N. Greville. *Generalized inverses: Theory and applications*. Vol. 15. Springer Science & Business Media, 2003.

Bernheim, B. D. "Rationalizable strategic behavior". *Econometrica: Journal of the Econometric Society*, 1984, pp. 1007–1028.

Bošanský, B. and J. Čermák. "Sequence-form algorithm for computing Stackelberg equilibria in extensive-form games". In: *Proceedings of the 29th AAAI Conference on Artificial Intelligence*. 2015, pp. 805–811.

Boyd, S. and L. Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.

Camerer, C. F. *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press, 2011.

Camerer, C. F., T.-H. Ho, and J.-K. Chong. "A cognitive hierarchy model of games". *The Quarterly Journal of Economics* 119:3, 2004, pp. 861–898.

Candogan, O. and K. Drakopoulos. "Optimal signaling of content accuracy". *Operations Research* 68:2, 2020, pp. 497–515.

Čermák, J., B. Bošanský, K. Durkota, V. Lisý, and C. Kiekintveld. "Using correlated strategies for computing Stackelberg equilibria in extensive-form games". In: *Proceedings of the 30th AAAI Conference on Artificial Intelligence*. 2016, pp. 439–445.

Černý, J., B. An, and A. N. Zhang. *Quantal correlated equilibrium in extensive form games*. Technical report. 2023.

– "Quantal correlated equilibrium in normal form games". In: *Proceedings of the 23rd ACM Conference on Economics and Computation*. 2022, pp. 210–239.

Černý, J., V. Lisý, B. Bošanský, and B. An. "Computing Quantal Stackelberg equilibrium in extensive-form games". *Proceedings of the 35th AAAI Conference on Artificial Intelligence* 35:6, 2021, pp. 5260–5268.

– "Dinkelbach-type algorithm for computing Quantal Stackelberg equilibrium". In: *Proceedings of the 29th International Joint Conference on Artificial Intelli-*

gence. Ed. by C. Bessiere. International Joint Conferences on Artificial Intelligence Organization, 2020, pp. 246–253.

Cheeseman, P., B. Kanefsky, and W. M. Taylor. "Where the really hard problems are". In: *Proceedings of the 12th International Joint Conference on Artificial Intelligence*. Morgan Kaufmann Publishers Inc., Sydney, New South Wales, Australia, 1991, pp. 331–337.

Conitzer, V. and T. Sandholm. "Computing the optimal strategy to commit to". In: *Proceedings of the 7th ACM Conference on Electronic Commerce*. New York, NY, USA, 2006, pp. 82–90.

Crawford, V. P., M. A. Costa-Gomes, and N. Iriberri. "Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications". *Journal of Economic Literature* 51:1, 2013, pp. 5–62.

Daskalakis, C., P. W. Goldberg, and C. H. Papadimitriou. "The complexity of computing a Nash equilibrium". *SIAM Journal on Computing* 39:1, 2009, pp. 195–259.

Davis, T., N. Burch, and M. Bowling. "Using response functions to measure strategy strength." In: *Proceedings of the 28th AAAI Conference on Artificial Intelligence*. 2014, pp. 630–636.

Dhillon, A. and J. F. Mertens. "Perfect correlated equilibria". *Journal of Economic Theory* 68:2, 1996, pp. 279–302.

Dinkelbach, W. "On nonlinear fractional programming". *Management Science* 13:7, 1967, pp. 492–498.

Eysenck, M. W. and M. T. Keane. *Cognitive psychology: A student's handbook*. Taylor & Francis, 2000.

Farina, G., C. Kroer, and T. Sandholm. "Online convex optimization for sequential decision processes and extensive-form games". In: *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*. 2019, pp. 1917–1925.

Forges, F. "Correlated equilibria and communication in games". *Complex Social and Behavioral Systems: Game Theory and Agent-Based Models*, 2020, pp. 107–118.

Garcia, C. B. and W. I. Zangwill. *Pathways to solutions, fixed points, and equilibria*. English. Prentice-Hall Series in Computational Mathematics. Englewood Cliffs, New Jersey: Prentice-Hall, Inc. XV, 479 p. $ 43.95 (1981). 1981.

Georg, K. "A note on stepsize control for numerical curve following". In: *Homotopy Methods and Global Convergence*. Springer, 1983, pp. 145–154.

Gigerenzer, G. and D. G. Goldstein. "Reasoning the fast and frugal way: Models of bounded rationality." *Psychological Review* 103:4, 1996, p. 650.

Gigerenzer, G. and R. Selten. *Bounded rationality: The adaptive toolbox*. MIT Press, 2002.

Goeree, J. K., C. A. Holt, and T. R. Palfrey. "Quantal response equilibria". In: *Behavioural and Experimental Economics*. Springer, 2010, pp. 234–242.

– *Quantal Response Equilibrium*. Princeton University Press, 2016.

Gonzalez, C., J. F. Lerch, and C. Lebiere. "Instance-based learning in dynamic decision making". *Cognitive Science* 27:4, 2003, pp. 591–635.

Gray, J. R. "A bias toward short-term thinking in threat-related negative emotional states". *Personality and Social Psychology Bulletin* 25:1, 1999, pp. 65–75.

Hager, W. W. and H. Zhang. "A survey of nonlinear conjugate gradient methods". *Pacific journal of Optimization* 2:1, 2006, pp. 35–58.

Halpern, J. Y., R. Pass, and L. Seeman. "Computational extensive-form games". In: *Proceedings of the 17th ACM Conference on Economics and Computation*. 2016, pp. 681–698.

Harsanyi, J. C. "Games with incomplete information played by "Bayesian" players, I–III Part I. The basic model". *Management Science* 14:3, 1967, pp. 159–182.

Hart, S. and A. Mas-Colell. "A simple adaptive procedure leading to correlated equilibrium". *Econometrica* 68:5, 2000, pp. 1127–1150.

Hastie, T. J. "Generalized additive models". In: *Statistical Models in S*. Routledge, 2017, pp. 249–307.

Ibaraki, T. "Integer programming formulation of combinatorial optimization problems". *Discrete Mathematics* 16:1, 1976, pp. 39–52.

Jiang, A. X., K. Leyton-Brown, and N. A. Bhat. "Action-graph games". *Games and Economic Behavior* 71:1, 2011, pp. 141–173.

Jiang, A. X., T. H. Nguyen, M. Tambe, and A. D. Procaccia. "Monotonic maximin: A robust Stackelberg solution against boundedly rational followers". In: *International Conference on Decision and Game Theory for Security*. Springer. 2013, pp. 119–139.

Kahneman, D. and A. Tversky. "Prospect theory: An analysis of decision under risk". In: *Handbook of the Fundamentals of Financial Decision Making: Part I*. World Scientific, 2013, pp. 99–127.

Kolodziej, S., P. M. Castro, and I. E. Grossmann. "Global optimization of bilinear programs with a multiparametric disaggregation technique". *Journal of Global Optimization* 57:4, 2013, pp. 1039–1063.

Kroer, C., G. Farina, and T. Sandholm. "Robust Stackelberg equilibria in extensive-form games and extension to limited lookahead". In: *Proceedings of 32nd AAAI Conference on Artificial Intelligence*. 2018.

Kuhn, H. W. "Extensive games and the problem of information". *Annals of Mathematics Studies* 28, 1953.

Lee, C. "Bounded rationality and the emergence of simplicity amidst complexity". *Journal of Economic Surveys* 25:3, 2011, pp. 507–526.

Leonardos, S., G. Piliouras, and K. Spendlove. "Exploration-exploitation in multi-agent competition: Convergence with bounded rationality". *Advances in Neural Information Processing Systems* 34, 2021.

Letchford, J. and V. Conitzer. "Computing optimal strategies to commit to in extensive-form games". In: *Proceedings of the 11th ACM Conference on Electronic Commerce*. 2010, pp. 83–92.

Ling, C. K., F. Fang, and J. Z. Kolter. "What game are we playing? End-to-end learning in normal and extensive form games". In: *Proceedings of the 27th International Joint Conference on Artificial Intelligence*. 2018, pp. 396–402.

Malhotra, N. K. "Information load and consumer decision making". *Journal of Consumer Research* 8:4, 1982, pp. 419–430.

Marchesi, A. and N. Gatti. "Trembling-hand perfection and correlation in sequential games". In: *Proceedings of the 35th AAAI Conference on Artificial Intelligence*. 2021, pp. 5566–5574.

Markowitz, H. "The utility of wealth". *Journal of Political Economy* 60:2, 1952, pp. 151–158.

McCormick, G. P. "Computability of global solutions to factorable nonconvex programs: Part I - Convex underestimating problems". *Mathematical Programming* 10:1, 1976, pp. 147–175.

McFadden, D. L. "Quantal choice analaysis: A survey". In: *Annals of Economic and Social Measurement, Volume 5, number 4*. NBER, 1976, pp. 363–390.

McKelvey, R. D., A. M. McLennan, and T. L. Turocy. "Gambit: Software tools for game theory", 2006.

McKelvey, R. D. and T. R. Palfrey. "Quantal response equilibria for extensive form games". *Experimental Economics* 1:1, 1998, pp. 9–41.

– "Quantal response equilibria for normal form games". *Games and Economic Behavior* 10:1, 1995, pp. 6–38.

Megiddo, N. and A. Wigderson. "On play by means of computing machines: preliminary version". In: *Proceedings of the 1986 Conference on Theoretical Aspects of Reasoning About Knowledge*. Morgan Kaufmann Publishers Inc. 1986, pp. 259–274.

Mertikopoulos, P. and W. H. Sandholm. "Learning in games via reinforcement and regularization". *Mathematics of Operations Research* 41:4, 2016, pp. 1297–1324.

Milec, D., J. Černý, V. Lisý, and B. An. "Complexity and algorithms for exploiting quantal opponents in large two-player games". In: *Proceedings of the 35th AAAI Conference on Artificial Intelligence*. 2020.

Myerson, R. B. "Refinements of the Nash equilibrium concept". *International Journal of Game Theory* 7:2, 1978, pp. 73–80.

Nagel, R. "Unraveling in guessing games: An experimental study". *The American Economic Review* 85:5, 1995, pp. 1313–1326.

Nash, J. F. "Non-cooperative games". *Annals of Mathematics*. Second 54:2, 1951.

Nesterov, Y. *Introductory lectures on convex optimization: A basic course*. 1st ed. Applied Optimization 87. Springer US, 2004.

Neumann, J. von. "Zur theorie der gesellschaftsspiele". *Mathematische Annalen* 100:1, 1928, pp. 295–320.

Nguyen, T. H., A. Sinha, S. Gholami, A. Plumptre, L. Joppa, M. Tambe, M. Driciru, F. Wanyama, A. Rwetsiba, and R. Critchlow. "Capture: A new predictive anti-poaching tool for wildlife protection". In: *Proceedings of the 15th Interna-*

tional Conference on Autonomous Agents and Multiagent Systems. 2016, pp. 767–775.

Papadimitriou, C. H. "On the complexity of the parity argument and other inefficient proofs of existence". *Journal of Computer and System Sciences* 48:3, 1994, pp. 498–532.

Paulhus, D. L. and K. M. Williams. "The dark triad of personality: Narcissism, Machiavellianism, and psychopathy". *Journal of Research in Personality* 36:6, 2002, pp. 556–563.

Pita, J., M. Jain, M. Tambe, F. Ordónez, and S. Kraus. "Robust solutions to Stackelberg games: Addressing bounded rationality and limited observations in human cognition". *Artificial Intelligence* 174:15, 2010, pp. 1142–1171.

Potts, W. J. "Generalized additive neural networks". In: *Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1999, pp. 194–200.

Powell, M. J. "A view of algorithms for optimization without derivatives". *Mathematics Today-Bulletin of the Institute of Mathematics and its Applications* 43:5, 2007, pp. 170–174.

Pratt, J. W. "Risk aversion in the small and in the large". In: *Uncertainty in Economics*. Elsevier, 1978, pp. 59–79.

Rasmusen, E. *Games and information: An introduction to game theory*. Blackwell, 2001.

Rong, J., T. Qin, B. An, and T.-Y. Liu. "Modeling bounded rationality for sponsored search auctions". In: *Proceedings of the 22nd European Conference on Artificial Intelligence*. 2016, pp. 515–523.

Roughgarden, T. "Algorithmic game theory". *Communications of the ACM* 53:7, 2010, pp. 78–86.

Rubinstein, A. *Modeling bounded rationality*. MIT Press, 1998.

Selten, R. "Reexamination of the perfectness concept for equilibrium points in extensive games". *International Journal of Game Theory* 4, 1975.

Shoham, Y. and K. Leyton-Brown. *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press, 2009.

Simon, H. A. *Administrative behavior*. Simon and Schuster, 2013.

– "Bounded rationality in social science: Today and tomorrow". *Mind & Society* 1:1, 2000, pp. 25–39.

Stackelberg, H. *Market structure and equilibrium*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 1–10.

Stengel, B. von and F. Forges. "Extensive-form correlated equilibrium: Definition and computational complexity". *Mathematics of Operations Research* 33:4, 2008, pp. 1002–1022.

Stengel, B. von and S. Zamir. *Leadership with commitment to mixed strategies*. Technical report. 2004.

Tambe, M. *Security and game theory: Algorithms, deployed systems, lessons learned*. Cambridge University Press, New York, NY, USA, 2011.

Turocy, T. L. "A dynamic homotopy interpretation of the logistic quantal response equilibrium correspondence". *Games and Economic Behavior* 51:2, 2005, pp. 243–263.

– "Computing sequential equilibria using agent quantal response equilibria". *Economic Theory* 42:1, 2010, pp. 255–269.

Velupillai, K. V. "Foundations of boundedly rational choice and satisficing decisions." *Advances in Decision Sciences*, 2010.

Vielma, J. P., S. Ahmed, and G. Nemhauser. "Mixed-integer models for nonseparable piecewise-linear optimization: Unifying framework and extensions". *Operations Research* 58:2, 2010, pp. 303–315.

Wahba, G. *Spline models for observational data*. Vol. 59. SIAM, 1990.

Wang, K., L. Xu, A. Perrault, M. K. Reiter, and M. Tambe. "Coordinating followers to reach better equilibria: End-to-end gradient descent for Stackelberg games". In: *Proceedings of the 36th AAAI Conference on Artificial Intelligence*. 2022, pp. 5219–5227.

Wright, J. R. and K. Leyton-Brown. "Level-0 meta-models for predicting human behavior in games". In: *Proceedings of the 15th ACM Conference on Economics and Computation*. 2014, pp. 857–874.

Yang, R., C. Kiekintveld, F. Ordonez, M. Tambe, and R. John. "Improving resource allocation strategy against human adversaries in security games". In: *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*. Barcelona, Catalonia, Spain, 2011, pp. 458–464.

Yang, R., F. Ordonez, and M. Tambe. "Computing optimal strategy against quantal response in security games". In: *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems*. 2012, pp. 847–854.

Yano, M., J. D. Penn, G. Konidaris, and A. T. Patera. *Math, numerics & programming (for mechanical engineers)*. MIT Press, 2013.

Zoutendijk, G. "Nonlinear programming, computational methods". *Integer and Nonlinear Programming*, 1970, pp. 37–86.