Quantal Correlated Equilibrium in Normal Form Games

JAKUB ČERNÝ, Nanyang Technological University, Singapore and A*STAR, Singapore
BO AN, Nanyang Technological University, Singapore
ALLAN N. ZHANG, Singapore Institute of Manufacturing Technology, A*STAR, Singapore

Correlated equilibrium is an established solution concept in game theory describing a situation when players condition their strategies on external signals produced by a correlation device. In recent years, the concept has begun gaining traction also in general artificial intelligence because of its suitability for studying coordinated multi-agent systems. Yet the original formulation of correlated equilibrium assumes entirely rational players and hence fails to capture the subrational behavior of human decision-makers. We investigate the analogue of quantal response for correlated equilibrium, which is among the most commonly used models of bounded rationality. We coin the solution concept the *quantal correlated equilibrium* and study its relation to quantal response and correlated equilibria. The definition corroborates with prior conception as every quantal response equilibrium is a quantal correlated equilibrium, and correlated equilibrium is its limit as quantal responses approach the best response. We prove the concept remains PPAD-hard but searching for an optimal correlation device is beneficial for the signaler. To this end, we introduce a homotopic algorithm that simultaneously traces the equilibrium and optimizes the signaling distribution. Empirical results on one structured and one random domain show that our approach is sufficiently precise and several orders of magnitude faster than a state-of-the-art non-convex optimization solver.

CCS Concepts: • **Theory of computation** → **Algorithmic game theory**; **Exact and approximate computation of equilibria**.

Additional Key Words and Phrases: Normal-form games, bounded rationality, quantal response equilibrium, correlated equilibrium, quantal correlated equilibrium

ACM Reference Format:
Jakub Černý, Bo An, and Allan N. Zhang. 2022. Quantal Correlated Equilibrium in Normal Form Games. In *Proceedings of the 23rd ACM Conference on Economics and Computation (EC '22), July 11–15, 2022, Boulder, CO, USA.* ACM, New York, NY, USA, 30 pages. https://doi.org/10.1145/3490486.3538350

1 INTRODUCTION

Robert Aumann introduced the concept of correlated equilibrium [4] in 1974 as a generalization of Nash equilibrium [50]. While in Nash equilibrium the players act upon their best interests simply based on the reasoning about their opponents' strategies, in correlated equilibrium they may condition their behavior also on an external private signal. The process of selecting and revealing the signals is traditionally entrusted to a mediator mechanism, a so-called correlation device. Together with an underlying normal-form game, a correlation device constitute an *extended game* in which a profile of one signal per player is sampled first, the players consequently receive their private signals and choose their strategies accordingly. Correlated equilibrium consists of all pairs of a correlation device's distribution over signals and players' strategies, such that the strategies form a Nash equilibrium in the extended game [23].

EC '22, July 11–15, 2022, Boulder, CO, USA.

ACM ISBN 978-1-4503-9150-4/22/07…$15.00
https://doi.org/10.1145/3490486.3538350

Throughout the years, correlated equilibrium became one of the most prominent concepts in game theory, because of its suitability for studying coordinated multiagent systems [3, 33], as well as appealing computational complexity that is provably polynomial. Signaling has also applications in fighting misinformation [13], terrorism [37], or cybercrime [36]. Besides, many of the contemporary breakthroughs in Nash equilibrium computation in two-player zero-sum games (e.g., in [10, 48]) or even multi-player games (e.g., in [11]) 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 [14, 22] or team-maxmin equilibrium with coordination device [20, 60]. With the new approaches, even large scenarios may be 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 [2, 8, 34, 58]. Perhaps the first attempt to introduce "boundedly rational" behavior into correlated equilibrium was through trembling-hand perfection [53]. The idea of trembles is that players may choose suboptimal actions with the same non-zero, yet vanishing probability, and was studied in context of correlated strategies in both normal-form [18] and extensive-form [41] 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, as the literature shows that instead of making uniform deviations, human players tend to discriminate between different alternatives more systematically [12]. Among the most renowned models of bounded rationality that address this issue is the quantal response [43, 44]. Quantal response assumes that players act stochastically, choosing higher-utility actions with higher probability, and is consistently regarded as one of the best predictors of human behavior in games, verified by numerous experiments [31, 58]. The behavioral model of quantal response, in context of the leader-follower equilibria [15, 16], is also a fundamental component of several algorithms successfully deployed in the real world [19, 58]. Despite its importance, to the best of our knowledge, no work that studies correlated quantal response strategies has ever been published.

1.1 Contributions

In this work, we investigate the amalgamation of quantal response and correlated equilibrium. We consider generalized Luce models of quantal behavior that represent the response as a function increasing in utility normalized over the set of player's actions. Luce models are able to capture a wide range of behavior (including the most common logit model [43], but also attitudes towards risk and loss [35], or subjective utilities (e.g., in [51])), and despite being studied in theory before [29], to the best of our knowledge, our work is the *first to formulate an algorithm computing quantal equilibria with Luce models*. To this end, 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 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, in Section 3, 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.

In Section 4, 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 Section 5, 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 (e.g., a government) aims to coordinate the retailers to streamline the economy, but is only capable of sending signals (e.g., taxation policies) because of the retailers' autonomy. 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.

2 PROBLEM DEFINITION

In a normal-form game $G = (N, A, u)$, we denote the finite set of players N, $n = |N|$, and for each player $i \in N$, the finite set of possible actions for i is A_i. The set A is then a Cartesian product of A_i. For a tuple $a \in A$, we define a real-valued utility function for player i as $u_i : A \to \mathbb{R}$. Pure (i.e., deterministic) strategies are a Cartesian product Π of Π_i, where Π_i coincides with A_i and each $\pi_i \in \Pi_i$ represents an action the player i commits to playing. Mixed (i.e., stochastic) strategies Δ are a Cartesian product of Δ_i, and each Δ_i is a probability simplex over Π_i. For any tuple or product, we use $-i$ to denote the opponents of player i and their corresponding sub-tuple or sub-product, e.g., for $a = (a_1, \ldots, a_n) \in A$, $a_{-i} = (a_1, \ldots, a_n) \backslash \{a_i\}$, and $\delta_{-i}(a_{-i}) = \prod_{j \in N, i \neq j} \delta_j(a_j), \delta_i \in \Delta_i$.

2.1 Standard solution concepts in normal-form games

We provide formal definitions of two equilibria our work is built upon: correlated equilibrium and quantal response equilibrium. We begin with correlated equilibrium, which describes a situation when the players are able to coordinate their strategies based on external signals. Then we proceed to quantal response equilibrium, in which the players' decisions are affected by systematic biases.

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 $G = (N, A, S, u)$, where Λ is a set of distributions over a Cartesian product S of finite sets of private signals S_i for each player. The extended game is a sequential extension of the underlying game in which a correlation device first samples a signal profile according to the signaling scheme λ that is considered public knowledge. The players then learn about their respective signals and play strategies conditioned on the signal. Deterministic strategies in the extended game define action to play for each signal that could be received. In other words, each $\pi_i \in \Pi_i$ 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 strategies B describe stochastic strategies in the extended game, where 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 case of Nashian correlated equilibrium, the tuple $(\lambda, (\beta_i)_{i \in N}), \lambda \in \Lambda, \beta_i \in B_i$ is hence a *correlated equilibrium* if it satisfies the inequalities

$$\sum_{a \in A} \sum_{s \in S} \lambda(s) \beta(a|s) u_i(a) \geq \sum_{a_{-i} \in A_{-i}} \sum_{s \in S} \lambda(s) \beta_{-i}(a_{-i}|s) u_i(m(s_i), a_{-i}) \qquad \forall i \in N, \forall m : S_i \to A_i. \quad \text{(CE)}$$

Note that when all players are able to play any pure, mixed or behavioral strategy, according to Kuhn's equivalence theorem, there exists an equivalent formulation of (CE) also in mixed strategies. The probability distribution over pure strategy profiles induced by λ and β is called a *correlated equilibrium distribution* [5]. In case the players attain a Nash equilibrium in the extended game, according to the Revelation principle [24], 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 polynomial computability and equivalence to correlated equilibrium constructed the standard way, this form is most commonly studied in the literature. However, the underlying equilibrium plays a crucial role. For example, in perfect correlated equilibrium, the direct and standard formulations do not coincide [18].

Example 2.1. Consider a variant of the Battle of Sexes game shown in the center of Figure 1. The tetrahedron on the left is the simplex of probability distributions on pure profiles in the game, with the corners corresponding to the profiles themselves. The inner polytope with 4 vertices is the set of direct correlated equilibria. Each vertex is one Nash equilibrium in the game.

Both forms of correlated equilibrium assume perfectly rational strategizing of all players, i.e., each player is capable of selecting and following the utility-maximizing option. Relaxing this assumption leads to a "statistical version" of best response called *quantal response*, which takes into account the inevitable error-proneness of humans and allows the players to make systematic errors. Quantal-responding of player i is commonly modeled through function $QR_i : \Delta_{-i} \to \Delta_i$ that is monotonically increasing in expected utility:

$$u_i(\delta_{-i}, a_i^k) \leq u_i(\delta_{-i}, a_i^l) \Rightarrow QR_i^k(\delta_{-i}) \leq QR_i^l(\delta_{-i}) \qquad \forall \delta_{-i} \in \Delta_{-i}, a_i^k, a_i^l \in A_i,$$

where $QR_i^j(\delta_{-i})$ is the probability of playing action a^j. In this work, we focus on a wide class of quantal response functions called *generalized Luce models* [29].

Definition 2.2. 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

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

Because q_i is strictly positive and increasing function, the corresponding QR is a valid quantal response function. We call such functions q_i *generators* of generalized Luce models. When quantal response functions of all players are continuous, by a direct application of the Brouwer's theorem we know that the quantal response dynamic has a fixed point, called *quantal response equilibrium*.

Example 2.3. The graph on the right of Figure 1 depicts the quantal-response dynamic when both players act according to a generator $q(x) = (x + (119 + \sqrt{401937})/332)^3$. The x-axis varies the probability of the first player taking action o using a thick solid line. The second player quantal-responds, followed by the first player again. The equilibrium is a fixed point of this dynamic, i.e., when the line touches or crosses the quadrant's axis. There are two quantal response equilibria, $\delta_1(o) \in \{0.0834, 0.825\}$[1]. The induced probability distributions over profiles are depicted in the tetrahedron on the left in hollow circles.

[1]The quantal generator was numerically approximated such that for $\delta_1(o) = 0.0834$ the dynamic touches the quadrant's axis without crossing it, resulting in two equilibria instead of three. The generator is not entirely exact, but due to the concept's continuity the existence of a close-enough "touching" generator is guaranteed.

2.2 Quantal correlated equilibria

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 call this form of quantal correlated equilibrium the per-signal equilibrium. Inserting bounded rationality into the model is simple using the standard construction.

Definition 2.4. 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 (S-QCE) if

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

Example 2.5. Consider the game in Figure 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, similarly as with the quantal response equilibrium. There are three per-signal equilibria, for $\beta_1(o) \in \{0.0833, 0.233, 0.772\}$.

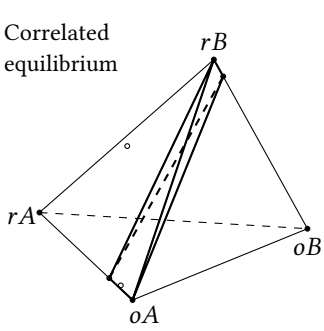
In our second formulation, instead of playing quantally for each signal independently, we assume the players act quantally over the whole set of pure strategies in the extended game. We refer to this formulation as the over-pure-strategies equilibrium.

Definition 2.6. 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 (Π-QCE) if

$$\begin{aligned} u_i(\pi_i) &= \sum_{(a_i,s_i)\in\pi_i} \sum_{\pi_{-i}\in\Pi_{-i}} \sum_{(a_{-i},s_{-i})\in\times\pi_{-i}} \lambda(s_i, s_{-i})\delta_{-i}(\pi_{-i})u_i(a_i, a_{-i}) \\ \delta_i(\pi_i) &= \frac{q_i(u_i(\pi_i))}{\sum_{\pi_i'\in\Pi_i} q_i(u_i(\pi_i'))}. \end{aligned} \tag{Π-QCE}$$

Example 2.7. Figure 1 shows an over-pure-strategies quantal correlated dynamic as well, using a thick dotted line. Similarly as in S-QCE, we consider a setting with one and two equally possible signals per player. The game contains a single over-pure-strategies equilibrium for $\delta_1(o) = 0.0833$.f

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), the formulation (S-QCE) 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 [27] over long-term, premeditated decisions. This behavior arises especially in conflicts [30] or when facing information overload caused by large decision space [40]. In the light of that, formulation (S-QCE) seems more natural then its counterpart (Π-QCE). For this reason, we will prefer the formulation (S-QCE) in the analysis and when we formulate an algorithm solving it, but the same approach could be applied also to formulation (Π-QCE).



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

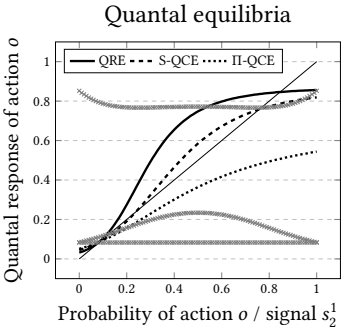


Fig. 1. Equilibria in a variant of the Battle of Sexes game. (Middle) Each row of the matrix is labeled by a strategy of the first player and every column is labeled by a strategy of the second player. The values denote the utilities of the first and the second player, respectively. (Left) The simplex of correlated equilibria. (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.

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 2.8. Let $G = (N, A, S, u)$ be a signaling game and $f : \Lambda \times \Sigma \to \mathbb{R}$ be a criterion function. The optimal QCE 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 B (for S-QCE) or Δ (for Π-QCE).

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.

3 PROPERTIES OF QUANTAL CORRELATED EQUILIBRIA

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

Proposition 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 λ.

Remark 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 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 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.*

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 3. *Let G be a normal-form game. Any quantal response equilibrium in G*

(1) may be extended into a per-signal quantal correlated equilibrium laying on a corner of the signaling simplex; and
(2) is a trivial over-pure-strategies quantal correlated equilibrium with a single signal per player.

Remark 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 (e.g., zero-sum games), 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 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 uncorrelted quantal equilibria.

This fact is also clear when looking at Figure 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 4. *Let $G = (N, A, S, u)$ be a signaling game and $C = \{(\lambda, QCE(\lambda)), \lambda \in \Lambda\}$, where QCE is either S-QCE or Π-QCE. Then 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 C is connected.*

Example 3.1. Figure 1 shows an example of the correspondence C depicted in gray x-lines. Here, C is projected on the coordinates corresponding to the probability of observing signal s_2^1 and probability of playing action o in $S - QCE(s_2^1)$. 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 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[2], the maximum is never reached in quantal response equilibrium.

[2]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 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.*

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. *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.*

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 results in this section we examine the concepts' computational complexity.

Proposition 7. *Let G be a signaling game of n players and $q_1, \dots, q_n$ be their respective generators. Let* QCE *be the problem of computing a quantal correlated equilibrium in G. Then* QCE *is PPAD-hard.*

Now we are ready to move to the introduction of a practical algorithm for computing a quantal correlated equilibrium.

4 HOMOTOPY METHOD FOR COMPUTING QUANTAL CORRELATED EQUILIBRIA

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: (i) they focus on the logit generator $q(x) = e^{\alpha x}$, and use its properties nontransferable to other generators in the Luce model; and (ii) they are able to compute only the normal-form representation of quantal response equilibrium which does not translate to the per-signal one as in S-QCE. We hence formulate a novel homotopy method optimizing quantal correlated equilibrium, making the following contributions: (i) we reformulate the generalized Luce model to improve robustness, (ii) we employ general product-separating functions to alleviate steepness of quantal generators; and (iii) we simultaneously trace the equilibrium and gradiently optimize 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 [45, 46]. Out of them, only the Karush-Kuhn-Tucker reformulation of per-player optimization is capable of converging in some general-sum games [57]. 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 [39], laminar regret minimization [21], or smooth Q-learning [38].

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 [25]. 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 [55]. Another homotopy for quantal response equilibrium was introduced in [52] in context of sponsored search auctions, using specific properties thereof. Neither could be used for quantal correlation because of their strict assumptions about game domains and quantal models.

4.1 Tracing the equilibrial correspondence path

We formulate a novel 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 KKT methods[3]. Moreover, each player may use a different generator, not necessarily a logit one[4], 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 4.1. 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

$$\hat{q}(x, t=0) = c, c \in \mathbb{R}$$
$$\hat{q}(x, t=1) = q(x).$$

Example 4.2. Perhaps the simplest way of creating parametric representations is in terms of *exponentiation*. Consider a logit generator $e^{\alpha x}$. One of is possible parametric representations is $e^{\alpha t x}$, 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 4.3. 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* [25]. 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 $e^{\alpha x}$ the newtonian representation looks as $te^{\alpha x} + (1-t)$.

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), we define a homotopy function $\overline{H}$ for QCE with

[3]For example, the experiments in [39] consider optimization over simple games (rock-paper-scissors, one-card poker, and a security game) with sequence-form payoff matrices of per-player size at most 16.

[4]Traversing the homotopic curve may even correspond to implementing different exploration-exploitation policies, e.g., the *Explore-Then-Exploit* [7], with an appropriate parametric representation.

strategy profile β and homotopic parameter t as

$$\overline{H}(\beta, t) = \left(H_i^{k,l}(\beta, t)\right)_{i \in N,\ a_i^k \in A_i,\ s_i^l \in S_i}$$
$$\overline{H}_i^{k,l}(\beta, t) = \hat{q}_i(u_i(a_i^k|s_i^l), t) - \beta_i(a_i^k|s_i^l) \sum_{a_i \in A_i} \hat{q}_i(u_i(a_i|s_i^l), t),$$

i.e., a reformulation of the second equation in (Π-QCE) with the generators substituted by their parametric representations. Note that 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) for $t = 1$ by the definition of parametric representations. Moreover,

Proposition 8. *For arbitrary λ, the solution for $t = 0$ is a uniform strategy for each signal.*

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. 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

$$\beta_i(a_i^0|s_i) = \frac{q_i(u_i(a_i^0|s_i))}{\sum_{a_i \in A_i} q_i(u_i(a_i|s_i))}, \qquad \beta_i(a_i^j|s_i) = \frac{q_i(u_i(a_i^j|s_i))}{\sum_{a_i \in A_i} q_i(u_i(a_i|s_i))}.$$

Because both equalities share the same sum $\sum_{a_i \in A_i} q_i(u_i(a_i|s_i))$, we may write

$$\beta_i(a_i^0|s_i) q_i(u_i(a_i^j|s_i)) = \beta_i(a_i^j|s_i) q_i(u_i(a_i^0|s_i)). \qquad (\star)$$

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 ($\star$). 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 ($\star$) then motivates a substitution of variables $\gamma = f_1(\beta)$ and the resulting homotopy is formulated as

$$H(\gamma, t) = \left(H_i^{k,l}(\gamma, t)\right)_{i \in N,\ a_i^k \in A_i,\ s_i^l \in S_i}$$
$$H_i^{k,l}(\gamma, t) = f_2(f_1(\hat{q}_i(u_i(a_i^0|s_i^l), t)), \gamma_i(a_i^k|s_i^l)) - f_2(f_1(\hat{q}_i(u_i(a_i^k|s_i^l), t)), \gamma_i(a_i^0|s_i^l))$$
$$H_i^{0,l}(\gamma, t) = \sum_{a_i^k \in A_i} f_1^{-1}(\gamma_i(a_i^k|s_i^l)) - 1.$$

Because H is a reformulation of system $\overline{H}$, a simple modification of Proposition 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.

THEOREM 4.4 ([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:*

$$(1)\ H'(c(s))c'(s) = 0, \qquad (2)\ \|c(s)\| = 1, \qquad (3)\ det\begin{pmatrix} H'(c(p)) \\ c'(p) \end{pmatrix} > 0.$$

As a consequence of Theorem 4.4, the curve c associated with the quantal correlated equilibrium homotopy may be regarded as a local solution of an initial value problem:

$$(i)\ (\gamma(p), t(p))' = c'(H'(\gamma(p), t(p))), \qquad (ii)\ (\gamma(0), t(0)) = (\gamma^{t=0}, 0),$$

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 literature 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 [1], 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[5], 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 $(\gamma, t)^{\iota+1}$ laying on the curve [9]. 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 steplenght h accordingly during iterations when c becomes more linear (or conversely, more curvy) 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 $t = 1$ [26]. 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 $H'^\top$ [1, 9]. 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

[5]Our experiments with lower-degree Runge-Kutta methods yielded similar results.

ALGORITHM 1: Predictor-corrector method for quantal correlated equilibrium

Input: $H, (\gamma, t)$ such that $H((\gamma, t)) = 0, \bar{\iota}$
Parameters: $h, \underline{h}, \bar{\iota}_{gn}, \underline{\epsilon}, \overline{\epsilon}, \epsilon_t, a_\kappa, a_f, a_\eta$
$\iota \leftarrow 0$
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`
    $\overline{(\gamma, t)} \leftarrow GaussNewton((\overline{\gamma, t}), Q, R), \quad \iota_{gn} \leftarrow \iota_{gn} + 1$
    $h,\ \|c|(\gamma, t)\|,\ f,\ accept,\ newton \leftarrow$
     $UpdateStep(h, \iota_{gn}, f, (\gamma, t), \overline{(\gamma, t)}, \|c|(\gamma, t)\|, \underline{\epsilon}, \overline{\epsilon}, a_\kappa, a_f, a_\eta, newton)$
    if $\|c|(\gamma, t)\| > \overline{\epsilon}_c$ ***or*** $\iota_{gn} > \overline{\iota_{gn}}$ **then** $accept \leftarrow False$
    if ***not*** *accept* **then** break
  end
  if *accept* **then** $(\gamma, t) \leftarrow \overline{(\gamma, t)},\ \iota \leftarrow \iota + 1$ **else** $h \leftarrow h/a_f$
end
return (γ, t)

numerical stability. Let z denote the last column of Q, then the tangent and the Moore-Penrose inverse may be obtained as

$$c' = sgn(det(Q)det(R))z, \qquad H'^{+} = Q \begin{pmatrix} R^{\top -1} \\ 0 \end{pmatrix}.$$

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.

$$\frac{\partial H_i^{k,l}}{\partial \gamma_i(a_i^k|s_i^l)} = \frac{\partial f_2(f_1(\hat{q}_i(u_i(a_i^0|s_i^l), t)), \gamma_i(a_i^k|s_i^l))}{\partial \gamma_i(a_i^k|s_i^l)}$$

$$\frac{\partial H_i^{k,l}}{\partial \gamma_i(a_i^0|s_i^l)} = -\frac{\partial f_2(f_1(\hat{q}_i(u_i(a_i^k|s_i^l), t)), \gamma_i(a_i^0|s_i^l))}{\partial \gamma_i(a_i^0|s_i^l)}$$

$$\begin{aligned}\frac{\partial H_i^{k,l}}{\partial \gamma_{i'}(a_{i'}^{j'}|s_{i'}^{l'})} = &\frac{\partial f_2(f_1(\hat{q}_i(u_i(a_i^0|s_i^l), t)), \gamma_i(a_i^k|s_i^l))}{\partial f_1(\hat{q}_i(u_i(a_i^0|s_i^l), t))} \frac{\partial f_1(\hat{q}_i(u_i(a_i^0|s_i^l), t))}{\partial \hat{q}_i(u_i(a_i^0|s_i^l), t))} \frac{\partial \hat{q}_i(u_i(a_i^0|s_i^l), t)}{\partial u_i(a_i^0|s_i^l)} \frac{\partial u_i(a_i^0|s_i^l)}{\partial \gamma_{i'}(a_{i'}^{j'}|s_{i'}^{l'})} \\ &- \frac{\partial f_2(f_1(\hat{q}_i(u_i(a_i^k|s_i^l), t)), \gamma_i(a_i^0|s_i^l))}{\partial f_1(\hat{q}_i(u_i(a_i^k|s_i^l), t))} \frac{\partial f_1(\hat{q}_i(u_i(a_i^k|s_i^l), t))}{\partial \hat{q}_i(u_i(a_i^k|s_i^l), t))} \frac{\partial \hat{q}_i(u_i(a_i^k|s_i^l), t)}{\partial u_i(a_i^k|s_i^l)} \frac{\partial u_i(a_i^k|s_i^l)}{\partial \gamma_{i'}(a_{i'}^{j'}|s_{i'}^{l'})}\end{aligned}$$

$$\begin{aligned}\frac{\partial H_i^{k,l}}{\partial t} = &\frac{\partial f_2(f_1(\hat{q}_i(u_i(a_i^0|s_i^l), t)), \gamma_i(a_i^k|s_i^l))}{\partial f_1(\hat{q}_i(u_i(a_i^0|s_i^l), t))} \frac{\partial f_1(\hat{q}_i(u_i(a_i^0|s_i^l), t))}{\partial \hat{q}_i(u_i(a_i^0|s_i^l), t))} \frac{\partial \hat{q}_i(u_i(a_i^0|s_i^l), t)}{\partial t} \\ &- \frac{\partial f_2(f_1(\hat{q}_i(u_i(a_i^k|s_i^l), t)), \gamma_i(a_i^0|s_i^l))}{\partial f_1(\hat{q}_i(u_i(a_i^k|s_i^l), t))} \frac{\partial f_1(\hat{q}_i(u_i(a_i^k|s_i^l), t))}{\partial \hat{q}_i(u_i(a_i^k|s_i^l), t))} \frac{\partial \hat{q}_i(u_i(a_i^k|s_i^l), t)}{\partial t},\end{aligned}$$

where $i' \neq i$ and for any k, including 0,

$$\frac{\partial u_i(a_i^k|s_i^l)}{\partial \gamma_{i'}(a_{i'}^{k'}|s_{i'}^{l'})} = \sum_{\substack{a_{-i}\in A_{-i} \\ a_{i'}^{k'}\in a_{-i}}} \sum_{\substack{s_{-i}\in S_{-i} \\ s_{i'}^{l'}\in s_{-i}}} \lambda(s_i, s_{-i}) u_i(a_i^k, a_{-i}) \frac{\partial f_1^{-1}(\gamma_{i'}(a_{i'}^{k'}|s_{i'}^{l'}))}{\partial \gamma_{i'}(a_{i'}^{k'}|s_{i'}^{l'})} \prod_{\substack{j\in -i\setminus i' \\ a_j\in a_{-i}, s_j\in s_{-i}}} f_1^{-1}(\gamma_j(a_j|s_j)).$$

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 1. Here, h is the initial stepleght with $\underline{h}$ being its minimum value, and $\bar{\iota}_{gn}$ is a maximum number of iterations of the Gauss-Newton method. $\underline{\epsilon}$ and $\overline{\epsilon}$ 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. The full description and pseudocode of the steplength update algorithm can be found in the appendix.

Remark 3. The predictor-corrector method may potentially diverge because of the Jacobian unboundedness. Contrary to general sequential games [56], 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, then the Jacobian is bounded. All f_1, f_2 and q_i's considered in this work satisfy this condition.

4.2 Finding locally optimal signaling scheme

There may be multiple curves spanning across the solution space of the 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 the previous section, we described how to traverse a specific, unique branch that starts with uniform behavioral strategies (i.e., a centroid of the strategy simplex) and gradually approaches the quantal correlated equilibrium for a given signaling scheme and quantal generators, moving across the whole domain of t. In the literature, this branch is commonly referred to as the *principal branch* [28, 52, 55][6]. 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 [38]. The exact same path is taken also by replicator dynamic, a standard algorithm of evolutionary game theory [55].

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 [52] or subrationality estimation for general normal-form games [42]. 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

[6]More specifically, for quantal functions that approach best response, the corresponding quantal response equilibrium approximates a unique *limiting Nash equilibrium* on the principal branch.

ALGORITHM 2: Gradient optimization of signaling scheme and equilibrium tracing

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** $accept \leftarrow False, \quad$ break
  end
  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** $\eta \leftarrow \Delta_r \cdot \eta$
end

suited optimization is the earlier mentioned formulation OPT:

$$\max_{\lambda \in \Lambda, \beta \in B} f(\lambda, \beta) \qquad s.t.\ \beta \in S - QCE(\lambda).$$

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 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, i.e., changing the scheme is a part of the optimization process, not the homotopical traversal. 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.

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, i.e., to compute $\beta_p'(\lambda)$. To this end, we may use the homotopy method, because in case we remember the internediate 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

$$\lambda^{\iota+1} \leftarrow P_{\Lambda}\left(\lambda^{\iota} + \eta f'(\lambda, \beta_p(\lambda))\right),$$

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 perform 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 2. 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 $\bar{\imath}$ 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 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 4.4, $(\beta(p), t(p)) : H_\lambda((\beta(p), t(p))) = 0$, there exists sufficiently small η such that a Gauss-Newton sequence $\{\mathcal{N}^i((\beta, t))\}_{i=0}^{\infty}$ converges to a point $(\beta', t') : H_{\lambda'}((\beta', t')) = 0$, $\lambda' = P_\Lambda(\lambda + \eta f'(\lambda, \beta(p)))$.*

Now we turn to the question of optimality of the found solution. Proposition 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. Moreover, note that because of compactness, convergence to local optima is still guaranteed.

5 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[7]. 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.

The implementation of Algoritm 1 is based on a general homotopy scheme described in [1] in Appendix P3 which is further used also in Gambit Library's quantal response equilibrium solver [42]. We set the parameters as $h = 0.35, \underline{h} = 10^{-8}, \bar{l}_{gn} = 100, \underline{\epsilon} = 10^{-4}, \overline{\epsilon} = 0.8, \epsilon_t = 10^{-4}, a_\kappa = 0.8, a_f = 0.8, a_\eta = 0.1, \eta = 0.8, \underline{\eta} = 10^{-5}, \Delta_a = 0.99, \Delta_r = 0.9$, and $\bar{\imath} = 10$. 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.

5.1 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 (e.g., in Stackelberg games it may be easier to solve a random game than a game with specific structure), 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. Full formal description of both games is in the appendix.

[7] According to results published at http://plato.asu.edu/ftp/minlp.html.

5.1.1 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*[8] – a maximization of a sum of players' utilities – formally defined as

$$welfare(\beta) = \sum_{i \in N} \sum_{a_i \in A_i} \sum_{s_i \in S_i} \beta_i(a_i|s_i) u_i(a_i|s_i).$$

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 $q(x) = e^x$. 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 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.

5.1.2 Supply chain games. In this game, the suppliers P choose a warehouse $h \in H$ to store a raw material, while the retailers R choose a place $f \in F$ 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. 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

$$\sum_{f \in F} \left(t(f) - \sum_{\substack{r \in R \\ \tau_{hf}(f) \in \tau_{sr}(r)}} \sum_{s_r \in S_r} \beta_r(f|s_r) \sum_{s_{-r} \in S_{-r}} \lambda(s_r, s_{-r}) \right)^2,$$

where functions τ identify a territory where a manufacture is located or the retailer is licensed to operate, respectively, described in detail in the appendix. Similarly as in randomly generated

[8]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.

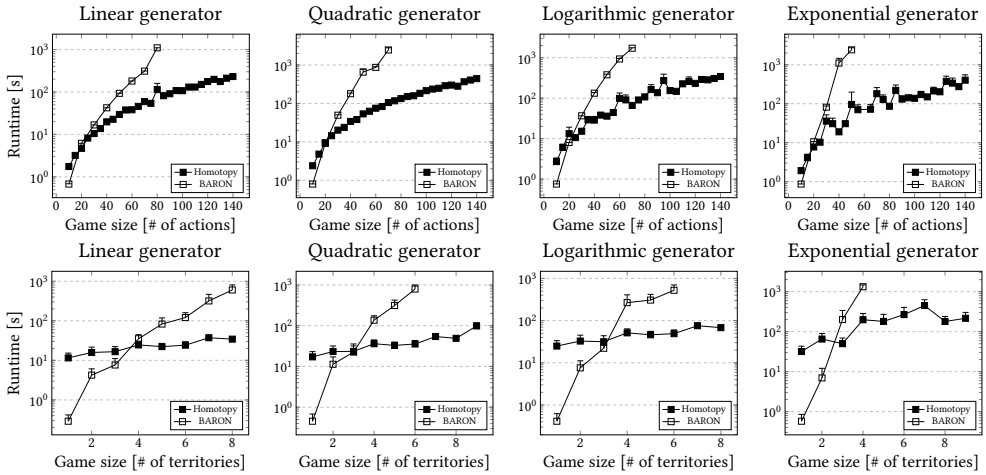


Fig. 2. Mean runtimes of computing (S-QCE) with fixed λ using BARON and the homotopy algorithm in (Top) randomly generated games and (Bottom) supply chain games. Every point shows also a standard error.

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 $f_1(x) = x$, $f_2(x, y) = xy$. 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 [32].

5.2 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 size used to illustrate scalability. For randomly generated games it is a number of actions per player, while in supply chain games it is a number of territories in the game. For each combination of setting × game size × 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.

5.2.1 Computing quantal correlated equilibrium. The upper line of Figure 2 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. 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 bottom line of Figure 2. 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 [6]. The same convergence holds also for other special points of interest on the branch, i.e., zero or extremal points of a smooth functional $c(p) \to \mathbb{R}$. We may hence expect similar scalability as presented here also for other criteria, e.g., maximum likelihood estimation along the path.

5.2.2 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 on the left in Figures 3 (maximizing social welfare) and 4 (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 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. On the right of both figures we present scalability comparison of different generators in two-player square games with 2 signals per player. The results indicate that quantal correlated equalibria with logarithmic generators consistently take the longest to compute, while the the most common logistic exponential generator is among the fastest.

	Quadratic generator			Logarithmic generator			Exponential generator		
#a	H [s]	B [s]	Δ	H [s]	B [s]	Δ	H [s]	B [s]	Δ
2	2±0	0±0	$-6 \cdot 10^6$	4±0	0±0	$-7 \cdot 10^6$	1±0	1±0	$1 \cdot 10^1$
3	3±0	5±1	$-8 \cdot 10^6$	12±3	1±0	$-2 \cdot 10^5$	2±0	15±3	$-6 \cdot 10^2$
4	6±1	23±3	$-6 \cdot 10^4$	12±2	1±0	$-5 \cdot 10^6$	2±0	208±62	$-7 \cdot 10^3$
5	8±1	345±240	$-1 \cdot 10^5$	19±2	3±0	$-6 \cdot 10^6$	3±0	898±191	$-8 \cdot 10^3$
6	11±0	481±191	$-6 \cdot 10^6$	25±0	10±3	$2 \cdot 10^2$	-	-	-
7	18±1	1052±212	$-8 \cdot 10^6$	35±1	11±2	$-3 \cdot 10^6$	-	-	-
8	23±3	1619±371	$-8 \cdot 10^6$	45±1	61±43	$-3 \cdot 10^6$	-	-	-

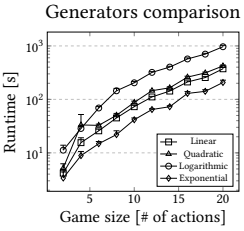


Fig. 3. The results of searching for a signaling scheme optimizing *social welfare* in **randomly generated games**. (Left) Comparison of the runtimes of the homotopy algorithm (H) and BARON (B), and deviations (Δ) from global solutions in smaller games. (Right) The runtimes of different generators in larger games. With the exception of the deviations, each value shows also a standard error.

The results in supply chain games can be found in Figure 5. 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. On the right side of the figure we compare scalability of individual generators in games with 3 suppliers, 2 retailers and 2 materials. As expected, the logarithmic one performs 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.

	Linear generator			Quadratic generator			Logarithmic generator			Exponential generator		
#a	H [s]	B [s]	Δ	H [s]	B [s]	Δ	H [s]	B [s]	Δ	H [s]	B [s]	Δ
2	2±0	0±0	$2 \cdot 10^2$	3±0	0±0	$-3 \cdot 10^7$	6±1	0±0	$-2 \cdot 10^6$	1±0	1±0	$-3 \cdot 10^3$
3	7±1	0±0	$1 \cdot 10^2$	7±0	9±1	$-2 \cdot 10^7$	14±4	2±0	$-1 \cdot 10^6$	3±0	51±13	$-1 \cdot 10^3$
4	15±4	0±0	$6 \cdot 10^4$	12±2	1266±472	$4 \cdot 10^7$	27±6	13±6	$-9 \cdot 10^7$	17±13	1191±466	$-7 \cdot 10^5$
5	16±3	7±2	$2 \cdot 10^3$	-	-	-	44±11	45±15	$-2 \cdot 10^3$	-	-	-
6	35±9	169±139	$2 \cdot 10^3$	-	-	-	61±12	159±50	$-8 \cdot 10^7$	-	-	-
7	126±85	267±204	$2 \cdot 10^3$	-	-	-	75±15	1156±422	$-4 \cdot 10^6$	-	-	-

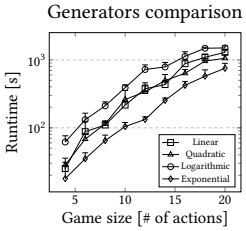


Fig. 4. The results of searching for a signaling scheme optimizing *Gini index* in **randomly generated games**: comparison of runtimes, deviations, and generators' scalability. The figure follows the same format as Figure 3.

	Linear generator			Quadratic generator			Logarithmic generator			Exponential generator		
#t	H [s]	B [s]	Δ	H [s]	B [s]	Δ	H [s]	B [s]	Δ	H [s]	B [s]	Δ
1	15±4	5±3	$-2 \cdot 10^7$	20±5	799±745	$-6 \cdot 10^7$	58±23	2±1	$-6 \cdot 10^7$	5±1	1167±730	$-1 \cdot 10^2$
2	23±4	525±411	$-3 \cdot 10^7$	32±8	3734±1193	$7 \cdot 10^7$	67±14	72±52	$-5 \cdot 10^7$	21±8	2990±1097	$-5 \cdot 10^3$
3	74±33	2746±995	$-8 \cdot 10^8$	62±14	4866±938	$-4 \cdot 10^5$	179±47	756±694	$-1 \cdot 10^6$	60±17	5665±890	$-1 \cdot 10^4$
4	209±154	2726±1084	$-4 \cdot 10^8$	-	-	-	261±74	1084±758	$-1 \cdot 10^4$	-	-	-

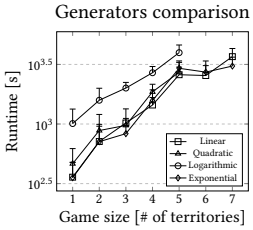


Fig. 5. The results of searching for a signaling scheme optimizing *manufacture allocation* in **supply chain games**: comparison of runtimes, deviations, and scalability. The figure follows the same format as Figure 3.

6 CONCLUSION

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: (i) we eliminated the normalization sum in Luce models, (ii) we reformulated the product of strategies using product-separating functions, and (iii) 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 (up to 300-times) than the state-of-the-art solver BARON and provides competitive solutions.

ACKNOWLEDGMENTS

Jakub would like to express his gratitude towards Géraldine Bouveret for discussing the theoretical foundations of quantal correlated equilibrium and Joanna Low for elevating the manuscript's prose. The authors also thank the anonymous reviewers for helping improving the concept's motivation and suggesting examining scalability with respect to signal sets cardinalities. This research is partially supported by the A*STAR CPPS – Contextual & Intelligent Programme, under the RIE2020 IAF-PP Grant No. A19C1a0018 and MOE Tier-1 project RG13/19 (S).

REFERENCES

[1] Eugene L Allgower and Kurt Georg. 2012. *Numerical continuation methods: an introduction*. Vol. 13. Springer Science & Business Media.

[2] Bo An, Fernando Ordóñez, Milind Tambe, Eric Shieh, Rong Yang, Craig Baldwin, Joseph DiRenzo III, Kathryn Moretti, Ben Maule, and Garrett Meyer. 2013. A deployed quantal response-based patrol planning system for the US Coast Guard. *Interfaces* 43, 5 (2013), 400–420.

[3] Itai Ashlagi, Dov Monderer, and Moshe Tennenholtz. 2008. On the value of correlation. *Journal of Artificial Intelligence Research* 33 (2008), 575–613.

[4] Robert Aumann. 1974. Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics* 1, 1 (1974), 67–96.

[5] Robert J Aumann. 1987. Correlated equilibrium as an expression of Bayesian rationality. *Econometrica: Journal of the Econometric Society* (1987), 1–18.

[6] Mordecai Avriel. 2003. *Nonlinear programming: analysis and methods*. Courier Corporation.

[7] Yu Bai and Chi Jin. 2020. Provable self-play algorithms for competitive reinforcement learning. In *International Conference on Machine Learning*. PMLR, 551–560.

[8] Anjon Basak, Jakub Černý, Marcus Gutierrez, Shelby Curtis, Charles Kamhoua, Daniel Jones, Branislav Bošanský, and Christopher Kiekintveld. 2018. An initial study of targeted personality models in the FlipIt game. In *International Conference on Decision and Game Theory for Security*. Springer, 623–636.

[9] Adi Ben-Israel and Thomas NE Greville. 2003. *Generalized inverses: theory and applications*. Vol. 15. Springer Science & Business Media.

[10] Noam Brown and Tuomas Sandholm. 2018. Superhuman AI for heads-up no-limit poker: Libratus beats top professionals. *Science* 359, 6374 (2018), 418–424.

[11] Noam Brown and Tuomas Sandholm. 2019. Superhuman AI for multiplayer poker. *Science* 365, 6456 (2019), 885–890.

[12] Colin F Camerer. 2011. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press.

[13] Ozan Candogan and Kimon Drakopoulos. 2020. Optimal signaling of content accuracy: Engagement vs. misinformation. *Operations Research* 68, 2 (2020), 497–515.

[14] Andrea Celli, Alberto Marchesi, Gabriele Farina, and Nicola Gatti. 2020. No-regret learning dynamics for extensive-form correlated equilibrium. In *Conference on Neural Information Processing Systems (NeurIPS)*.

[15] Jakub Černý, Viliam Lisý, Branislav Bošanský, and Bo An. 2020. Dinkelbach-type algorithm for computing Quantal Stackelberg equilibrium. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, Christian Bessiere (Ed.). International Joint Conferences on Artificial Intelligence Organization, 246–253.

[16] Jakub Černý, Viliam Lisý, Branislav Bošanský, and Bo An. 2021. Computing Quantal Stackelberg equilibrium in extensive-form games. *Proceedings of the AAAI Conference on Artificial Intelligence* 35, 6 (May 2021), 5260–5268.

[17] Constantinos Daskalakis, Paul W Goldberg, and Christos H Papadimitriou. 2009. The complexity of computing a Nash equilibrium. *SIAM J. Comput.* 39, 1 (2009), 195–259.

[18] Amrita Dhillon and Jean François Mertens. 1996. Perfect correlated equilibria. *Journal of Economic Theory* 68, 2 (1996), 279–302.

[19] Fei Fang, Thanh H Nguyen, Rob Pickles, Wai Y Lam, Gopalasamy R Clements, Bo An, Amandeep Singh, Brian C Schwedock, Milind Tambe, and Andrew Lemieux. 2017. PAWS – A deployed game-theoretic application to combat poaching. *AI Magazine* (2017).

[20] Gabriele Farina, Andrea Celli, Nicola Gatti, and Tuomas Sandholm. 2021. Connecting optimal ex-ante collusion in teams to extensive-form correlation: Faster algorithms and positive complexity results. In *International Conference on Machine Learning*.

[21] Gabriele Farina, Christian Kroer, and Tuomas Sandholm. 2019. Online convex optimization for sequential decision processes and extensive-form games. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 1917–1925.

[22] Gabriele Farina, Christian Kroer, and Tuomas Sandholm. 2021. Better regularization for sequential decision spaces: Fast convergence rates for Nash, correlated, and team equilibria. In *ACM Conference on Economics and Computation*.

[23] Françoise Forges. 1986. An approach to communication equilibria. *Econometrica* 54, 6 (1986), 1375–85.

[24] Françoise Forges. 2020. Correlated equilibria and communication in games. *Complex Social and Behavioral Systems: Game Theory and Agent-Based Models* (2020), 107–118.

[25] CB Garcia and WI Zangwill. 1981. Pathways to Solutions, Fixed Points, and Equilibria.

[26] K Georg. 1983. A note on stepsize control for numerical curve following. In *Homotopy methods and global convergence*. Springer, 145–154.

[27] Gerd Gigerenzer and Daniel G Goldstein. 1996. Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review* 103, 4 (1996), 650.

[28] Jacob K Goeree, Charles A Holt, and Thomas R Palfrey. 2010. Quantal response equilibria. In *Behavioural and Experimental Economics*. Springer, 234–242.

[29] Jacob K Goeree, Charles A Holt, and Thomas R Palfrey. 2016. *Quantal Response Equilibrium*. Princeton University Press.

[30] Jeremy R Gray. 1999. A bias toward short-term thinking in threat-related negative emotional states. *Personality and Social Psychology Bulletin* 25, 1 (1999), 65–75.

[31] Philip A Haile, Ali Hortaçsu, and Grigory Kosenok. 2008. On the empirical content of quantal response equilibrium. *American Economic Review* 98, 1 (2008), 180–200.

[32] Albert Xin Jiang, Kevin Leyton-Brown, and Navin AR Bhat. 2011. Action-graph games. *Games and Economic Behavior* 71, 1 (2011), 141–173.

[33] Albert Xin Jiang, Ariel D Procaccia, Yundi Qian, Nisarg Shah, and Milind Tambe. 2013. Defender (mis) coordination in security games. In *Twenty-Third International Joint Conference on Artificial Intelligence*.

[34] Michael Johanson and Michael Bowling. 2009. Data biased robust counter strategies. In *Artificial Intelligence and Statistics*. 264–271.

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

[36] Madasamy Kaliappan and Balasubramanian Paramasivan. 2015. Enhancing secure routing in mobile ad hoc networks using a dynamic bayesian signalling game model. *Computers & Electrical Engineering* 41 (2015), 301–313.

[37] Harvey E Lapan and Todd Sandler. 1993. Terrorism and signalling. *European Journal of Political Economy* 9, 3 (1993), 383–397.

[38] Stefanos Leonardos, Georgios Piliouras, and Kelly Spendlove. 2021. Exploration-exploitation in multi-agent competition: Convergence with bounded rationality. *Advances in Neural Information Processing Systems* 34 (2021).

[39] Chun Kai Ling, Fei Fang, and J Zico Kolter. 2018. 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*. 396–402.

[40] Naresh K Malhotra. 1982. Information load and consumer decision making. *Journal of Consumer Research* 8, 4 (1982), 419–430.

[41] Alberto Marchesi and Nicola Gatti. 2021. Trembling-hand perfection and correlation in sequential games. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 5566–5574.

[42] Richard D McKelvey, Andrew M McLennan, and Theodore L Turocy. 2006. Gambit: Software tools for game theory. (2006).

[43] Richard D. McKelvey and Thomas R. Palfrey. 1995. Quantal response equilibria for normal form games. *Games and Economic Behavior* 10, 1 (1995), 6–38.

[44] Richard D McKelvey and Thomas R Palfrey. 1998. Quantal response equilibria for extensive form games. *Experimental Economics* 1, 1 (1998), 9–41.

[45] Panayotis Mertikopoulos and William H Sandholm. 2016. Learning in games via reinforcement and regularization. *Mathematics of Operations Research* 41, 4 (2016), 1297–1324.

[46] Panayotis Mertikopoulos and Zhengyuan Zhou. 2019. Learning in games with continuous action sets and unknown payoff functions. *Mathematical Programming* 173, 1 (2019), 465–507.

[47] David Milec, Jakub Černý, Viliam Lisý, and Bo An. 2021. Complexity and algorithms for exploiting quantal opponents in large two-player games. *Proceedings of the AAAI Conference on Artificial Intelligence* 35, 6 (May 2021), 5575–5583.

[48] Matěj Moravčík, Martin Schmid, Neil Burch, Viliam Lisý, Dustin Morrill, Nolan Bard, Trevor Davis, Kevin Waugh, Michael Johanson, and Michael Bowling. 2017. DeepStack: Expert-level artificial intelligence in no-limit poker. *Science* (2017).

[49] Anna Nagurney. 2021. Supply chain game theory network modeling under labor constraints: Applications to the Covid-19 pandemic. *European Journal of Operational Research* 293, 3 (2021), 880–891.

[50] John F. Nash. 1951. Non-Cooperative Games. *Annals of Mathematics* 54, 2 (1951).

[51] Thanh Nguyen, Rong Yang, Amos Azaria, Sarit Kraus, and Milind Tambe. 2013. Analyzing the effectiveness of adversary modeling in security games. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 27. 718–724.

[52] Jiang Rong, Tao Qin, Bo An, and Tie-Yan Liu. 2016. Modeling bounded rationality for sponsored search auctions. In *Proceedings of the Twenty-second European Conference on Artificial Intelligence*. 515–523.

[53] Reinhard Selten. 1975. Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory* 4 (1975).

[54] Wen Jun Tan, Allan N Zhang, and Wentong Cai. 2019. A graph-based model to measure structural redundancy for supply chain resilience. *International Journal of Production Research* 57, 20 (2019), 6385–6404.

[55] Theodore L Turocy. 2005. A dynamic homotopy interpretation of the logistic quantal response equilibrium correspondence. *Games and Economic Behavior* 51, 2 (2005), 243–263.

[56] Theodore L Turocy. 2010. Computing sequential equilibria using agent quantal response equilibria. *Economic Theory* 42, 1 (2010), 255–269.

[57] Kai Wang, Lily Xu, Andrew Perrault, Michael K Reiter, and Milind Tambe. 2021. Coordinating followers to reach better equilibria: End-to-end gradient descent for Stackelberg games. *arXiv preprint arXiv:2106.03278* (2021).

[58] Rong Yang, Fernando Ordonez, and Milind Tambe. 2012. Computing optimal strategy against quantal response in security games. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 2*. 847–854.

[59] Sisi Yin, Tatsushi Nishi, and Ignacio E Grossmann. 2015. Optimal quantity discount coordination for supply chain optimization with one manufacturer and multiple suppliers under demand uncertainty. *The International Journal of Advanced Manufacturing Technology* 76, 5-8 (2015), 1173–1184.

[60] Youzhi Zhang, Bo An, and Jakub Černý. 2021. Computing ex ante coordinated team-maxmin equilibria in zero-sum multiplayer extensive-form games. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 5813–5821.

A PROOFS

Proposition 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 λ.

Proof. We construct two simple reductions from G to (1) a normal-form game for showing the relation to quantal response equilibrium; and to (2) an extensive-form game to show a relation to an agent quantal response equilibrium.

(1) Let $G' = (N, A', u')$ share the same players with G and $A' = (\Pi_1, \dots, \Pi_n)$, i.e., the pure strategies of G. Utility u' is then defined as $u'_i(a) = \sum_{(a_1|s_1,\dots a_n|s_n)\in\times a} \lambda(s_1, \dots, s_n)u(a_1, \dots, a_n)$ for each action profile $a \in A'$. The expected utilities hence correspond to the definition of utility in over-pure-strategies quandal correlated equilibrium. The quantal response equilibrium in G' is a fixed point of the dynamic and hence satisfies the definition of Π-QCE in G.
(2) For definitions of extensive-form games and agent quantal correlated equilibrium, see [44]. In extensive-form games, the states a player cannot distinguish are grouped into mutually disjunctive information sets. 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 agent quantal response equilibrium, each player acts according to quantal response 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 S-QCE.

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

Proposition 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 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 obtain

$$\beta_2(a_2|s_2) = \frac{\sum_{\pi\in\Pi_2:(a_2,s_2)\in\pi} \prod_{(a'_2,s'_2)\in\pi\setminus(a_2,s_2)} q_2(u_2(a'_2, s'_2))q_2(u_2(a_2, s_2))}{\left(\sum_{a'_2\in A_2} q_2(u_2(a'_2, s_2))\right)\left(\sum_{\pi\in\Pi_2:(a_2,s_2)\in\pi} \prod_{(a'_2,s'_2)\in\pi\setminus(a_2,s_2)} q_2(u_2(a'_2, s'_2))\right)}.$$

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

$$u_1(a_1) = \sum_{a_2\in A_2} \sum_{s_2\in S_2} \lambda(s_2)\beta(a_2|s_2)u(a_1, a_2) = \sum_{a_2\in A_2} \sum_{s_2\in S_2} \sum_{\pi\in\Pi_2:(a_2,s_2)\in\pi} \lambda(s_2)\delta_2(\pi)u(a_1, a_2) = u_1(\pi = a),$$

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. □

Proposition 3. *Let G be a normal-form game. Any quantal response equilibrium in G*

(1) may be extended into a per-signal quantal correlated equilibrium laying on a corner of the signaling simplex; and

(2) is a trivial over-pure-strategies quantal correlated equilibrium with only 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.

□

Proposition 4. *Let $G = (N, A, S, u)$ be a signaling game and $C = \{(\lambda, QCE(\lambda)), \lambda \in \Lambda\}$, where QCE is either S-QCE or Π-QCE. Then 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 C is connected.*

Proof. We proceed similarly as in Theorem 3 in [43]. We observe that as all generators are continuous, both formulations of quantal correlated equilibria may be written as zeros of continuous systems[9] of variables (λ, β) or (λ, δ). Because the systems are continuous, C is closed. As both λ and β, δ are bounded, C is compact. Therefore, the correspondence is upper hemicontinuous. When there exists only one quantal correlated equilibrium for any λ, then C is connected because it is an image of a connected space by a continuous correspondence. □

Proposition 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 a fully rational expected utility that is always positive and negatively correlated with utilities of all players. Then the signaler's utility in quantal correlated equilibrium is greater than in quantal response equilibrium.*

Proof. Assume that λ is a corner of the signaling simplex. According to Proposition 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

[9] We use this fact in the next section to define the equilibrial homotopy.

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. □

Proposition 6. *Let q_p be a parametric generator continuous in p 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 $\{\beta_1, \beta_2, \dots\}$ be a sequence of corresponding quantal correlated equilibria with generators q_{p_t} for a fixed signaling scheme λ. Then $\beta^* = \lim_{t\to\infty} \beta_t$ is a correlated equilibrium for λ.*

Proof. According to Proposition 1, the quantal correlated equilibria are quantal response equilibria in the extended games. By the the same reasoning as in Theorem 2 in [43], 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. 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. □

Proposition 7. *Let G be a signaling game of n players and $q_1, \dots, q_n$ be their respective generators. Let* QCE *be the problem of computing a quantal correlated equilibrium in G. Then* QCE *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 [17]. 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 1 in [47], 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 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. □

Proposition 8. *For arbitrary λ, the solution for $t = 0$ is a uniform strategy for each signal.*

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 $\beta_i^{t=0}(a_i|s_i) = 1/|A_i|$. □

Proposition 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 4.4, $(\beta(p), t(p)) : H_\lambda((\beta(p), t(p))) = 0$, 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 2 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 of [1], 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 laying 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 $H_{\lambda'}$. □

B STEPLENGTH ADAPTATION ALGORITHM

For completeness, in Algorithm 3 we present a simple method for updating the steplength h. The method is a variant of (6.1.10) from [1]. 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.

ALGORITHM 3: Method UpdateStep for adapting the steplength in predictor-corrector

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** $newton \leftarrow True$

$f \leftarrow \max(f, a_f \sqrt{\|c|(\gamma, t)\| / \overline{\epsilon}})$

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** $h,\ \|c|(\gamma, t)\|^\iota,\ f,\ False,\ newton$

end

$f \leftarrow \max(f, a_f \sqrt{\kappa / a_\kappa})$

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 $h,\ \|c|(\gamma, t)\|^\iota,\ f,\ True,\ newton$

C ORIGINAL HOMOTOPY METHOD FAILURE

Here we present an example of an equilibrium tracing failure. Assume that two players who act according to logit generators $q = exp$ engage in a game depicted in Figure 6. In case we trace the quantal correlated equilibrium with trivial single actions using the original homotopy introduced at the beginning of section 4.1, 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 2	
		A	*B*
Player 1	*o*	-2,23	-28,-13
	r	28,10	-22,-9

Fig. 6. An example of a normal-form game where the original homotopy method fails to reach the equilibrium.

D EXPERIMENTAL DOMAINS

In this section, we provide the full description of the experimental domains.

D.1 Randomly generated normal-form games

We construct general-sum games with actions spaces of possibly different size for each player. The utilities are generated uniformly randomly from the interval $[-1, 5]$. 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*[10] – a maximization of a sum of players' utilities – formally defined as

$$welfare(\beta) = \sum_{i \in N} \sum_{a_i \in A_i} \sum_{s_i \in S_i} \beta_i(a_i|s_i) u_i(a_i|s_i).$$

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 $q(x) = e^x$. We set $C = 3$ 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 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.

D.2 Supply chain games

Our definition of a game on supply chains is inspired by [49, 54, 59]. 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. Function $\delta : H \times F \to \mathbb{R}$ serves to identify a distance between a manufacture

[10] 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.

ALGORITHM 4: Material allocation and pricing

Input: action profile π
Output: facility utilization $\mathcal{A}$, material prices $\mathcal{Z}$
for $t \in T$ **do**
  $A_h \leftarrow \left[\pi_p \in \pi \,\middle|\, p \in P,\ \tau_{hf}(\pi_p) = t\right], \qquad A_f \leftarrow \left[\pi_r \in \pi \,\middle|\, r \in R,\ \tau_{hf}(\pi_r) = t\right]$
  while $|A_h| > 0$ *and* $|A_f| > 0$ **do**
    for $f \in A_f$ **do**
      $\overline{\delta}[f] \leftarrow 0, \quad \overline{P}[f] = [], \quad \overline{\mu} \leftarrow \mu_f \upharpoonright f, \quad \overline{A}_h \leftarrow sort\_ascending(A_h, \delta \upharpoonright f)$
      for $h \in \overline{A}_h$ **do**
        $\overline{\delta}[f] \leftarrow \overline{\delta}[f] + \delta(h, f), \quad \overline{P}[f] \leftarrow \overline{P}[f] \cup h, \quad \overline{\mu}(\mu_h(h)) := \overline{\mu}(\mu_h(h)) - 1$
        if $\sum_{m \in M} \overline{\mu}(m) = 0$ **then** break
      end
      if $\sum_{m \in M} \overline{\mu}(m) > 0$ **then** $\overline{\delta}[f] \leftarrow \infty$
    end
    $f^* \leftarrow arg\min_{f \in A_f} \overline{\delta}[f]$
    if $\overline{\delta}[f^*] = \infty$ **then** break
    $A_f \leftarrow A_f \backslash f, \qquad A_h \leftarrow A_h \backslash \overline{P}[f^*], \qquad \overline{\delta}[f^*] \leftarrow \infty$
  end
  for $h \in H \,|\, \tau_{hf}(h) = t$ **do** $\mathcal{A}(h) \leftarrow \left(|\{p \in P \,|\, \pi_p = h\}| - |[h' \in A_h \,|\, h' = h]|\right) / |\{p \in P \,|\, \pi_p = h\}|$
  for $f \in F \,|\, \tau_{hf}(f) = t$ **do** $\mathcal{A}(f) \leftarrow \left(|\{r \in R \,|\, \pi_r = f\}| - |\left[f' \in A_f \,\middle|\, f' = f\right]|\right) / |\{r \in R \,|\, \pi_r = f\}|$
  for $m \in M$ **do**
    $\mathcal{Z}(t, m) \leftarrow \zeta(t, m) - |[h \in A_h \,|\, \mu_h(h) = m]| / |\left\{p \in P \,\middle|\, \mu_h(\pi_p) = m, \tau_{hf}(\pi_p) = t\right\}|$
    $$- \sum_{f \in A_f} \mu_f(f, m) \quad / \sum_{r \in R,\, \tau_{hf}(\pi_r) = t} \mu_f(\pi_r, m)$$
  end
end

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 utilities are given by

$$u_p(\pi) = \mathcal{A}(\pi_p)\mathcal{Z}\left(\tau_{hf}^{-1}(\pi_p), \mu_h(\pi_p)\right) - \zeta_b\left(\tau_{hf}^{-1}(\pi_p), \mu_h(\pi_p)\right) - \zeta_m(p, \pi_p) - \zeta_h(\pi_p)|\left\{i \in P \,\middle|\, \pi_i = \pi_p\right\}|$$
$$u_r(\pi) = \mathcal{A}(\pi_r)\rho(\pi_r) - \upsilon(\pi_r) - \sum_{m \in M} \mu_f(\pi_r, m)\mathcal{A}(\pi_r)\mathcal{Z}\left(\tau_{hf}^{-1}(\pi_r), m\right),$$

where $\mathcal{A}$ is relative utilization of warehouses and manufactures, and $\mathcal{Z}$ is material pricing computed by Algorithm 4. 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 a 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 v 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 [32].

E ADDITIONAL EXPERIMENTS

This section presents two sets of additional experiments suggested by an anonymous reviewer. We would like to point out that we observed two principal issues with BARON: (1) it solves smaller instances within a couple of minutes, but due to the concept's non-convex formulation, after just slightly increasing the action space or adding more players BARON does not terminate within several hours, and (2) it quickly runs out of the 32GB of memory we used in the experiments. The homotopy method seems to suffer from neither. The experiments in the main text were set up so that the sizes of the game instances gradually increase in order to control and show BARON's scalability, which required keeping the number of signals low. Our intention was also to assess the performance over a broader spectrum of domains, generators, etc., rather than focus on specific large instances. In this section, we hence first examine scalability with respect to different cardinalities of signal sets, and then also the algorithm's performance on larger games.

E.1 Varying signal spaces

We ran experiments tracing the quantal correlated equilibrium in both randomly generated normal form games and supply chain games. For every setting, we randomly sampled 10 game instances, and terminated all running processes after 2 hours.

In randomly generated normal form games, we assumed a setting with 3 players choosing from 20, 20, and 2 actions, respectively. The utility range was set to $[-1, 5]$, i.e., the same as in the main-text experiments. 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 7. 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 in the main text, we assume

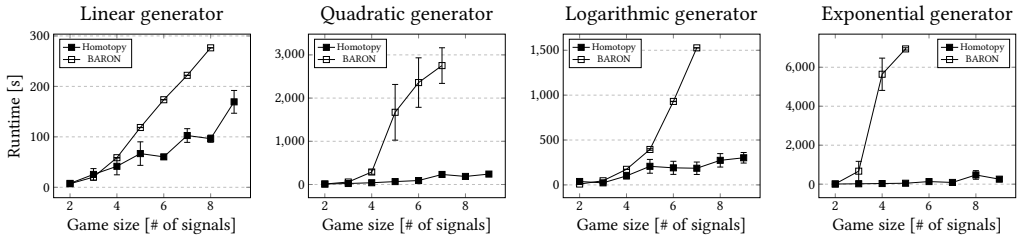


Fig. 7. Mean runtimes of computing (S-QCE) with fixed λ using BARON and the homotopy algorithm in **randomly generated games**. Every point shows also a standard error.

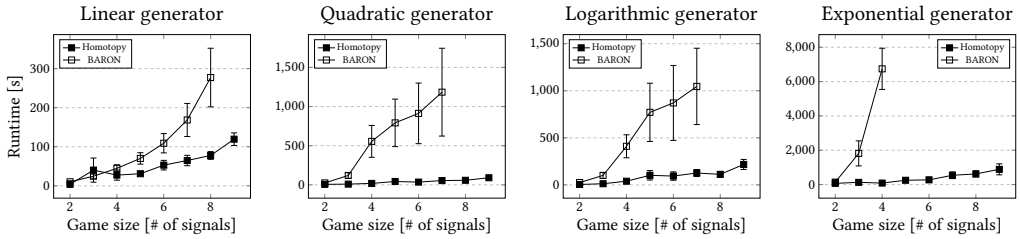


Fig. 8. Mean runtimes of computing (S-QCE) with fixed λ using BARON and the homotopy algorithm in **supply chain games**. Every point shows also a standard error.

only the manufacturers receive one of the $\#s$ signals each. The results could be found in Figure 8. Contrary 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.

E.2 Larger games

Finally, in Figure 9 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 3-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.

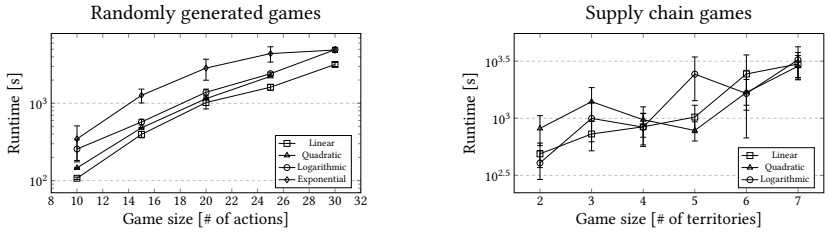


Fig. 9. Scalability of the homotopy method across different generators in (Left) randomly generated games and (Right) supply chain games. Every point shows also a standard error.