Information Theory — The Bridge Connecting Bounded Rational Game Theory and Statistical Physics

David H. Wolpert*

April 28, 2005

Abstract

A long-running difficulty with conventional game theory has been how to modify it to accommodate the bounded rationality of all real-world players. A recurring issue in statistical physics is how best to approximate joint probability distributions with decoupled (and therefore far more tractable) distributions. This paper shows that the same information theoretic mathematical structure, known as Product Distribution (PD) theory, addresses both issues. In this, PD theory not only provides a principled formulation of bounded rationality and a set of new types of mean field theory in statistical physics; it also shows that those topics are fundamentally one and the same.

1 Introduction

In noncooperative game theory, one has a set of N players, each choosing its strategy x_i independently, by sampling a distribution $q_i(x_i)$ over those strategies. Each player i also has her own utility function $g_i(x)$, specifying how much reward she gets for every possible joint-strategy x of all N players. Let $q_{(i)}(x_{(i)})$ mean the joint probability distribution of all players other than i, i.e., $\prod_{j \neq i} q_j(x_j)$. Then the "goal" of each player i is to set q_i to so that, conditioned on $q_{(i)}$, the expected value of i's utility is as high as possible.

Conventional game theory assumes each player i is "fully rational", able to solve for that optimal q_i, and that she then uses that distribution. It is primarily concerned with analyzing such equilibria of the game [3, 4, 5, 6]. In the real world, this assumption of full rationality almost never holds, whether the players are humans, animals, or computational agents [7, 8, 9, 10, 11, 12, 13, 14, 15]. This is due to the cost of computation of that optimal distribution, if nothing

*D. Wolpert is with NASA Ames Research Center, Moffett Field, CA, 94035 dhw@ptolemy.arc.nasa.gov

else. This real-world **bounded rationality** is one of the major impediments to applying conventional game theory in the real world.

More generally, consider any scientific scenario, in which one wishes to make predictions about a particular physical system. To make those predictions it is necessary to first have *some* information / data concerning the system, to serve as the basis of one's prediction. Without such information, science can say nothing, and to pretend otherwise is erroneous. This is true even when the physical system is a set of human players engaged in a game: To make any predictions concerning the players, one must first be provided (or obtain through observation) some information concerning them and the game. Together with known scientific laws, only that provided information should be used in making one's prediction. So in particular, unless one explicitly is provided the information that the players in a game are fully rational, to simply *assume* that they are violates one of the fundamental tenets of how science is done.

This paper shows how Shannon's information theory [16, 17, 18] provides a principled way to modify conventional game theory to accommodate bounded rationality. This is done by following information theory's prescription that, given only partial knowledge concerning the distributions the players are using, we should use the minimum information (Maxent) principle to infer those distributions. Doing so results in the principle that the bounded rational equilibrium is the minimizer of a certain set of coupled Lagrangian functions of the joint distribution, $q(x) = \prod_i q_i(x_i)$. This mathematical structure is a special instance of Product Distribution (PD) theory [19, 20, 21, 22, 23, 24, 11].

In addition to showing how to formulate bounded rationality, PD theory provides many other advantages to game theory. Its formulation of bounded rationality explicitly includes a term that, in light of information theory, is naturally interpreted as a cost of computation. PD theory also seamlessly accommodates multiple utility functions per player. It also provides many powerful techniques for finding (bounded rational) equilibria, and helps address the issue of multiple equilibria. Another advantage is that by changing the coordinates of the underlying space of joint moves x, the same mathematics describes a type of bounded rational cooperative game theory, in which the moves of the players are transformed into contracts they all offer one another.

Perhaps the most succinct and principled way of deriving statistical physics is as the application of the Maxent principle. In this formulation, the problem of statistical physics is cast as how best to infer the probability distribution over a system's states when one's prior knowledge consists purely of the expectation values of certain functions of the system's state [25, 18]. For example, this prescription says we should infer that the probability distribution p governing the system is the Boltzmann distribution when our prior knowledge is the system's expected energy. This is known as the "canonical ensemble". Other ensembles arise when other expectation values are added to one's prior knowledge. In particular, if the number of particles in the system is uncertain, but one knows its expectation value, one arrives at the "grand canonical ensemble".

One major difficulty with working with these ensembles is that under them the particles of the system are statistically coupled with one another. For high-dimensional systems, this can make statistical physics calculations very difficult. Accordingly, a large body of work has been produced under the rubric of Mean Field (MF) theory, in which the ensemble is approximated with a distribution in which the particles are independent [26]. In an MF approximation, a product distribution q governs the joint state of the particles — just as a product distribution governs the joint strategy of the players in a game.

MF approximations are usually derived in an ad hoc manner. The principled way to derive a MF approximation (or any other kind) to a particular ensemble is to specify a distance measure saying how close two probability distributions are, and then solve for the q that is closest to the distribution being approximated, p. To do this one needs to specify the distance measure. How best to measure distances between probability distributions is a topic of ongoing controversy and research [27]. The most common way to do so is with the infinite limit log likelihood of data being generated by one distribution but misattributed to have come from the other. This is known as the Kullback-Leibler (KL) distance [16, 28, 17]. It is far from being a metric. In particular, it is not symmetric under interchange of the two distributions being compared.

It turns out that the simplest MF theories minimize the KL distance from q to p. However it can be argued it is the KL distance from p to q that is the most appropriate measure, not the KL distance from q to p. Using that distance, the optimal q is a new kind of approximation not usually considered in statistical physics.

For the canonical ensemble, the type of KL distance arising in simple MF theories turns out to be identical to the maxent Lagrangian arising in bounded rational game theory. This shows how bounded rational (independent) players are formally identical to the particles in the MF approximation to the canonical ensemble. Under this identification, the moves of the players play the roles of the states of the particles, and particle energies are translated into player utilities. The coordinate transformations which in game theory result in cooperative games are, in statistical physics, techniques for allowing the canonical ensemble to be more accurately approximated with a product distribution.

This identification raises the potential of transferring some of the powerful mathematical techniques that have been developed in the statistical physics community (e.g., extensions of mean field theory [26] or cavity methods [29]) to noncooperative game theory. In also suggests translating some of the other ensembles of statistical physics to game theory, in addition to the canonical ensemble. As an example, in the grand canonical ensemble the number of particles is variable, which, after a MF approximation, corresponds to having a variable number of players in game theory. Among other applications, this provides us with a new framework for analyzing games in evolutionary scenarios, different from evolutionary game theory. Finally, much work has been done in statistical physics on approximations that are higher-order than mean-field, introducing extra random variables that allow for some statistical dependencies coupling the variables. The associated generalization of PD theory is a full-blown theory of Probability Lagrangians.

In the next section noncooperative game theory and information theory are cursorily reviewed. Then bounded rational game theory is derived, and its many advantages are discussed. The following section starts with a cursory review of the information-theoretic derivation of statistical physics. After that is a discussion of the two kinds of KL distance and the MF theories they induce, and a discussion of coordinate systems. This section also includes a discussion on translating a MF version of the grand canonical ensemble into a new kind of evolutionary game theory.

Miscellaneous proofs can be found in the appendix.

As discussed in the physics section, the maxent Lagrangian and associated Boltzmann solution at the core of this paper has been investigated for an extremely long time in the context of many-particle systems. The use of the Boltzmann distribution over possible moves also has a long history in the Reinforcement Learning (RL) literature, i.e., in the design of algorithms for a player involved in an iterated game with Nature [30, 31]. Related work has considered multiple players [32, 33]. In particular, some of that work has been done in the context of "mechanism design" of many players, i.e., in the context of designing the utility functions of the players to induce them to maximize social welfare [34, 35, 36, 37]. In all of this RL work the Boltzmann distribution is usually motivated either as an *a priori* reasonable way to trade off exploration and exploitation, as part of Markov Chain Monte Carlo procedure, or by its asymptotic convergence properties [38].

In addition, independent of the work reported in this paper, the maxent Lagrangian and/or the Boltzmann distribution has previously been muted as a way to model human players [39, 40, 10]. Some of that work has explicitly noted the relation between the Boltzmann distribution and statistical physics [41]. However the motivation of the maxent Lagrangian and Boltzmann distribution in that work is *ad hoc*, based on particular simple models of human decision-making and/or of player interactions. There is no use of information theory to derive the maxent Lagrangian from first principles. Due to this, no connection is made in that previous work between the maxent Lagrangian and the cost of computation, no extension is made to other kinds of prior knowledge concerning the game, there is no recognition of how to modify the Lagrangian for multiple cost functions, there is no extension to the grand canonical ensemble and therefore variable numbers of players, and there is no development of rationality operators, or the relation between semi-coordinate transformations and cooperative game theory. Ultimately, this lack of theoretical underpinnings is also why that previous work did not note the formal identity between the game theory of actual bounded rational human players and MFT.

Finally, it's important to note that PD theory also has many applications in science beyond those considered in this paper. For example, see [21, 22, 42, 43, 44] for work relating the maxent Lagrangian to distributed control and to distributed optimization. See [43] for algorithms for speeding up convergence to bounded rational equilibria. Some of those algorithms are related to simulated and deterministic annealing [28]. In [20] others of those algorithms are related to Stackelberg games, and more generally to the problem of finding the optimal control hierarchy for team of players with a common goal, i.e., finding an optimal organization chart. See also [45, 46, 47] for work showing, respectively, how to use PD theory to improve Metropolis-Hastings sampling, how to relate it to the mechanism design work in [34, 35, 36, 37], and how to extend it to continuous move spaces and time-extended strategies.

2 PD theory as Bounded Rational Noncooperative Game Theory

This section motivates PD theory as a way of addressing several of the shortcomings of conventional noncooperative game theory.

2.1 Review of noncooperative game theory

In noncooperative game theory one has a set of N **players**. Each player i has its own set of allowed **pure strategies**. A **mixed strategy** is a distribution $q_i(x_i)$ over player i's possible pure strategies. Each player i also has a **utility function** g_i that maps the pure strategies adopted by all N of the players into the real numbers. So given mixed strategies of all the players, the expected utility of player i is $E(g_i) = \int dx \ \prod_j q_j(x_j) g_i(x)$ [1].

This basic framework can be elaborated to model many interactions between biological organisms, and in particular between human beings. These interactions range from simple abstractions like the famous prisoner's dilemma to iterated games like chess, to international relations [48, 3, 4].

Much of noncooperative game theory is concerned with **equilibrium concepts** specifying what joint-strategy one should expect to result from a particular game. In particular, in a **Nash equilibrium** every player adopts the mixed strategy that maximizes its expected utility, given the mixed strategies of the other players. More formally, $\forall i, q_i = \text{argmax}_{q'_i} \int dx \ q'_i \prod_{j \neq i} q_j(x_j) \ g_i(x)$.

Several very rich fields have benefited from a close relationship with noncooperative game theory. Particular examples are evolutionary game theory (in which the set of N players is replaced by an infinite set of reproducing organisms) and cooperative game theory (in which players choose which **coalitions** of other players to join) [49, 6]. Game theory as a whole is also closely related to economics, in particular the field of mechanism design, which is concerned with how to induce the set of players to adopt a socially desirable joint-strategy [50, 3, 51, 52].

2.2 Problems with conventional noncooperative game theory

A number of objections to the Nash equilibrium concept have been resolved. In particular, it was Nash who proved that every game has at least one Nash

[1] Throughout this paper, the integral symbol will be interpreted in the appropriate measure-theoretic terms, e.g., as Lebesgue integrals, point-sums, etc.

equilibrium if one expands the realm of discourse to include mixed strategies. (The same is not true for pure strategies.) Other objections have been more or less resolved through numerous **refinements** of the Nash equilibrium concept.

However there are several major problems with the concept that are still outstanding. One of them is the possible multiplicity of equilibria; this multiplicity means the Nash equilibrium concept cannot be used to specify the joint strategy that is actually adopted in a real world game. (Some refinements of the Nash equilibrium concept attempt to address this problem, though none has succeeded.) Another problem is that while calculating Nash equilibria is straightforward in many simple games (e.g., 2 players in a zero-sum game), calculating them in the general case can be a very difficult computational multi-criteria optimization problem. Yet another problem is that there is no general way to extend the concept to allow each player to have multiple utility functions.

Another set of problems has to do with the connection between the Nash equilibrium concept and real world games. In the traditional game theory view, the Nash equilibrium of a game arises all at once, by each player performing the needed calculations beforehand. This requires in particular that each player know the other players' payoff functions. When such an assumption is too egregious, somewhat contorted alternatives like correlated equilibria or Bayesian equilibria are invoked [2].

Perhaps the major example of such a problem with the connection between the Nash equilibrium concept and the real world is its assumption of **full rationality**. This is the assumption that every player i can both calculate what the strategies $q_{j \neq i}$ will be and then calculate its associated optimal distribution. In other words, it is the assumption that every player will calculate the entire joint distribution $q(x) = \prod_j q_j(x_j)$. If for no other reasons than computational limitations of real humans, this assumption is essentially untenable. This problem is just as severe if one allows statistical coupling among the players [53, 3].

Under the rubric of "behavioral economics" a large body of empirical lore has been generated characterizing the bounded rationality of humans. Much (but hardly all) of that lore is encapsulated in formalisms like Prospect theory. Similarly, much has been learned about the empirical behavior of (bounded rational) machine learning computer algorithms playing games with one another [7, 13]. None of this work has resulted in a full mathematical theory of bounded rationality however.

There have also been numerous theoretical attempts to incorporate bounded rationality into noncooperative game theory by modifying the Nash equilibrium concept. Some of them assume essentially that every player's mixed strategy is its Nash-optimal strategy with some form of noise superimposed [6]. Others explicitly model the humans, typically as computationally limited automata,

[2] The "all at once" view of Nash equilibrium is not adopted in Evolutionary game theory. There the Nash equilibrium arises through a sequence of multiple games, as the fixed point of the replicator dynamics. That dynamics is an *ad hoc* learning rule, i.e., a model for how humans update their strategies. Interestingly, a bounded rational version of that very dynamics "falls out" of PD theory, without any need for the fuzzy arguments Evolutionary game theory uses to justify it as a model for human learning. See [46].

and assume the automata perform optimally subject to those computational limitations [10]. Both approaches, while providing insight, are very *ad hoc* when viewed as models of games involving real-world organisms or real-world (i.e., non-trivial) machine learning algorithms.

The rest of this section shows how statistical inference can be used to extend game theory to avoid many of these shortcomings. In a nutshell, the perspective adopted here is purely empiricist — you are a scientist, with limited information about a particular game, and must predict what strategies are being followed in the game. The prediction you make is the associated equilibrium concept.

This is an exercise in statistical inference, an old and well-understood field. No modeling of the players and their thought processes need arise in such inference. However if we are provided hard data (!) concerning how the players behave, those should be taken into account in the inference. In addition, models concerning the behaviors of players can be used, if desired, for example via Bayesian inference. Note that any priors arising in that inference are set by *us*, the scientists predicting the game, not by the players involved in the game. If we believe the players to be Bayesians, that simply means that our inference algorithm must involve probability distributions over the possible priors of the players.

The particular inference algorithm explored in this paper is the standard (non-Bayesian) one arising from information theory. (As discussed below, others are more appropriate in certain circumstances.) The difficulty of calculating equilibria is addressed in the sections below on solving for the distributions of PD theory. In adition, the sections after this one present some other extensions of game theory, in particular to allow for a variable number of players. (Games with variable number of players arise in many biological scenarios as well as economic ones.)

2.3 Review of the minimum information principle

Shannon was the first person to realize that based on any of several separate sets of very simple desiderata, there is a unique real-valued quantification of the amount of syntactic information in a distribution $P(y)$. He showed that this amount of information is (the negative of) the Shannon entropy of that distribution, $S(P) = -\int dy \; P(y) ln[\frac{P(y)}{\mu(y)}]$ [3].

So for example, the distribution with minimal information is the one that doesn't distinguish at all between the various y, i.e., the uniform distribution. Conversely, the most informative distribution is the one that specifies a single possible y. Note that for a product distribution, entropy is additive, i.e., $S(\prod_i q_i(y_i)) = \sum_i S(q_i)$.

Say we are given some incomplete prior knowledge about a distribution $P(y)$. How should one estimate $P(y)$ based on that prior knowledge? Shannon's result

[3] μ is an *a priori* measure over y, often interpreted as a prior probability distribution. Unless explicitly stated otherwise, in this paper we will always assume it is uniform, and not write it explicitly. See [25, 18, 16].

tells us how to do that in the most conservative way: have your estimate of $P(y)$ contain the minimal amount of extra information beyond that already contained in the prior knowledge about $P(y)$. Intuitively, this can be viewed as a version of Occam's razor. This approach is called the minimum information (or "maxent") principle. It has proven extremely useful in domains ranging from signal processing to image processing to supervised learning [17].

2.4 Maxent Lagrangians

Much of the work on equilibrium concepts in game theory adopts the perspective of an external observer of a game. We are told something concerning the game, e.g., the moves sets and utility functions of the separate players, information sets, etc., and from that wish to predict what joint strategy will be followed by real-world players of the game. Say that in addition to such information, we are told the expected utilities of the players. What is our best estimate of the distribution q that generated those expected utility values? By the maxent principle, it is the distribution with maximal entropy, subject to those expectation values.

To formalize this, for simplicity assume a finite number of players, and a finite number of possible moves (pure strategies) for each player. To agree with the convention in other fields, from now on we implicitly flip the sign of each g_i so that the associated player i wants to minimize that function rather than maximize it. Intuitively, this flipped $g_i(x)$ is the "cost" to player i when the joint-strategy is x, rather than its utility then.

So our prior knowledge is that the players are independent, that their cost functions are the $\{g_i\}$, and that their expected utilities are given by the set of values $\{\epsilon_i\}$. The maxent estimate of the q for that prior knowledge is given by the minimizer of the Lagrangian

$$\begin{aligned} \mathscr{L}(q) &\triangleq \sum_i \beta_i [E_q(g_i) - \epsilon_i] - S(q) \\ &= \sum_i \beta_i [\int dx \prod_j q_j(x_j) g_i(x) - \epsilon_i] - S(q) \end{aligned} \tag{1}$$

where the subscript on the expectation value indicates that it evaluated under distribution q, and the $\{\beta_i\}$ are Lagrange parameters implicitly set by the constraints on the expected utilities [4].

Solving, we find that the mixed strategies minimizing the Lagrangian are related to each other via

$$q_i(x_i) \propto e^{-E_{q_{(i)}}(G|x_i)} \tag{2}$$

where the overall proportionality constant for each i is set by normalization, and $G \triangleq \sum_i \beta_i g_i$, and the subscript $q_{(i)}$ on the expectation value indicates that

[4]Throughout this paper the terms in any Lagrangian that restrict distributions to the unit simplices are implicit. The other constraint needed for a Euclidean vector to be a valid probability distribution is that none of its components are negative. This will not need to be explicitly enforced in the Lagrangian here.

it is evaluated according to the distribution $\prod_{j\neq i} q_j$. In Eq. 2 the probability of player i choosing pure strategy x_i depends on the effect of that choice on the utilities of the other players. This reflects the fact that our prior knowledge concerns all the players equally.

If we wish to focus only on the behavior of player i, it is appropriate to modify our prior knowledge. To see how to do this, first consider the case of maximal prior knowledge, in which we know the actual joint-strategy of the players. For this case, trivially, the maxent principle says we should "estimate" q as that joint-strategy (it being the q with maximal entropy that is consistent with our prior knowledge). The same conclusion holds if our prior knowledge also includes the expected cost of player i.

Now modify this maximal set of prior knowledge by removing from it specification of player i's strategy. So our prior knowledge is the mixed strategies of all players other than i, together with player i's expected cost. We can incorporate the prior knowledge of the other players' mixed strategies directly into our Lagrangian, without introducing Lagrange parameters. That **maxent Lagrangian** is

$$\begin{aligned}\mathscr{L}_i(q_i) &\triangleq \beta_i[E_q(g_i) - \epsilon_i] - S_i(q_i) \\ &= \beta_i[\int dx \prod_j q_j(x_j) g_i(x) - \epsilon_i] - S_i(q_i).\end{aligned}$$

All of these Lagrangians (one for each i) are jointly solved at a q given by a set of coupled **Boltzmann distributions**:

$$q_i^B(x_i) \propto e^{-\beta_i E_{q_{(i)}}(g_i|x_i)} \tag{3}$$

where the $\{\beta_i\}$ are Lagrange parameters enforcing our constraints in the usual way. Following Nash, we can use Brouwer's fixed point theorem to establish that for any fixed set of non-negative values $\{\beta_i\}$, there must exist at least one product distribution given by the product of these Boltzmann distributions (one term in the product for each i).

The first term in $\mathscr{L}_i$ is minimized by a perfectly rational player. The second term is minimized by a perfectly *irrational* player, i.e., by a perfectly uniform mixed strategy q_i. So β_i in the maxent Lagrangian explicitly specifies the balance between the rational and irrational behavior of the player. In particular, for $\beta_i \to \infty$, by minimizing the Lagrangians we recover the Nash equilibria of the game. More formally, in that limit the set of q that simultaneously minimize the Lagrangians is the same as the set of delta functions about the Nash equilibria of the game. The same is true for Eq. 2.

The $\beta < \infty$ solutions of Eq. 3 can also be viewed as "equilibra" in the conventional game theory sense, of being a self-consistent set of mixed strategies of the players. To see this, posit that for each player there is a rule (implicit or otherwise) for how it sets its mixed strategy. Say that each player's rule for forming its mixed strategy is based on the expected costs of each of that player's pure strategies. Have each player's rule take the form of a Boltzmann distribution over those expected costs for each of the player's possible pure strategies. (Such a rule may reflect cost of computation (see below), desire by the player to explore as well as exploit, inherent psychological biases, etc.) Then the $\beta < \infty$ solutions of Eq. 3 constitute a joint mixed strategy where all the players follow their separate rules in a globally consistent manner.

Eq. 2 is just a special case of Eq. 3, where all player's share the same cost function G. (Such games are known as **team games**.) Due to this, our guarantee of the existence of a solution to the set of maxent Lagrangians implies the existence of a solution of the form Eq. 2.

Typically players aren't close to perfectly self-defeating. Almost always they will be closer to minimizing their expected cost than maximizing it. For prior knowledge consistent with such a case, the β_i are all non-negative. Examples of games and their associated bounded rational equilibria can be found below in Sec. 2.11, after the discussion of rationality operators.

Finally, our prior knowledge often will not consist of exact specification of the expected costs of the players, even if that knowledge arises from watching the players make their moves. Such other kinds of prior knowledge are addressed in several of the following subsections.

2.5 Alternative interpretations of Lagrangians

There are numerous alternative interpretations of these results. For example, change our prior knowledge to be the entropy of each player i's strategy, i.e., how unsure it is of what move to make. Now we cannot use information theory to make our estimate of q. Given that players try to minimize expected cost, a reasonable alternative is to predict that each player i's expected cost will be as small as possible, subject to that provided value of its entropy and the other players' strategies. The associated Lagrangians are $\alpha_i[S(q_i) - \sigma_i] - E(g_i)$, where σ_i is the provided entropy value for player i. This is equivalent to the maxent Lagrangian, and in particular has the same form of solution, Eq. 3.

Another alternative interpretation involves **world cost** functions, which are quantifications of the quality of a joint pure strategy x from the point of view of an external observer (e.g., a system designer, the government, an auctioneer, etc.). A particular class of world cost functions are "social welfare functions", which can be expressed in terms of the cost functions of the individual players. Perhaps the simplest example is $G(x) = \sum_i \beta_i g_i(x)$, where the β_i serve to trade off how much we value one player's cost vs. anothers. If we know the value of this social welfare function, but nothing else, then maxent tells us to minimize the Lagrangian of Eq. 1.

An important aspect of any of these interpretations is that typically one does not have to explicitly specify the values in one's "prior knowledge", e.g., ϵ_i. This is because typically the associated Lagrange parameters are montonic functions of those "prior knowledge" values, e.g., β_i is monotically increasing in ϵ_i [43]. So it suffices to only specify the values of the Lagrange parameters; having "prior knowledge" of an expected value is purely notional. For example, we can take our prior knowledge to be a value of β_i (which sets ϵ_i) rather than a value of ϵ_i (which sets the equilibrium value for β_i). This is formalized in the subsection on rationality operators, where the prior knowledge is explicitly formulated as the values of Lagrange parameters.

2.6 Bounded rational game theory

In many situations we have prior knowledge different from (or in addition to) expected values of cost functions. This is particularly true when the players are human beings (so that behavioral economics studies can be brought to bear) or simple computational algorithms. To apply information theory in such situations, we simply need to incorporate that prior knowledge into our Lagrangian(s).

To give a simple example, say that we know that the players all want to ensure not just a low expected cost, but also that the actual cost doesn't vary too much from one sample of q to the next. We can formalize this by saying that in addition to expected costs, our prior knowledge includes variances in the costs. Given the expected values of the costs, such variances are specified by the expected values of the squares of the cost. Accordingly, all our prior knowledge is in the form of expectation values. Modifying Eq. 3 appropriately, we arrive at the solution

$$q_i(x_i) \quad \propto \quad e^{-E_{q_{(i)}}(\alpha_i(g_i-\lambda_i)^2|x_i)}. \tag{4}$$

where the Lagrange parameters α_i and λ_i are given by the provided expectations and variances of the costs of the players.

Eq. 4 is our best guess for what the actual mixed strategy of player i is, in light of our prior knowledge concerning that player. Note that this formula directly reflects the fact that player i does make her decision based only on minimizing cost, i.e., maximizing utility. In this, we are directly incorporating the possibility that the player violates the axioms of utility theory — something not allowed in conventional game theory. Other behavioral economics phenomena like risk aversion can be treated in a similar fashion.

A variant of this scenario would have our prior knowledge only give the variances of the costs of the players and not their expected costs. In this cost the Lagrangian must involve a term quadratic in q, in addition to the entropy term and a term linear in q. (See the subsection on multiple cost functions.) More generally, our prior knowledge can be any nonlinear function of q. In addition, even if we stick to prior knowledge that is linear in q, that knowledge can couple the cost functions of the players. For example, if we know that the expected difference in cost of players i and j is ϵ, the associated Lagrange constraint term is $\int dxq(x)[g_i(x) - g_j(x) - \epsilon]$. In this situation our prior knowledge couples the strategies of the players, even though those players are independent. See the discussions on constrained optimization in [23, 21].

2.7 Cost of computation

As mentioned above, bounded rationality is an unavoidable consequence of the cost of computation to player i of finding its optimal strategy. Unfortunately, one cannot simply incorporate that cost into g_i, and then presume that the player acts perfectly rationally for this new g_i. The reason is that this cost is associated with the entire distribution $q_i(x_i)$ that player i calculates; it not associated with some particular joint-strategy formed by sampling such a distribution.

How might we quantify the cost of calculating q_i? The natural approach is to use information theory. Indeed, that cost arises naturally in the bounded rationality formulation of game theory presented above. To see how, for each player i define

$$f_i(x, q_i(x_i)) \triangleq \beta_i g_i(x) + \ln[q_i(x_i)].$$

Then we can write the maxent Lagrangian for player i as

$$\mathscr{L}_i(q) = \int dx \; q(x) f_i(x, q_i(x_i)). \tag{5}$$

Now in a bounded rational game every player sets its strategy to minimize its Lagrangian, given the strategies of the other players. In light of Eq. 5, this means that we *can* interpret each player in a bounded rational game as being perfectly rational for a cost function that incorporates its computational cost. To do so we simply need to expand the domain of "cost functions" to include probability values as well as joint moves.

Similar results hold for non-maxent Lagrangians. All that's needed is that we can write such a Lagrangian in the form of Eq. 5 for some appropriate function f_i.

2.8 Geometry of the Lagrangian

In this subsection we investigate the surface $\mathscr{L}_i$ with β_i and ϵ_i both treated as fixed parameters. (So in particular, $E_q(g_i)$ need not equal ϵ_i.) For simplicity, we consider team games where all $g_i = G$.

In general the geometry of any such surface $\mathscr{L}_i$ is not trivially related to that of the underlying $G(x)$, even in the limit where all $\beta_i \to \infty$. For example, in general $G(x)$ can have a local minimum at $x = x'$ while the associated $\mathscr{L}_i(q)$ does not have a local minimum at a q that is a delta function about x'. This will be the case for example if $G(x'') < G(x')$ for an x'' that does not neighbor x'; shifting an infinitesimal amount of probability mass from the delta function centered at x' to a new one centered at x'' will reduce $E(G)$ and therefore (given our choice of the β_i) it will reduce $\mathscr{L}_i(q)$.

To investigate the geometry of this Maxent Lagrangian we must be careful to distinguish between the various possible parameterizations of distributions. To begin, let $\mathcal{P}$ be the space of distributions (product or otherwise) over our variables. In addition define $\mathcal{Q} \subset \mathcal{P}$ as the set of product distributions over X. Also for all $i = 1, \ldots ..., n$, define $o(X_i)$ as the number of possible values x_i

(for simplicity taken to be finite), and write $o(X) = \prod_i o(X_i)$. So the simplex $\mathcal{P} \subset \mathbb{R}_+^{o(X)}$.

Let t be an arbitrary element of $\mathbb{R}_+^{o(X)}$, not necessarily normalized. The entropy with its domain extended to all t, $-\int dx\ t(x)\text{ln}t(x)$, is a concave function over $\mathbb{R}_+^{o(X)}$. Therefore it is also concave over any subset of $\mathbb{R}_+^{o(X)}$, and in particular the entropy of normalized vectors $p \in \mathcal{P}$ with components $p(x)$ is concave. In addition $\int dx\ G(x)t(x)$ is a convex function of t. So the Maxent Lagrangian $L(p \in \mathcal{P})$ is a convex function. Since $\mathcal{P}$ is a convex space, this means the Maxent Lagrangian has at most one minimum over $\mathcal{P}$.

However we are restricting attention to $\mathcal{Q}$ (to reduce the number of parameters we're going to be using, if for no other reason). In general this is not a convex subspace of $\mathcal{P}$; if $p \in \mathcal{Q}$ and $p' \in \mathcal{Q}$, then distributions on the line connecting p and p' may not be. So even though the Maxent Lagrangian is convex over $\mathcal{Q}$, we do not have guarantees of a single local minimum. (See the appendix.)

There are other ways to parameterize product distributions however. In particular, consider the space $\mathbb{R}_+^{\sum_i o(X_i)}$. The most appropriate way to express a product distribution q is as a vector in this space, since that way we can assign each separate distribution q_i to a different set of components. For example, we can parameterize any $q \in \mathcal{Q}$ in terms of vectors in the space $\mathbb{R}_+^{\sum_i o(X_i)}$ via the **Naive Distributed Parameterization** (NDP), $q(x) = \prod_i q_i(x_i)$.[5]

Unlike $\mathcal{Q}$ itself, its pre-image in $\mathbb{R}_+^{o(X_i)}$ under the NDP is convex; it is the set of $\{q_i\}$ such that $\int dx_i\ q_i(x_i) = 1\ \forall i$, i.e., the n-fold Cartesian product of unit simplices. In addition, say we express the entropy via the NDP as $S(\{q_i\}) = -\int dx\ [\prod_i q_i(x_i)]\text{ln}(q_i(x_i))$. If we restrict each q_i to be normalized, then $S(\{q_i\}) = -\sum_i \int dx_i\ q_i(x_i)\text{ln}(q_i(x_i))$, and is concave, as before.

As an aside, note that if we allow $\{q_i\}$ to range over all $\mathbb{R}_+^{\sum_i o(X_i)}$, then the entropy as just defined is not concave in general:

$$-\int dx\ [\prod_i q_i(x_i)]\text{ln}(q_i(x_i) \quad \neq \quad -\sum_i \int dx_i\ q_i(x_i)\text{ln}(q_i(x_i))$$

for non-normalized $\{q_i\}$, the Hessian of S has off-diagonal entries, and its eigenvalues can have mixed signs. To avoid ever confronting this issue, it is expedient to simply redefine how we extend the domain of the entropy function for the NDP, as

$$S(\{q_i\}) \quad \triangleq \quad -\sum_i \int dx_i\ q_i(x_i)\text{ln}(q_i(x_i)).$$

Return now to the issue of the convexity of our particular (normalized distributions) optimization problem under the NDP. While with the NDP our feasible

[5]Note that $\mathbb{R}_+^{\sum_i o(X_i)}$ is not resticted to normalized vectors. Accordingly, this parameterization of $\mathcal{Q}$ is many-to-one, i.e., many $\mathbb{R}_+^{\sum_i o(X_i)}$ vectors with components $\{q_i(x_i)\}$ map to the same $q \in \mathcal{Q}$.

set is convex as desired, and the entropy term of the Maxent Lagrangian is concave as desired, now expected G presents problems: $\int dx\ G(x) \prod_i q_i(x_i)$ is a multilinear function of the components of $\{q_i\}$, and therefore is not convex.

We can fix this while still using the space $\mathbb{R}_+^{\sum_i o(X_i)}$, so long as we change from the NDP. For example, since geometric means are concave functions, $\int dx \prod_i [r_i(x_i)]^{1/n} G(x)$ is a convex function of $\{r_i\}$, so long as $G(x) \leq 0\ \forall x$. Unfortunately, in this new parameterization the subtracted entropy term in the Lagrangian, $\sum_i \int dx_i\ [r_i(x_i)]^{1/n} \ln([r_i(x_i)]^{1/n})$, is concave rather than convex. A more important problem is that the set of r such that $\prod_i [r_i(x_i)]^{1/n}$ is normalized is not convex. So for simplicity, from now on we will use the NDP, and write "q" as shorthand for $\{q_i\}$.

Consider L as a function of q, with β and γ both treated as fixed parameters. (So in particular, $E_q(g)$ need not equal γ.) First, say that $q_{(i)}$ is held fixed, with only q_i allowed to vary. This makes $E(g_i)$ be linear in q_i. In addition, entropy is a concave function, and the unit simplex is a convex region. Accordingly, the Lagrangian of Eq. 3 has a unique local minimum over q_i. So there is no issue of choosing among multiple minima when all of $q_{(i)}$ is fixed. Nor is there any problem of "getting trapped in a local minimum" in a computational search for that minimum. Indeed, in this situation we can just jump directly to that global optimum, via Eq. 3. All of this is also true if we are considering the Lagrangian $\mathscr{L}_{j \neq i}$ rather than $\mathscr{L}_i$; the function from i's strategy to j's Lagrangian has a single optimum, interior to i's simplex.

Now introduce the shorthand

$$[U]_{i,p}(x_i) \triangleq \int dx_{(i)} U(x_i, x_{(i)}) p(x_{(i)} \mid x_i),$$

so that $[g_i]_{i,q_{(i)}}(x_i)$ is player i's effective cost function, $E_{q_{(i)}}(g_i \mid x_i)$. Consider the value $E_{q_i^B}([g_i]_{i,q_{(i)}})$. This is the value of $E(g_i)$ at i's bounded rational equilibrium for the fixed $q_{(i)}$, i.e., it is the value at the minimum over q_i of $\mathscr{L}_i$. View that value as a function of β_i. One can show that this is a decreasing function. In fact, its derivative just equals the negative of the variance of $[g_i]_{i,q_{(i)}}(x_i)$ evaluated under distribution $q_i^B(x_i)$. Since $E(g_i)$ is bounded below (for bounded g_i), this means that that variance must go to zero for large enough β_i. So as β_i grows, $q_i^B(x_i) \to 0$ for all x_i that don't minimize $E_{q_{(i)}}(g_i \mid x_i)$. In other words, in that limit, q_i becomes Nash-optimal.

One can generalize this result as

$$\frac{\partial E(g_i)}{\partial \beta_j} \quad = \quad -\mathrm{Cov}([g_i]_{j,q^B_{(j)}}, [g_j]_{j,q^B_{(j)}}).$$

In this equation Cov(., .) is just the usual covariance, here evaluated between the two functions $E(g_i \mid x_j)$ and $E(g_j \mid x_j)$, taken as functions of the random variable x_j (which is distributed according to q_j^B).

This result has a very intuitive explanation. β_j measures how much player j's mixed strategy is weighted towards optimizing j's cost function, i.e., how rational player j is. (See the subsection on rationality operators below.) So the result says that the first-order dependence of i's cost function on player j's rationality is just given by the covariance of the two players' cost functions. The more independent those cost functions are (under q_j^B), the less that making j more rational will affect i's cost.

Next consider varying over all $q \in \mathcal{Q}$, the space of all product distributions q. Over this space, the $E(g_i)$ term in $\mathscr{L}_i$ is multilinear. So $\mathscr{L}_i$ is not a simple convex function of q. This is true even for a team game, with shared β_i, for which case every i has the same Lagrangian. So we do not have the guarantees of a single local minimum provided by convexity even in this case.

To further analyze the shape of the team game Lagrangian as a function of q, we start with the following lemma, which extends the technique of Lagrange parameters to off-equilibrium points:

Lemma 1: Consider the set of all vectors leading from $x' \in \mathbb{R}^n$ that are, to first order, consistent with a set of constraints over $\mathbb{R}^n$. Of those vectors, the one giving the steepest ascent of a function $V(x)$ is $\vec{u} = \nabla V + \sum_i \lambda_i \nabla f_i$, up to an overall proportionality constant, where the λ_i enforce the first order consistency conditions, $\vec{u} \cdot \nabla f_i = 0 \;\; \forall i$.

Note that the gradient of entropy is infinite at the border of $\mathcal{Q}$, since at least one $\ln(q_i)$ term will be negative infinite there. Combined with Lemma 1, this can be used to establish that at the edge of $\mathcal{Q}$, the steepest descent direction of any player's Lagrangian points into the interior of $\mathcal{Q}$ (assuming finite β and $\{g_i\}$). (This is reflected in the equilibrium solutions Eq. 3.) Accordingly, whereas Nash equilibria can be on the edge of $\mathcal{Q}$ (e.g., for a pure strategy Nash equilibrium), in bounded rational games any equilibrium must lie in the interior of $\mathcal{Q}$. In other words, any equilibrium (i.e., any local minimum) of a bounded rational game has non-zero probability for all joint moves. So just as when only varying a single q_i, we never have to consider extremal mixed strategies in searching for equilibria over all $\mathcal{Q}$. We can use local descent schemes instead [21, 23, 43].

Lemma 1 can also be used to construct examples of games with more than one bounded rational equilibrium (just like there are games with more than Nash equilibrium). One can also show that for every player i and any point q interior to $\mathcal{Q}$, there are directions in $\mathcal{Q}$ along which i's Lagrangian is locally convex. Accordingly, no player's Lagrangian has a local maximum interior to $\mathcal{Q}$. So if there are multiple local minima of i's Lagrangian, they are separated by saddle points across ridges. In addition, the uniform q is a solution to the set of coupled equations Eq. 3 for a team game, but typically is not a local minimum, and therefore must be a saddle point.

Say we modify the Lagrangians to be defined for all possible p, not just those that are product distributions. For example the Lagrangian of Eq. 1 becomes

$$\mathscr{L}(p) \triangleq \sum_i \beta_i [\int dx \; g_i(x) p(x) - \epsilon_i] - S(p).$$

The first term in this Lagrangian is linear in p. Since entropy is a concave function of the Euclidean vector p over the unit simplex, this means that the overall Lagrangian is a convex function of p over the space of allowed p. This means there is a *unique* minimum of the Lagrangian over the space of all possible legal p. Furthermore, as mentioned previously, for finite β at least one of the derivatives of the Lagrangian is negative infinite at the border of the allowed region of p. This means that the unique minimum of the Lagrangian is interior to that region, i.e., is a legal probability distribution.

In general this optimal p will not be a product distribution, of course. Rather the strategy choices of the players are typically statistically coupled, under this p. Such coupling is very suggestive of various stochastic formulations of non-cooperative game theory. Coupling also arises in cooperative game theory, in which binding contracts couple the moves of the players [48, 6].

Similarly, as is proven in the appendix, the Lagrangian $\mathscr{L}(p) = \beta \sum_i [E_p(g_i)]^2 - S(p)$ is convex over the manifold of legal p, assuming non-negative β. So the model of mechanism design introduced below in Sec. 2.9 has a unique equilibrium — if we allow the players to be statistically coupled.

2.9 Multiple cost functions per player

Say player i has several different cost functions $\{g_i^j\}$ and wants to choose a strategy that will do well at all of them. In the case of pure strategies we can simply "roll up" the cost functions into an aggregate function and employ that in a conventional, single-cost-function-per-player game theoretic analysis. An aggregate cost function like $\frac{\sum_j g_i^j(x)}{\sum_j 1}$ would not necessarily work, since it may be that the pure strategy x minimizing that sum results in a relatively large value for one of the $g_i^j(x)$. However by construction, minimizing a function like $\max_j g_i^j(x)$ will ensure that no particular cost function is favored over the others. Player i will perform well according to such an aggregate function iff it performs well according to all of the constituent g_i^j.

One might think that for mixed strategies one could similarly roll up the cost functions and say that player i works to minimize an aggregate cost function. However especially when player i has many cost functions, it may be that performance according to one or more of the constituent cost functions is quite bad even though the performance according to this average function is good. In particular, it may be that player i has relatively low value of the expectation of the maximum of its cost functions, even though the maximum of the expected costs is quite high [6]. More generally, we cannot ensure that the expected costs

[6] This can even occur if all players other than i are playing pure strategies. For example, say that the number of cost functions is one less than N, the number of potential moves available to player i'. Say that for the pure strategies of the other players, we can write $g_i^j(x) = a\delta_{x_i,x_j} + b\delta_{x_i,N}$ where $a > b$. Then $E(\max_j g_i^j(x))$ is minimized by the mixed strategy $q_i(x_i) = \delta_{x_i,N}$, which results in $E(g_i^j) = b$ for all j. So the worst-case (over cost functions) expected cost for this mixed strategy is b. On the other hand the uniform strategy results in $E(g_i^j) = (a+b)/N$ for all j, i.e., for this mixed strategy the worst-case expected cost is $(a+b)/N$. That difference in worst-case expected costs may be very large; the q_i optimizing $E(\max_j g_i^j(x))$ is very different from the one optimizing $\max_j [E_q(g_i^j)]^2$, giving a very different value of $\max_j [E_q(g_i^j)]^2$.

of player i, $E_q(g_i^j) = \int dx\; g_i^j(x) q_i(x) q_{(i)}(x_{(i)})$, all have good values by appropriately defining an aggregate g_i and requiring only that $\int dx\; g_i(x) q_i(x) q_{\ (i)}(x_{(i)})$ is good. Instead, we must redefine the goal of "minimizing expected costs".

One way to reformulate our goal proceeds by analogy with the goal typically ascribed to a player in pure strategy games. This analogy is based on viewing the cost function for player i as controlled by a fictional player in a meta-game. The concept of dominant strategies in conventional game theory involves having player i choose a pure strategy to minimize the worst case (over other players' moves) cost to i, i.e., to minimize $\max_{x_{(i)}} g^i(x_i, x_{(i)})$. Here the analogy would be for the player to choose a mixed strategy to minimize the worst case (over moves by the fictional player) expected cost, i.e., to minimize $\max_j E_q(g_i^j)$.

A similar solution, appropriate when all of the cost functions are nowhere-negative, is for player i to minimize $\sum_j [E_q(g_i^j)]^2$. Due to the convexity of the squaring operator such minimization will help ensure that no single expectation value $E_q(g_i^j)$ is too high [7]. Indeed, consider increasing the power we raise the costs to, getting the function $[\sum_j [E_q(g_i^j)]^n]^{1/n}$. Minimizing this for large n will approximate the lim-sup norm, which would force all g_i^j to have the same (as low as possible) expectation value.

As far as the math is concerned, $\sum_j [E_q(g_i^j)]^2$ is just a "Lagrangian" of q, one that is convex like the Lagrangian in Eq. 3. If we wish, we can modify such a Lagrangian to incorporate bounded rationality, to force the solution to be interior to $\mathcal{Q}$, getting Lagrangians like $\sum_j \beta_j [E_q(g_i^j)]^2 - S(q_i)$, where the β_j determine the relative rationalities of player i according to its various cost functions.

These kinds of Lagrangians can also model the process of mechanism design, where there is an external designer who induces the players to adopt a desirable joint-strategy [3]. As an example, "desirable" sometimes means that no single player's expected cost is high. A system that meets this goal fairly well can be modeled with a Lagrangian involving terms like $\sum_i [E_q(g_i)]^2$.

2.10 Rationality operators

Often our prior knowledge will not concern expected costs. In particular, this is usually true if our prior knowledge is provided to us before the game is played, rather than afterward. In such a situation, prior knowledge will more likely concern the "intelligences" of the players, i.e., how close they are to being rational. In particular, if we want our prior knowledge concerning player i to be relatively independent of what the other players do, we cannot use i's expected cost as

[7] Choosing q_i to minimize the expectation value $\int dx\; q(x) \sum_j [g_i^j(x)]^2$ will do a roughly similar thing to minimizing $\sum_j [E_q(g_i^j)]^2$, in that it will help ensure that $q_i(x_i)$ is small where the individual $E(g_j \mid x_i)$ are large. However it will also favor having small variances in the value of the costs, perhaps at the expense of the expected values of the costs: $E_q(\sum_j [g_i^j]^2) = \sum_j ([E_q(g_i^j)]^2 + \mathrm{Var}_q(g_i^j))$. In accord with conventional game theory and the axiomatization of utility, here we assume players are interested in expected costs (negatie utilities), not variances in those costs.

our prior knowledge. Our prior knowledge will often concern how peaked i's mixed strategy is about whichever of those of its moves minimizing its cost (or how peaked we can assume it to be), not the associated minimal cost values.

Formally, the problem faced by player i is how to set its mixed strategy $q_i(x_i)$ so as to maximize the expected value of its **effective cost function**, $E(g_i \mid x_i)$. Generalizing, what we want is a *rationality* operator $R(U, p)$ that measures how peaked an arbitrary distribution $p(y)$ is about the minimizers of an arbitrary cost function $U(y)$, $\text{argmin}_y U(y)$.

The naive choice of such an operator might be the entropy of p. However this choice would falsely characterize a player i as irrational if its Nash optimal strategy were not a pure strategy, but rather had extended support across x_i. To correct this problem we could use the entropy of the U-space distribution, $P(u) \triangleq \int dx \delta(U(x) - u) p(x)$. However this choice still has problems. In particular, it doesn't quantify how peaked $P(u)$ is about low U values, just how peaked it is in general.

Formally, we make two requirements of R:

1. If $p(y) \propto e^{-\beta U(y)}$, for non-negative β, then it is natural to require that the peakedness of the distribution — its rationality value — is β.

2. We also need to also specify something of $R(U, p)$'s behavior for non-Boltzmann p. It will suffice to require that of the p satisfying $R(U, p) = \beta$, the one that has maximal entropy is proportional to $e^{-\beta U(y)}$. In other words, we require that the Boltzmann distribution maximizes entropy subject to a provided value of the rationality operator.

As an illustration, a natural choice for $R(U, p)$ would be the β of the Boltzmann distribution that "best fits" p. Information theory provides us such a measure for how well a distribution p_1 is fit by a distribution p_2. This is the **Kullback-Leibler distance** [16, 28]:

$$KL(p_1 \mid\mid p_2) \triangleq S(p_1 \mid\mid p_2) - S(p_1) \tag{6}$$

where $S(p_1 \mid\mid p_2) \triangleq -\int dy \; p_1(y) \ln[\frac{p_2(y)}{\mu(y)}]$ is known as the **cross entropy** from p_1 to p_2 (and as usual we implicitly choose uniform μ).

The KL distance is always non-negative, and equals zero iff its two arguments are identical. In addition, $KL(\alpha p^1 + (1 - \alpha) p^2 \mid\mid p^2)$ is an increasing function of $\alpha \in [0.0, 1.0]$, i.e., as one moves along the line from p^1 to p^2, the KL distance from p^1 to p^2 shrinks.[8] The same is true for $KL(p^2 \mid\mid \alpha p^1 + (1 - \alpha) p^2)$. In addition, those two KL distances are identical to 2nd order about $\alpha = 0$. However they differ as one moves away from $\alpha = 0$ in general; KL distance is not a symmetric function of its arguments. In addition, it does not obey the triangle inequality, although it obeys a variant [16]. Despite these shortcomings, it is by far the most common way to measure the distance between two distributions.

[8]This follows from the fact that the second derivative with respect to α is non-negative for all α, combined with the fact that KL distance is never negative and equals 0 when $\alpha = 0$.

Define $N(U) \triangleq \int dy\, e^{-U(y)}$, the normalization constant for the distribution proportional to $e^{-U(y)}$. (This is called the **partition function** in statistical physics.) Then using the KL distance, we arrive at the rationality operator

$$\begin{aligned} R_{KL}(U,p) &\triangleq \text{argmin}_{\beta} KL(p \mid\mid \frac{e^{-\beta U}}{N(\beta U)}) \\ &= \text{argmin}_{\beta}[\beta \int dy\, p(y)U(y) + \ln(N(\beta U))]. \end{aligned}$$

In the appendix it is proven that R_{KL} respects the two requirements of rationality operators. Note that the argument of the argmin is globally convex (as a function of the minimizing variable β). In addition its second derivative is given by the variance (over y) of the Boltzmann distribution $e^{-\beta U(y)}/N(\beta U)$. This typically makes numerical evaluation of R_{KL} quite fast.

The quantity $\ln(N(\beta U))$ appearing in the second equation, when scaled by β^{-1}, is called the **free energy**. It is easy to verify that it equals the Lagrangian $E_p(U) - S(p)/\beta$ if p is given by the Boltzmann distribution $p(y) \propto e^{-\beta U(y)}$.

Say our prior knowledge is $\{\rho_i\}$, the rationalities of all the players for their associated effective cost functions. Then the Lagrangian for our prior knowledge is

$$\mathscr{L}(q) \;=\; \sum_i \lambda_i [R([g_i]_{i,q}, q_i) - \rho_i] \;-\; S(q). \tag{7}$$

where the λ_i are the Lagrange parameters. Just as before, there is an alternative way to motivate this Lagangian: if our prior knowledge consists of the entropy of the joint system, and we assume each player will have maximal rationality subject to that prior knowledge, we are led to the Lagrangian of Eq. 7.

Say $q_{(i)}$ is fixed. Then for the KL rationality operator, the set of all q_i with the same rationality value ρ_i also share the same value of $E_q(g_i)$. Moreover, that expected value of g_i is exactly the one that would arise if $q_i(x_i)$ were a Boltzmann distribution over values $E(g_i \mid x_i)$ with temperature ρ_i^{-1}.

More formally, it is shown in the appendix that for the Kullback-Leibler rationality operator, we can replace any constraint of the form $R([g_i]_{i,q}, q_i) = \rho_i$ with $E_q(g_i) = \int dx\; g_i(x) \frac{e^{-\rho_i E(g_i \mid x_i)}}{N(\rho_i g_i)} q_{(i)}(x_{(i)})$. For fixed $q_{(i)}$, the set of all q_i obeying the one constraint is identical to the set of all q_i obeying the other one. So knowing that player i has KL rationality ρ_i is equivalent to knowing that the actual expected value of g_i under q_i equals the "ideal expected value", in which q_i is replaced by the Boltzmann distribution of Eq. 3 with $\beta = \rho_i$. However since this knowledge does not specify $q_{(i)}$, it does not specify that ideal expected value of g_i. This contrasts with the prior knowledge underlying the Lagrangian in Eq. 1, in which we know the actual numerical value of $E_q(g_i)$.

Now return to the consideration of rationality operators in general, not just KL rationality operators. Just as before, we can focus on player i by augmenting our prior knowledge to include the strategies of all the other players. The associated Lagrangian is

$$\mathscr{L}_i(q_i) = \lambda_i [R([g_i]_{i,q}, q_i) - \rho_i] \;-\; S(q_i). \tag{8}$$

(The prior knowledge concerning the strategies of the other players is implicit, arising in the effective cost function.) It is shown in the appendix that the set of all the Lagrangians in Eq. 8 (one for each player) are minimized simultaneously by any distribution q^g obeying the equation

$$\begin{aligned} q^g &= \prod_i \frac{e^{-\rho_i [g_i]_{i,q^g}}}{N(\rho_i [g_i]_{i,q^g})} \\ &= \frac{\prod_i e^{-\rho_i [g_i]_{i,q^g}}}{N(\sum_i \rho_i [g_i]_{i,q^g})} \end{aligned}$$

where the second equation follows trivially from the first. It is also shown in the appendix that there is at least one distribution solving this equation and therefore simultaneously minimizing all the Lagrangians.

Such a q^g obeys all the constraints in the Lagrangian in Eq. 7. This means that there exists a solution to Eq. 7's associated primal problem, i.e, there is at least one feasible point of that problem.

Note that the Lagrangian $\mathscr{L}_i$ of Eq. 8 for player i arises in response to prior knowledge specific to player i. Changing from one player and its Lagrangian to another changes the associated prior knowledge. The same is true for the Lagrangians in Eq. 3.

In contrast, the Lagrangian of Eq. 7 arises for a single centralized body of prior knowledge, namely the set of all players' rationalities. For that single body of knowledge, the equilibrium of the game is the solution to a *single*-objective optimization problem. This contrasts with the conventional formulation of full rationality game theory, where the equilibrium is cast as a solution to a multi-objective optimization problem (one objective per player). Furthermore, as usual, for finite β at least one of the derivatives of the Lagrangian of Eq. 7 is negative infinite at the border of the allowed region of product distributions (i.e., at the border of the Cartesian product of unit simplices). Accordingly, all solutions lie in the interior of that region. This can be a big advantage for finding such solutions numerically, since it allows one to use local descent algorithms.

2.11 Examples of bounded rational equilibria

It can be difficult to write down a set of cost functions and associated rationalities β_i and then solve for the associated bounded rational equilibrium. Starting with expected costs rather than rationalities (so the β_i are not specified up-front but instead are Lagrange parameters that we must solve for) can be even more tedious. However there is a simple alternative way to construct examples of games and their bounded rational equilibria. In this alternative one starts with a particular mixed strategy q and then solves for a game for which q is a bounded rational equilibrium, rather than the other way around.

To illustrate this, consider a 2-person noncooperative single-stage game. Let each player have 3 possible moves. Indicate each players' three possible moves by the numerals 0, 1, and 2. Say the (bounded rational) mixed strategy equilibrium is

$$\begin{aligned} &q_1(0) = 1/2,\ q_1(1) = 1/4,\ q_1(2) = 1/4; \\ &q_2(0) = 2/3,\ q_2(1) = 1/4,\ q_2(2) = 1/12\ . \end{aligned} \tag{9}$$

Now we know that at the equilibrium, $q_1(x_1) \propto e^{-\beta_1 E(g_1 | x_1)}$, where β_1 is player 1's rationality, and g_1 is her cost function (the negative of her utility function). This means for example that

$$\begin{aligned} \exp -(\beta_1[E(g_1 \mid x_1 = 0) - E(g_1 \mid x_1 = 1)]) &= \frac{q_1(0)}{q_1(1)} = 2; \\ \beta_1[E(g_1 \mid x_1 = 0) - E(g_1 \mid x_1 = 1)] &= -\ln(2). \end{aligned} \tag{10}$$

We have a similar equation for the remaining independent difference in expectation values for player 1. The analogous pair of equations for player 2 also hold.

Now define the vectors $\mathbf{g}_{i;j}(.) \triangleq g_i(x_i = j, .)$. So for example $\mathbf{g}_{1;0} = (g_1(x_1 = 0, x_2 = 0), g_1(x_1 = 0, x_2 = 1), g_1(x_1 = 0, x_2 = 2))$. Then we can express our equations compactly as four dot product equalities:

$$\begin{aligned} \beta_1(\mathbf{g}_{1;0} - \mathbf{g}_{1;1}) \cdot q_2 &= -\ln(2), \\ \beta_1(\mathbf{g}_{1;0} - \mathbf{g}_{1;2}) \cdot q_2 &= -\ln(2)\ ; \\ \beta_2(\mathbf{g}_{2;0} - \mathbf{g}_{2;1}) \cdot q_1 &= -\ln(8/3), \\ \beta_2(\mathbf{g}_{2;0} - \mathbf{g}_{2;2}) \cdot q_1 &= -\ln(8)\ . \end{aligned} \tag{11}$$

Note that we can absorb each β_i into its associated g_i; all that matters is their product. We can now plug in for the vectors q_1 and q_2 from Eq. 9 and simply write down solutions for the four three-dimensional vectors $\mathbf{g}_{i,j}$. If desired, we can then evaluate the associated expected values of the cost functions for the two players.

Note that the variables in the first pair of equalities in Eq. 11 are independent of those in the second pair. In other words, whereas the Boltzmann equations giving q for a specified set of g_i are a set of coupled equations, the equations giving the g_i for a specified q are not coupled. Note also that our equations for the $\mathbf{g}_{i;j}$ are (extremely) underconstrained. This illustrates how compressive the mapping from the g_i to the associated equilibrium q is. Bear in mind though that that mapping is also multi-valued in general; in general a single set of cost functions can have more than one equilibrium, just like it can have more than one Nash equilibrium.

The generalization of this example to arbitrary numbers of players with arbitrary move spaces is immediate. As before, indicate the moves of every player by an associated set of integer numerals starting at 0. Let the subscript (i) on a vector indicate all components but the i'th one. Also absorb the rationalities β_i into the associated g_i.

Now specify q and the vectors $g_i(x_i = 0, .)$ (one vector for each i) to be anything whatsoever. Then for all players i, the only associated constraint on the i'th cost function concerns certain projections of the vectors $g_i(x_i > 0, .)$ (one projection for each value x_i):

$$\begin{aligned}\int dx'_{(i)} g_i(x_i, x'_{(i)}) \prod_{j \neq i} q_j(x'_j) \;&=\; -\ln(\frac{q_i(0)}{q_i(x_i)}) \qquad + \\ &\int dx'_{(i)} g_i(0, x'_{(i)}) \prod_{j \neq i} q_j(x'_j) \\ &\qquad\qquad\qquad \forall i, x_i\end{aligned}$$

i.e.,

$$\begin{aligned}\mathbf{g}_i(x_i, .) \cdot q_{(i)} = -\ln(\frac{q_i(0)}{q_i(x_i)}) \quad &+ \quad \mathbf{g}_i(0, .) \cdot q_{(i)} \\ &\forall i, x_i.\end{aligned} \tag{12}$$

2.12 Semi-coordinate systems

Consider a multi-stage game like chess, with the stages (i.e., the instants at which one of the players makes a move) delineated by t. Now strategies are what are set by the players before play starts. So in such a multi-stage game the strategy of player i, x_i, must be the set of t-indexed maps taking what that player has observed in the stages $t' < t$ into its move at stage t. Formally, this set of maps is called player i's **normal form** strategy.

The joint strategy of the two players in chess sets their joint move-sequence, though in general the reverse need not be true. In addition, one can always find a joint strategy to result in any particular joint move-sequence. Typically there is overlap in what the players in chess have observed at stages preceding the current one. This means that even if the players' strategies are statistically independent, their move sequences are statistically coupled. In such a situation, by parameterizing the space of joint-move-sequences z with joint-strategies x, we shift our focus from the coupled distribution $P(z)$ to the decoupled product distribution, $q(x)$. This is the advantage of casting multi-stage games in terms of normal form strategies.

More generally, any onto mapping $\zeta : x \to z$, not necessarily invertible, is called a **semi-coordinate system**. The identity mapping $z \to z$ is a trivial example of a semi-coordinate system. Another example is the mapping from joint-strategies in a multi-stage game to joint move-sequences is an example of a semi-coordinate system. So changing the representation space of a multi-stage game from move-sequences z to strategies x is a semi-coordinate transformation of that game.

We can perform a semi-coordinate transformation even in a single-stage game. Say we restrict attention to distributions over spaces of possible x that are product distributions. Then changing $\zeta(.)$ from the identity map to some other function means that the players are no longer independent. After the transformation their strategy choices — the components of z — are statistically coupled, even though we are considering a product distribution.

Formally, this is expressed via the standard rule for transforming probabilities,

$$P_z(z) \triangleq \zeta(P_x) \triangleq \int dx P_x(x)\delta(z - \zeta(x)), \tag{13}$$

where $\zeta(.)$ is the mapping from x to z, and P_x and P_z are the distributions across x-space and z-space, respectively. To see what this rule means geometrically, let $\mathcal{P}$ be the space of all distributions (product or otherwise) over z's. Recall that $\mathcal{Q}$ is the space of all product distributions over x, and let $\zeta(\mathcal{Q})$ be the image of $\mathcal{Q}$ in $\mathcal{P}$. Then by changing $\zeta(.)$, we change that image; different choices of $\zeta(.)$ will result in different manifolds $\zeta(\mathcal{Q})$.

As an example, say we have two players, with two possible strategies each. So z consists of the possible joint strategies, labeled $(1,1), (1,2), (2,1)$ and $(2,2)$. Have the space of possible x equal the space of possible z, and choose $\zeta(1,1) = (1,1)$, $\zeta(1,2) = (2,2)$, $\zeta(2,1) = (2,1)$, and $\zeta(2,2) = (1,2)$. Say that q is given by $q_1(x_1 = 1) = q_2(x_2 = 1) = 2/3$. Then the distribution over joint-strategies z is $P_z(1,1) = P_x(1,1) = 4/9$, $P_z(2,1) = P_z(2,2) = 2/9$, $P_z(1,2) = 1/9$. So $P_z(z) \neq P_z(z_1)P_z(z_2)$; the strategies of the players are statistically coupled.

Such coupling of the players' strategies can be viewed as a manifestation of sets of potential binding contracts. To illustrate this return to our two player example. Each possible value of a component x_i determines a pair of possible joint strategies. For example, setting $x_1 = 1$ means the possible joint strategies are $(1,1)$ and $(2,2)$. Accordingly such a value of x_i can be viewed as a set of proffered binding contracts. The value of the other components of x determines which contract is accepted; it is the intersection of the proffered contracts offered by all the components of x that determines what single contract is selected. Continuing with our example, given that $x_1 = 1$, whether the joint-strategy is $(1,1)$ or $(2,2)$ (the two options offered by x_1) is determined by the value of x_2.

Binding contracts are a central component of cooperative game theory. In this sense, semi-coordinate transformations can be viewed as a way to convert noncooperative game theory into a form of cooperative game theory.

While the distribution over x uniquely sets the distribution over z, the reverse is not true. However so long as our Lagrangian directly concerns the distribution over x rather than the distribution over z, by minimizing that Lagrangian we set a distribution over z. In this way we can minimize a Lagrangian involving product distributions, even though the associated distribution in the ultimate space of interest is not a product distribution.

The Lagrangian we choose over x should depend on our prior information, as usual. If we want that Lagrangian to include an expected value over z's (e.g., of a cost function), we can directly incorporate that expectation value into the Lagrangian over x's, since expected values in x and z are identical: $\int dz P_z(z)A(z) = \int dx P_x(x)A(\zeta(x))$ for any function $A(z)$. (Indeed, this is the standard justification of the rule for transforming probabilities, Eq. 13.)

However other functionals of probability distributions can differ between the two spaces. This is especially common when $\zeta(.)$ is not invertible, so the space of possible x is larger than the space of possible z. For example, in general the entropy of a $q \in \mathcal{Q}$ will differ from that of its image, $\zeta(q) \in \zeta(\mathcal{Q})$ in such a case. (The prior probability μ in the definition of entropy only gives us invariance when the two spaces have the same cardinality.) A correction factor is necessary to relate the two entropies [46].

In such cases, we have to be careful about which space we use to formulate our Lagrangian. If we use the transformation $\zeta(.)$ as a tool to allow us to analyze bargaining games with binding contracts, then the direct space of interest is actually the x's (that is the place in which the players make their bargaining moves). In such cases it makes sense to apply all the analysis of the preceding sections exactly as it is written, concerning Lagrangians and distributions over x rather than z (so long as we redefine cost functions to implicitly pre-apply the mapping $\zeta(.)$ to their arguments). However if we instead use $\zeta(.)$ simply as a way of establishing statistical dependencies among the strategies of the players, it may make sense to include the entropy correction factor in our x-space Lagrangian.

An important special case is where the following three conditions are met: Each point z is the image under $\zeta(.)$ of the same number of points in x-space, n; $\mu(x)$ is uniform (and therefore so is $\mu(z)$); and the Lagrangian in x-space, $\mathscr{L}_x$, is a sum of expected costs and the entropy. In this situation, consider a z-space Lagrangian, $\mathscr{L}_z$, whose functional dependence on P_z, the distribution over z's, is identical to the dependence of $\mathscr{L}_x$ on P_x, except that the entropy term is divided by n [9]. Now the minimizer $P^*(x)$ of $\mathscr{L}_x$ is a Boltzmann distribution in values of the cost function(s). Accordingly, for any z, $P^*(x)$ is uniform across all n points $x \in \zeta^{-1}(z)$ (all such x have the same cost value(s)). This in turn means that $S(\zeta(P_x)) = nS(P_z)$ So our two Lagrangians give the same solution, i.e., the "correction factor" for the entropy term is just multiplication by n.

2.13 Entropic prior game theory

Finally, it is worth noting that in the real world the information we are provided concerning the system often will not consist of *exact* values of functionals of q, be those values expected costs, rationalities, or what have you. Rather that knowledge will be in the form of data, D, together with an associated likelihood function over the space of q. For example, that knowledge might consist of a bias toward particular rationality values, rather than precisely specified values:

$$P(D \mid q) \propto e^{-\alpha \sum_i [R_{KL}([g_i]_{i,q}) - \rho_i]^2}.$$

where α sets the strength of the bias.

The extension of the minimum information principle to such situations uses the **entropic prior**, $P(q) \propto e^{-\gamma S(q)}$. Bayes' theorem is then invoked to get the posterior distribution [18]:

$$P(q \mid D) \propto e^{-\sum_i \alpha_i [R_{KL}([g_i]_{i,q}) - \rho_i]^2 - \gamma S(q)}.$$

[9]For example, if $\mathscr{L}_x(P_x) = \beta E_{P_x}(G(\zeta(.))) - S(P_x)$, then $\mathscr{L}_z(P_z) = \beta E_{P_z}(G(.)) - S(P_z)/n$, where P_x and P_z are related as in Eq. 13.

The **Bayes optimal** estimate for q, under a quadratic penalty term, is then given by $E(q \mid D)$. The maxent principle for estimating q is given by this estimate under the limit of all α_i going to infinity. For finite α solving for $E(q \mid D)$ can be quite complicated though. For simplicity, such cases are not considered here.

More generally, it may well be that the entropic prior itself is inappropriate in a particular situation, and some other prior should be used. This is standard in Bayesian statistics; how to best infer a distribution q from provided information (or more generally a posterior distribution over such q) depends on the details of the full statistical scenario. See for example [54].

3 PD theory and statistical physics

There are many connections between bounded rational game theory — PD theory — and statistical physics. This should not be too surprising, given that many of the important concepts in bounded rational game theory, like the Boltzmann distribution, the partition function, and free energy, were first explored in statistical physics. This section discusses some of these connections.

3.1 Background on statistical physics

Statistical physics is the physics of systems about which we have incomplete information. An example is knowing only the expected value of a system's energy (i.e., its temperature) rather than the precise value of the energy. The statistical physics of such systems is known as the **canonical ensemble**. Another example is the **grand canonical ensemble** (GCE). There the number of particles of various types in the system is also uncertain. As in the canonical ensemble, in the GCE what knowledge we do have takes the form of expectation values of the quantities about which we are uncertain, i.e., the number of particles of the various types that the system contains, and the energy the system.

Traditionally these kinds of ensembles were analyzed in terms of "baths" of the uncertain variable that are connected to the system. For example, in the canonical ensemble the system is connected to a heat bath. In the GCE the system is also connected to a bath of particles of the various types.

Such analysis showed that for the canonical ensemble the probability of the system being in the particular state x is given by the Boltzmann distribution over the associated value of the system's energy, $G(x)$, with β interpreted as the (inverse) temperature of the system: $p(x) \propto e^{-\beta G(x)}$. This result is independent of the details characteristics of the physical system; all that is important is the **Hamiltonian** $G(x)$, and temperature β.

Note that once one knows $p(x)$ and $G(x)$, one knows the expected energy of the system. It is $G(x)$ that is a fixed property of the system, whereas β can vary. Accordingly, specifying β is exactly equivalent to specifying the expected energy of the system.

In the case of the GCE, x implicitly specifies the number of particles of the various types, as well as their precise state. The analysis for that case showed that $p(x) \propto e^{-\beta G(x) - \sum_i \mu_i n_i}$. In this formula β is again the inverse temperature, n_i is the number of particles of type i, and $\mu_i > 0$ is the **chemical potential** of each particle of type i.

Jaynes was the first to show that these results of conventional statistical physics could be derived without recourse to artificial notions like "baths", simply by using the maxent principle. In particular, he used the exact reasoning in Sec. 2.6 to derive the fact that the canonical ensemble is governed by the Boltzmann distribution.

3.2 Mean field theory and PD theory

In practice it can be quite difficult to evaluate this Boltzmann distribution, due to difficulty in evaluating the partition function. For example, in a **spin glass**, x is an N-dimensional vector of bits, one per particle, and $G(x) = \sum_{i,j} H_{i,j} x_i x_j$. So the partition function is given by $\int dx e^{-\sum_{i,j} H_{i,j} x_i x_j}$, where H is a symmetric real-valued matrix, and as before we use $\int$ to indicate the integral according to the appropriate measure (here a point-sum measure). In general, evaluating this sum for large numbers of spins cannot be done in closed form.

Mean Field (MF) theory is a technique for getting around this problem by approximating the partition function. Intuitively, it works by treating all the particles as independent. It does this by replacing some of the values of the state of a particle in the Hamiltonian by its average state. For example, in the case of the spin glass, one approximates $\sum_{i,j} H_{i,j}[x_i - E(x_i)][x_j - E(x_j)] \approxeq 0$, where the expectation values are evaluated according to the associated exact Boltzmann distribution, i.e., one assumes that fluctuations about the means are relatively negligible. This then means that

$$G(x) \approxeq \sum_{i,j} H_{i,j} 2 x_i E(x_j) \;-\; \sum_{i,j} H_{i,j} E(x_i) E(x_j),$$

The second sum in this approximation cancels out when we evaluate the associated approximate Boltzmann distribution, leaving us with the distribution

$$\begin{aligned} p^{\beta U}(x) \quad &\approxeq \quad P^{\beta U}(x) \triangleq \frac{e^{-\beta \sum_{i,j} H_{i,j} 2 x_i E(x_j)}}{\int dx \; e^{-\beta \sum_{i,j} H_{i,j} 2 x_i E(x_j)}} \\ &= \quad \prod_i \frac{e^{-\alpha_i x_i}}{\int dx_i \; e^{-\alpha_i x_i}}, \end{aligned}$$

where

$$\alpha_i \triangleq 2\beta \sum_j H_{i,j} E(x_j).$$

This approximation $P^{\beta U}(x)$ is far easier to work with than the exact Boltzmann distribution, $p^{\beta U}(x) = \frac{e^{-\beta G(x)}}{N(\beta U)}$, since each term in the product is for a single spin by itself. In particular, if we adopt this approximation we can use numerical techniques to solve the associated set of simultaneous equations

$$E(x_i) = \frac{\partial}{\partial_{\alpha_i}}[\int dx_i \; e^{-\alpha_i x_i}] \quad \forall i$$

for the $E(x_i)$ (so that those $E(x_i)$ are no longer exactly equal to the expected values of the $\{x_i\}$ under the distribution $p^{\beta U}(x)$). Given those $E(x_i)$ values, we can then evaluate the associated approximate Boltzmann distribution explicitly.

The mean field approximation to the Boltzmann distribution is a product distribution, and in fact is identical to the product distribution q^g of bounded rational game theory, for the team game where $g_i(x) = 2\beta G(x) \; \forall i$. Accordingly, the "mean field theory" approximation for an arbitrary Hamiltonian U can be taken to be the associated team game q^g, which is defined for any U.

This bridge between bounded rational game theory and statistical physics means that many of the powerful tools that have been developed in statistical physics can be applied to bounded rational game theory. In particular, much work in statistical physics has been done with approximating distributions that are higher order than products, allowing for coupling between the variables. The associated extension of PD theory is a full-blown theory of Probability Lagrangians.

Finally, this bridge can be used to apply PD theoretic techniques in statistical physics rather than vice-versa. In particular, it is shown elsewhere [20, 21] that if one replaces the identical cost function of each player in a team game with different cost functions, then the bounded rational equilibrium of that game can be numerically found far more quickly. In the context of statistical physics, this means that numerically solving for a MF approximation may be expedited by assigning a different Hamiltonian to each particle.

3.3 Information-theoretic misfit measures

The proper way to approximate a target distribution p with a distribution from a set $\mathcal{C}$ is to first specify a misfit measure saying how well each member of $\mathcal{C}$ approximates p, and then solve for the member with the smallest misfit. This is just as true when $\mathcal{C}$ is the set of all product distributions as when it is any other set.

How best to measure distances between probability distributions is a topic of ongoing controversy and research [27]. The most common way to do so is with the infinite limit log likelihood of data being generated by one distribution but misattributed to have come from the other. This is known as the **Kullback-Leibler distance** [16, 28, 17]:

$$KL(p_1 \,||\, p_2) \triangleq S(p_1 \,||\, p_2) - S(p_1) \tag{14}$$

where $S(p_1 \,||\, p_2) \triangleq -\int dx \; p_1(x) \ln[\frac{p_2(x)}{\mu(x)}]$ is known as the **cross entropy** from p_1 to p_2 (and as usual we implicitly choose uniform μ). The KL distance is always non-negative, and equals zero iff its two arguments are identical. However it it is far from being a metric. In addition to violating the triangle inequality, it is not symmetric under interchange of its arguments, and in numerical applications has a tendency to blow up. (That happens whenever the support of p_1 includes points outside the support of p_2.)

Nonetheless, this is by far the most popular measure. It is illuminating to use it as our misfit measure. As shorthand, define the "pq distance" as $KL(p \mid\mid q)$, and the "qp distance" as $KL(q \mid\mid p)$, where p is our target distribution and q is a product distribution. Then it is straightforward to show that the qp distance from q to target distribution $p^{\beta U}$ is just the maxent Lagrangian, up to irrelevant overall constants. In other words, the q minimizing the maxent Lagrangian — the distribution arising in MF theory — is the q with the minimal qp distance to the associated Boltzmann distribution.[10]

However the qp distance is the (infinite limit of the negative log of the) likelihood that distribution p would attribute to data generated by distribution q. It can be argued that a better measure of how well q approximates p would be based on the likelihood that q attributes to data generated by p. This is the pq distance. Up to an overall additive constant (of the canonical distribution's entropy), the pq distance is

$$KL(p \mid\mid q) = -\sum_i \int dx \; p(x) \ln[q_i(x_i)].$$

This is equivalent to a team game where each coordinate i has the "Lagrangian"

$$L_i^*(q) \quad \triangleq \quad -\int dx_i \; p_i(x_i) \ln[q_i(_i)],$$

where $p_i(x_i)$ is the marginal distribution $\int dx_{(i)} p(x)$.

The minimizer of this is just $q_i = p_i \; \forall i$, i.e., each q_i is set to the associated marginal distribution of p. So in particular, when our target distribution is the canonical ensemble distribution $p^{\beta U}$, the optimal q according to pq distance is the set of marginals of $p^{\beta U}$. Note that unlike the solution for qp distance, here the solution for each q_i is independent of the $q_{(i)}$. So we don't have a game theory scenario; we do not need to pay attention to the $q_{(i)}$ when estimating each separate q_i. Correspondingly, whereas there are many local minima of the team game Lagrangian studied above, $q \in \mathcal{Q} \rightarrow KL(q \mid\mid p^{\beta U})$, there is only one, global minimum of $q \in \mathcal{Q} \rightarrow KL(p^{\beta} \mid\mid q)$.

Another difference between the two kinds of KL distance is how the associated optimal product distributions are typically calculated numerically. The product distribution that optimizes the maxent Lagrangian is usually found via derivative-based traversal of that Lagrangian, or techniques like (mixed) Brouwer updating[20, 21, 24, 22, 42]. In contrast, the integral giving each marginal distribution of p is usually found via adaptive importance sampling

[10] Note that *any* distribution $\pi(x)$ can be written as a Boltzmann distribution, simply by identifying $\beta U(x) = \ln(\pi(x))$; the issues involved in approximating the Boltzmann distribution are generic to the general problem of approximating distributions.

of the associated integral, with the proposal distribution for the integral to approximate p_i set adaptively, as $q_{(i)}$[20].

It is possible to motivate yet other choices for the q that best approximates $p^{\beta U}$. To derive one of them, start with Lemma 1, with $\mathbb{R}^n$ set to the space of real-valued functions over the set of x's (so that n is the number of possible x). Have a single constraint f that restricts us to $\mathcal{P}$, the unit simplex in $\mathbb{R}^n$, i.e., that restricts us to the set of functions that (assuming they are nowhere-negative) are probability distributions. Choose V to be the associated Lagrangian, $\mathscr{L}(p) = \beta E_p(G) - S(p)$, p being a point in our constrained submanifold of $\mathbb{R}^n$. Note that this p can be *any* distribution over the x's, including one that couples the components $\{x_i\}$.

Say we are at some current product distribution q. Then we can apply Lemma 1 with the choices just outlined to tell us what direction to move from q in $\mathcal{P}$ so as to reduce the Lagrangian. In general, taking a step in that direction will result in a distribution p' that is not a product distribution. However we can solve for the product distribution that is closest to that p', and move to that product distribution. By iterating this procedure we can define a search over the submanifold of product distributions. We can then solve for the product distribution at which this search will terminate.

To do this, of course, we must define what we mean by "closest". Say that we choose to measure closeness by pq distance. Then the terminating production distribution is the one for which the marginals of $\nabla L + \lambda \nabla f$ all equal 0. For each i, this means that

$$\int dx_{(i)}[\beta G(x) + \ln(p(x)) + 1 + \lambda] = 0$$

at the equilibrium product distribution p. Writing out $p = \prod_i q_i$ and evaluating gives

$$q_i(x_i) \propto \exp\left(-\beta \frac{\int dx_{(i)} G(x)}{\int dx_{(i)} 1}\right). \tag{15}$$

This is akin to the q^g of a bounded rational game, except that each player/particle i sets its distribution by evaluating conditional expected U with a uniform distribution over the $x_{(i)}$, rather than with $q_{(i)}$.

3.4 Semi-coordinate transformations

Let's say there are numerical difficulties with our finding a q that is local minimization of the maxent Lagrangian. That q might still be a poor fit to $p(x)$ if it is far from the global minimizer of the Lagrangian. Furthermore, even the global minimizer might be a poor fit, if $p(x)$ simply can't be well-approximated by a product distribution.

There are many techniques for improving the fit of a product distribution to a target distribution in machine learning and statistics [28]. To give a simple example, say one wishes to approximate the target distribution in $\mathbb{R}^N$ with a product of Gaussians, one Gaussian for each coordinate. Even if the target distribution is a Gaussian, if it is askew, then one won't be able to do a good job of approximating it with a product of Gaussians. However one can use Principal Components Analysis (PCA) to find how to rotate one's coordinates so that a product of Gaussians fits the target exactly.

Similar techniques can address both the issue of breaking free of local minima of the Lagrangian, and improving the accuracy of the best product distribution approximation to p. More precisely, identify x with the variables z discussed in Sec. 2.12. Then consider changing the map $\zeta(.) : x \rightarrow z$ from the identity map. This will in general change the mapping from P_x to $\mathscr{L}_z(\zeta(P_x))$. So if $\mathscr{L}_z$ is the Lagrangian we are interested in, the mapping from product distributions over x can be changed by changing $\zeta(.)$, in general.

As an example, consider the case where the space of x's is identical to the space of z's, and consider all possible bijective transformations $\zeta(.)$. Entropy is the same in both spaces for any ζ, i.e., $S(P_z) = S(\zeta(P_x)) = S(P_x)$. So for fixed P_x, the entropy in z-space is independent of $\zeta(.)$. However if we fix P_x and change $\zeta(.)$ the expected values of utilities will change. So $\mathscr{L}_z(\zeta(P_x))$ does depend on $\zeta(.)$, as claimed.

This means that by changing $\zeta(.)$ while leaving q_x unchanged, we will in general change whether we are at a local minimum of $\mathscr{L}_z(\zeta(q_x))$. Furthermore, such a change will change how closely the global minimizer of $\mathscr{L}_z(\zeta(q_x))$ approximates any particular target distribution. Indeed, some such transformation will always transform a team game to have a strictly convex maxent Lagrangian, with only one (bounded rational) equilibrium, an equilibrium that is in the interior of the region of allowed q and that has the lowest possible value of the Lagrangian. In the worst case, we can get this behavior by transforming to the semi-coordinate system in which x is one-dimensional, so that any $p(z)$ — coupling its variables or not — can be expressed as a $q(x) = q_1(x_1)$.

Note that unlike with PCA, semi-coordinate transformations can be used for non-Euclidean semi-coordinates (i.e., when neither x's nor z's are Euclidean vectors). They also can be guided by numerous measures of the goodness of fit to the target distribution (e.g., KL distance), in contrast to PCA's restriction to assuming a Gaussian likelihood.

3.5 Bounded rational game theory for variable number of players

The bridge between statistical physics and bounded rational game theory have many uses beyond the practical ones alluded to the previous subsection. In particular, it suggests extending bounded rational game theory to ensembles other than the canonical ensemble. As an example, in the GCE the number of particles of the various allowed types is uncertain and can vary. The bounded rational game theory version of that ensemble is a game in which the number of players of various types can vary.

We can illustrate this by extending a simple instance of evolutionary game theory [6] to incorporate bounded rationality and allow for a finite total number of players. Say we have a finite population of players, each of which has one of m' possible **types**. (Types are sometimes called **feature vectors** in the literature.) Each player i in the population is randomly paired with a different player j, and they each choose a strategy for a two-person game. The set of strategies each of those players can choose among is fixed by its respective attribute vector. The cost player i receives depends on the attribute vectors of itself and of j, in addition to their joint strategy. Finally, to reflect this dependence, we allow each player to vary its strategy depending on the attribute vector of its opponent; we call player i's **meta-strategy** the mapping from its opponent's attribute vector to i's strategy [11].

As an example, in biological contexts a player can be a creature, with the associated type being the creature's physical attributes. Its strategy would then be the behavior it exhibits when interacting with another creature. In general, it might modify that behavior depending on the characteristics of the player it is interacting with; this is reflected in the player's meta-strategy. In non-biological contexts, often one instead interprets the population of (meta)-strategies of all players sharing the same type as random samples of the mixed strategy a single higher-level "player". In this interpretation, in the limit of infinite populations, the relative frequencies of meta-strategies among all players sharing the type T are interpreted as the probabilities comprising a mixed strategy of a single player associated with T.

We encode an instance of this scenario in an x with a countably infinite number of dimensions. $x_{i,0} \triangleq n_i(x)$ specifies the number of players of type i, with $\vec{n}(x)$ being the vector of the number of players of all types. For $1 < j \leq x_{i,0}$, $x_{i,j} \triangleq s_{i,j}(x)$ is the meta-strategy selected by the j'th player of type i. If its opponent is the j'th player of type T', the cost to the i'th player of type T is $g_{T,i,T',j}(x) \triangleq g_{T,i,T',j}(s, s', n_T, n_{T'})$, where s and s' are the two players' respective meta-strategies.[12] To enforce consistency between the index numbers i, j and the associated numbers of players, we set $g_{T,i,T',j}(s, s', \vec{n}) = 0$ if either $i > n_T$ or $j > n_{T'}$.

To start, as an example we parallel the GCE, and presume that for each type we know the expected number of players having that type, and the expected cost averaged over all players having that type. Also stipulate that the distribution over x is a product distribution, q. Then our first class of prior information specifies the values of

$$\sum_{k>0} k \; q_{T,0}(k) = \sum_{x_{T,0}} x_{T,0} \; q_{T,0}(x_{T,0}).$$

Our second class is of the averaged expected costs of each player engaged in a game with a randomly chosen opponent (other than itself). They specify the

[11]Note that it is trivial to replace meta-strategies with strategies throughout the analysis below: simply restrict attention to meta-strategies that do not vary with the opponent's attribute vector.

[12]Note that this broad notation allows the cost function to vary with the individual players, in addition to varying with the players' types and strategies and with the number of players sharing those types.

values of

$$\textstyle\sum_{\vec{n}:n_T>0} q(\vec{n}) \sum_{T':n_{T'}>0} [\frac{n_{T'}-\delta_{T,T'}}{(\sum_{T''} n_{T''})-1}] \sum_{j,k} \int ds_T \, ds_{T'}$$

$$[1-\delta_{T,T'}\delta_{j,k}] \frac{q_{T,j}(s_T) q_{T',k}(s_{T'}) g_{T,j,T',k}(s_T, s_{T'}, \vec{n})}{n_T(n_{T'} - \delta_{T,T'})}$$

respectively, for all types T. (The sums over j and k all implicitly extend from 1 to ∞, and the delta functions are Kronecker deltas that prevent a player from playing itself.) Using the consistency character of g's and dividing, this becomes

$$\textstyle\sum_{x_{1,0}} \ldots \sum_{x_{m',0}} \sum_{T'} \sum_{j,k} \int dx_{T,j} \, dx_{T',k}$$

$$\{[1-\delta_{T,T'}\delta_{j,k}] \, [\prod_{i=1}^{m'} q_{i,0}(x_{i,0})] \quad \times$$

$$\frac{q_{T,j}(x_{T,j}) \; q_{T',k}(x_{T',k}) \; g_{T,j,T',k}(x)}{x_{T,0} \; [\sum_{T''} x_{T'',0} - 1]}\}.$$

We can write these expressions as expectation values, over x, of $2m'$ functions. The first m' of these functions are the functions $n_T(x) = x_{T,0}$ (one function for each T), i.e., the type-counts. Integrating out many components of q, the second m' functions are

$$c_T(x) \triangleq \frac{\sum_{T',j,k}\{[1-\delta_{T,T'}\delta_{j,k}] \; g_{T,j,T',k}(x)\}}{x_{T,0} \; [(\sum_{T''} x_{T'',0}) - 1]}$$

(Roughly speaking, $c_T(x)$ is the average over all players (T, j) of type T, of the average over all players other than (T, j), of the cost to (T, j), as delineated in x.) Accordingly, the maxent principle directs us to minimize the Lagrangian

$$\mathscr{L}(q) = -\sum_T [\mu_T(E(n_T) - N_T) + \beta_T(E(c_T) - C_T)] \; - \; S(q)$$

where the integers $\{N_T\}$ and real numbers $\{C_T\}$ are our prior information. In the usual way, the solution for each pair $(i \in \{1, \ldots, m'\}, j \geq 0)$ is

$$q_{i,j}(x_{i,j}) \propto e^{-E([\sum_{T'} \mu_{T'} n_{T'} - \beta_{T'} c_{T'}] \mid x_{i,j})},$$

where the values of the Lagrange parameters are all set by our prior information.

This distribution is analogous to the one in the GCE. As usual, one can consider variants of it by focusing on one variable at a time, having prior knowledge in the form of rationality values, etc. In addition, even if we stay in this random-2-player games scenario, there is no reason for us to restrict attention to prior information paralleling that of the GCE. As with bounded rational game theory with a fixed number of players, our prior information can concern nonlinear functions of q, couple the cost functions, etc.

In particular, in evolutionary game theory we do not know the expected number of players having each type, nor their average costs. In addition, the equilibrium concept stipulates that all players will have type T if a particular condition holds. That condition is that the addition of a player of type other than T to the population results in an expected cost to that added player that is greater than the associated expected cost to the players having type T. This provides a model of the phenotypic interactions underlying natural selection.

We can encapsulate evolutionary game theory in a Lagrangian by appropriately replacing each pair of GCE-type constraints (one pair for each type) with a single constraint. As an example of how to do this, consider the (single) constraint for type each T that

$$E(\frac{n_T}{\sum_{T'} n_{T'}}) = E([\frac{\max_{T'} c_{T'} - c_T}{\max_{T'}(c_{T'}) - \min_{T'}(c_{T'})}]^\gamma) \tag{16}$$

for some positive real value γ.

For finite γ, the entropy term in the Lagrangian ensures that no $q_{T,0}(x_{T,0})$ equals 0 exactly. So for no T is the expectation value in the lefthand side of this constraint exactly 0, i.e., all n_T will have non-zero expectation. Now consider the limit of infinite γ. In that limit, any distribution q minimizing the Lagrangian is non-infinitesimal only for what are essentially the **evolutionarily stable strategies** (ESS's) of conventional evolutionary game theory. These are the pairs (type T, associated metastrategy distribution $q_{T,j}$) that are best performing, in the sense that no other pair has a lower expected cost function value. (Technically, this is a slight modification of the conventional definition of an ESS.) Since players of the same type are interchangable, an equivalent definition of an ESS is as a type T and an associated "rolled up" probability distribution over meta-strategies s_T, given by the values (one for each s_T)

$$\lim \frac{\sum_i q_{T,i}(x_{T,i})\delta_{x_{T,i},s_T}}{\sum_i q_{T,i}(x_{T,i})} \tag{17}$$

Since it has $\gamma \to \infty$, the ESS gives the condition introduced just above as underling the equilibrium concept of conventional evolutionary game theory. In this, the distribution for finite γ can be viewed as a "bounded rational" extension of conventional evolutionary game theory. In that extension (type, strategy) pairs are allowed even if they don't have the lowest possible cost, so long as their cost is close to the lowest possible. [13] Since ths extension gives you an entire distribution over the $\{n_T\}$, it also provides a way to deal with stocfhastic fluctuations in numbers of (type, meta-strategy) pairs.

There is always a solution to this Lagrangian (unlike the case in conventional full rationality evolutionary game theory). The technique of Lagrange parameters provides that solution for each pair ($i \in \{1, \ldots, m'\}, j \geq 0$) in the

[13] Many other parameterized constraints will result in this kind of relation between the parameter value and the resultant Lagrangian-minimizing distribution. The one in Eq. 16 was chosen simply for pedagogical clarity.

usual way:

$$q_{i,j}(x_{i,j}) \propto e^{-E(\sum_{T'} \alpha_{T'} f_{T'}(x) \mid x_{i,j})}$$

where the Lagrange parameters enforce our constraint, and

$$f_{T'}(x) \triangleq \frac{n_{T'}}{\sum_{T''} n_{T''}} - [\frac{\max_{T''} c_{T''} - c_{T''}}{\max_{T''}(c_{T''}) - \min_{T''}(c_{T''})}]^{\gamma}.$$

More general forms of conventional evolutionary game theory allow games with more than two players, and localization via network structures delineating how players are likely to be grouped to play a game. Other elaborations have each player not know the exact attribute vectors of all its opponents, but only an "information structure" providing some information about those opponents' attribute vectors. All such extensions can be straightforwardly incorporated into the current bounded rational analysis. Many other extensions are simple to make as well. For example, since the cost functions have all components of $\vec{n}$ in their argument lists, they can depend on the total size of the population. This allows us to model the effect on population size of finite environmental resources.

It is straightforward to extend all of the foregoing to have games involving more than two players at once. Note that if we change how we encode the number of players of the various types and their joint meta-strategy in x, we change the form of the expectations in Eq. 16. This reflects the fact that by changing the encoding we change the implication of using a product distribution. Formally, such a change in the encoding is a change in the semi-coordinate system. See Sec. 2.12.

4 Appendix

This appendix provides proofs absent from the main text.

4.1 $\beta \sum_i [E_p(g_i)]^2 - S(p)$ is convex over the unit simplex

Proof: Since $S(p)$ is concave over the unit simplex, and the unit simplex is a hyperplane, it suffices to prove that $\sum_i [E_p(g_i)]^2$ is convex over all of Euclidean space. Since a weighted average of convex functions is convex, we only need to prove that any single function of the form $[\int dx\ p(x)f(x)]^2$ is convex. The Hessian of this function is $2f(x)f(x')$. Rotate coordinates so that f is a basis vector, i.e., so that f is proportional to a delta function. This doesn't change the eigenvalues of the Hessian. After this change though, the Hessian is diagonal, with one non-zero entry on the diagonal, which is non-negative. So its eigenvalues are zero and a non-negative number. **QED.**

4.2 R_{KL} is a rationality operator

Proof: Since KL distance only equals 0 when its arguments match and is never negative, requirement (1) of rationality operators holds for R_{KL}. Next, since $R_{KL} = \text{argmin}_\beta[\beta \int dy\ p(y)U(y) + \ln(N(\beta U))]$, we know that $E_p(U) = -\frac{1}{N(\beta U)} \frac{\partial N(\beta U)}{\partial \beta}|_{\beta = R_{KL}(U,p)}$. Accordingly, all p with the same rationality have the same expected value $E_p(U)$. Using the technique of Lagrange parameters then readily establishes that of those distributions having the same expected U, the one with maximal entropy is a Boltzmann distribution. Furthermore, by requirement (1), we know that for a Boltzmann distribution the exponent β must equal the rationality of that distribution. **QED.**

4.3 Alternative form of a constraint on R_{KL}

Proof: Let $f\{\alpha, v\}$ be any function that is monotonically decreasing in its (real-valued) first argument. Then any constraint $R([g_i]_{i,q}, q_i) - \rho_i = 0$ is satisfied iff the constraint $f\{R([g_i]_{i,q}, q_i), q_{(i)}\} - f\{\rho_i, q_{(i)}\} = 0$ is satisfied. Choose

$$\begin{aligned} f\{\alpha, q_{(i)}\} &= -\frac{\partial \ln(N(\beta[g_i]_{i,q}))}{\partial \beta}|_{\beta=\alpha} \\ &= \frac{\int dx_i [g_i]_{i,q} e^{-\alpha [g_i]_{i,q}(x_i)}}{N(\alpha [g_i]_{i,q})}. \end{aligned}$$

Differentiating this quantity with respect to α gives the negative of the variance of $[g_i]_{i,q}$ under the Boltzmann distribution $\frac{e^{-\alpha[g_i]_{i,q}}}{N(\alpha[g_i]_{i,q})}$. Since variances are non-negative, this derivative is non-positive, which establishes that f is monotonically decreasing in its first argument.

Evaluating,

$$f\{\rho_i, q_{(i)}\} = \int dx\ g_i(x) \frac{e^{-\rho_i E(g_i | x_i)}}{N(\rho_i g_i)} q_{(i)}(x_{(i)}).$$

In addition, from the equation defining R_{KL}, we know that

$$-\frac{\ln(N(\beta U(x_i)))}{\partial \beta}|_{\beta = R_{KL}(U,q_i)} = \int dx_i q_i(x_i) U(x_i)$$

for any function U. Plugging in $U = [g_i]_{i,q}$, we see that

$$\begin{aligned} f\{R([g_i]_{i,q}, q_i), q_{(i)}) &= \int dx_i q_i(x_i) [g_i]_{i,q}(x_i) \\ &= E_q(g_i).\textbf{QED}. \end{aligned}$$

4.4 q^g minimizes the Lagrangians of Eq. 8

Proof: Following Nash, we can use Brouwer's fixed point theorem to establish that for any non-negative $\{\rho_i\}$, there must exist at least one product distribution solving the equation for q^g. The constraint terms in all the $\mathscr{L}_i$ of Eq. 8 are zero for this distribution. By requirement (2), we also know that given $q^g_{(i)}$ (and therefore $[g_i]_{i,q^g}$), there is no q_i with rationality ρ_i that has lower entropy than q^g_i. Accordingly, no q_i will have a lower value of $\mathscr{L}_i$ than $q_{(i)}$. Since this holds for all i, q^g minimizes all the Lagrangians in Eq. 8 simultaneously. **QED.**

4.5 Derivation of Lemma 1

Proof: Consider the set of $\vec{u}$ such that the directional derivatives $D_{\vec{u}} f_i$ evaluated at x' all equal 0. These are the directions consistent with our constraints to first order. We need to find the one of those $\vec{u}$ such that $D_{\vec{u}} g$ evaluated at x' is maximal.

To simplify the analysis we introduce the constraint that $|\vec{u}| = 1$. This means that the directional derivative $D_{\vec{u}} V$ for any function V is just $\vec{u} \cdot \nabla V$. We then use Lagrange parameters to solve our problem. Our constraints on $\vec{u}$ are $\sum_j u_j^2 = 1$ and $D_{\vec{u}} f_i(x') = \vec{u} \cdot \nabla f_i(x') = 0 \;\; \forall i$. Our objective function is $D_{\vec{u}} V(x') = \vec{u} \cdot \nabla V(x')$.

Differentiating the Lagrangian gives

$$2\lambda_0 u_i + \sum_i \lambda_i \nabla f = \nabla V \;\; \forall i.$$

with solution

$$u_i = \frac{\nabla V - \sum_i \lambda_i \nabla f}{2\lambda_0}.$$

λ_0 enforces our constraint on $|\vec{u}|$. Since we are only interested in specifying $\vec{u}$ up to a proportionality constant, we can set $2\lambda_0 = 1$. Redefining the Lagrange parameters by multiplying them by -1 then gives the result claimed. **QED.**

4.6 Proof of claims following Lemma 1

i) Define $f_i(q) \triangleq \int dx_i q_i(x_i)$, i.e., f_i is the constraint forcing q_i to be normalized. Now for any q that equals zero for some joint move there must be an i and an x'_i such that $q_i(x'_i) = 0$. Plugging into Lemma 1, we can evaluate the component of the direction of steepest descent along the direction of player i's probability of making move x'_i:

$$\frac{\partial \mathscr{L}_i}{\partial q_i(x_i)} + \lambda \frac{\partial f_i}{\partial q_i(x_i)} =$$
$$\beta E(g_i \mid x_i) + \ln(q_i(x_i)) - \frac{\int dx''_i [\beta E(g_i \mid x''_i) + \ln(q_i(x''_i))]}{\int dx''_i 1}$$

Since there must some x''_i such tha $q_i(x''_i) \neq 0$, $\exists x_i$ such that $\beta E(g_i \mid x''_i) + \ln(q_i(x''_i))$ is finite. Therefore our component is negative infinite. So $\mathscr{L}_i$ can be reduced by increasing $q_i(x'_i)$. Accordingly, no q having zero probability for some joint move x can be a minimum of i's Lagrangian.

ii) As a simple example of multiple minima for the team game, consider a problem where all $X_i = \mathbb{B}$. Say that $G(x)$ has identically-valued local minimum at $x = (1, 1, \ldots)$ and $x = (2, 2, \ldots)$, and that $G(x)$ for all other x is the same (large) value. Consider the limit where all $\beta_i \to \infty$. Then one q that minimizes $\mathscr{L}$ has all q_i being delta functions about $x = 1$, and another local minimum has them all being a delta function about $x = 2$. Consider changing all q_i infinitesimally away from the delta function giving the first of these local minima. Doing this will introduce extra probability at *all* vertices of the hypercube, not just at $x = (2, 2, \ldots)$. Accordingly, such an infinitesimal change to q raises $E(G)$ rather than lowers it; one cannot make an infinitesimal change to this q that will lower the Lagrangian. The same is true for the other local minimum of the Lagrangian.

To construct a bounded rational game with multiple equilibria, note that at any (necessarily interior) local minimum q, for each i,

$$\beta E(g_i \mid x_i) + \ln(q_i(x_i)) = \\ \beta \int dx_{(i)} g_i(x_i, x_{(i)}) \prod_{j \neq i} q_j(x_j) + \ln(q_i(x_i))$$

must be independent of x_i, by Lemma 1. So say there is a component-by-component bijection $T(x) \triangleq (T_1(x_1), T_2(x_2), \ldots)$ that leaves all the $\{g_j\}$ unchanged, i.e., such that $g_j(x) = g_j(T(x))\ \forall x, j$ [14].

[14] As an example, consider a congestion team game in which all players have the same set of possible moves, G being a function only of the bit string indexed by $k \in \mathbb{N}$, $\{N(x, k)\}$, where $N(x, k) = 1$ iff there is a move that is shared by exactly k of the players when the joint move is x. In this case T just permutes the set of possible moves in the same way for all players.

Define q' by $q'(x) = q(T(x))\ \forall x$. Then for any two values x_i^1 and x_i^2,

$$
\begin{aligned}
&\beta E_{q'}(g_i \mid x_i^1) + \ln(q'_i(x_i^1)) \\
&\qquad - \beta E_{q'}(g_i \mid x_i^2) \ + \ \ln(q'_i(x_i^2)) \\
&\qquad\qquad = \\
&\beta \int dx_{(i)} g_i(x_i^1, x_{(i)}) \prod_{j \neq i} q_j(T(x_j)) \ + \ \ln(q_i(T(x_i^1))) \\
&\quad - \beta \int dx_{(i)} g_i(x_i^2, x_{(i)}) \prod_{j \neq i} q_j(T(x_j))) \ + \ \ln(q_i(T(x_i^2))) \\
&\qquad\qquad = \\
&\beta \int dx_{(i)} g_i(x_i^1, T^{-1}(x_{(i)})) \prod_{j \neq i} q_j(x_j) \ + \ \ln(q_i(T(x_i^1))) \\
&\quad - \beta \int dx_{(i)} g_i(x_i^2, T^{-1}(x_{(i)})) \prod_{j \neq i} q_j(x_j)) \ + \ \ln(q_i(T(x_i^2))) \\
&\qquad\qquad = \\
&\beta \int dx_{(i)} g_i(T(x_i^1), x_{(i)})) \prod_{j \neq i} q_j(x_j) \ + \ \ln(q_i(T(x_i^1))) \\
&\quad - \beta \int dx_{(i)} g_i(T(x_i^2), x_{(i)})) \prod_{j \neq i} q_j(x_j)) \ + \ \ln(q_i(T(x_i^2))) \\
&\qquad\qquad = \\
&\beta E_q(g_i \mid T(x_i^1)) \ + \ \ln(q_i(T(x_i^1))) \\
&\qquad - \beta E_q(g_i \mid T(x_i^2)) \ + \ \ln(q_i(T(x_i^2)))
\end{aligned}
$$

where the invariance of g_i was used in the penultimate step. Since q is a local minimum though, this last difference must equal 0. Therefore q' is also a local minimum.

Now choose the game so that $\forall i, x_i, T(x_i) \neq x_i$. (Our congestion game example has this property.) Then the only way the transformation $q \rightarrow q(T)$ can avoiding producing a new product distribution is if $q_i(x_i) = q_i(x'_i)\ \forall i, x_i, x'_i$, i.e., q is uniform. Say the Hessians of the players' Lagrangians are not all positive definite at the uniform q. (For example have our congestion game be biased away from uniform multiplicities.) Then that q is not a local minimum of the Lagrangians. Therefore at a local minimum, $q \neq q(T)$. Accordingly, q and $q(T)$ are two distinct equilibria.

iii) To establish that at any q there is always a direction along which any player's Lagrangian is locally convex, fix all but two of the $\{q_i\}$, q_0 and q_1, and fix both q_0 and q_1 for all but two of their respective possible values, which we can write as $q_0(0), q_0(1), q_1(0)$, and $q_1(1)$, respectively. So we can parameterize the set of q we're considering by two real numbers, $x \triangleq q_0(0)$ and $y \triangleq q_1(0)$. The 2×2

Hessian of $\mathscr{L}_i$ as a function of x and y has the entries

$$\begin{array}{cc} \frac{1}{x} + \frac{1}{a-x} & \alpha \\ \alpha & \frac{1}{y} + \frac{1}{b-y} \end{array}$$

where $a \triangleq 1 - q_0(0) - q_0(1)$ and $b \triangleq 1 - q_1(0) - q_1(1)$, and α is a function of g_i and $\prod_{j \neq 0,1} q_j$. Defining $s \triangleq \frac{1}{x} + \frac{1}{a-x}$ and $t \triangleq \frac{1}{y} + \frac{1}{b-y}$, the eigenvalues of that Hessian are

$$\frac{s + t \pm \sqrt{4\alpha^2 + (s-t)^2}}{2}.$$

The eigenvalue for the positive root is necessarily positive. Therefore along the corresponding eigenvector, $\mathscr{L}_i$ is convex at q. **QED.**

iv) There are several ways to show that the value of $E_{q_i^B}([g_i]_{i,q_{(i)}})$ must shrink as β_i grows. Here we do so by evaluating the associated derivative with respect to β_i.

Define $N(U) \triangleq \int dy \; e^{-U(y)}$, the normalization constant for the distribution proportional to $e^{-U(y)}$. View the x_i-indexed vector q_i^B as a function of β_i, g_i and $q_{(i)}$. So we can somewhat inelegantly write $E(g_i) = E_{q_i^B(\beta_i, g_i, q_{(i)}), q_{(i)}}(g_i)$. Then one can expand

$$\begin{aligned} \frac{\partial E(g_i)}{\partial \beta_i} &= -\frac{\partial^2 \ln(N(\beta_i [g_i]_{i,q^B_{(i)}}))}{\partial \beta_i^2} \\ &= -\mathrm{Var}([g_i]_{i,q^B_{(i)}}) \end{aligned}$$

where the variance is over possible x_i, sampled according to $q_i^B(x_i)$. **QED.**

Acknowledgements: I would like to thank Stefan Bieniawski, Bill Macready, Stephane Airiau, Chiu Fan Lee, George Judge, Chris Henze, and Ilan Kroo for helpful discussion.

References

[1] D. Challet and N. F. Johnson, "Optimal combinations of imperfect objects," *Phys. Rev. Let.*, vol. 89, p. 028701, 2002.

[2] G. Korniss and M. N. et al., "Suppressing roughness of virtual times in parallel discrete-event simulations," *Science*, vol. 299, p. 677, 2003.

[3] D. Fudenberg and J. Tirole, *Game Theory.* Cambridge, MA: MIT Press, 1991.

[4] T. Basar and G. Olsder, *Dynamic Noncooperative Game Theory.* Philadelphia, PA: Siam, 1999, second Edition.

[5] M. Osborne and A. Rubenstein, *A Course in Game Theory.* Cambridge, MA: MIT Press, 1994.

[6] R. Aumann and S. Hart, *Handbook of Game Theory with Economic Applications.* North-Holland Press, 1992.

[7] R. Axelrod, *The Evolution of Cooperation.* NY: Basic Books, 1984.

[8] T. Sandholm and V. R. Lesser, "Coalitions among computationally bounded agents," *Artificial Intelligence*, vol. 94, pp. 99–137, 1997.

[9] A. Neyman, "Bounded complexity justifies cooperation in the finitely repeated prisoner's dilemma," *Economics Letters*, vol. 19, pp. 227–230, 1985.

[10] D. Fudenberg and D. K. Levine, *The Theory of Learning in Games.* Cambridge, MA: MIT Press, 1998.

[11] W. B. Arthur, "Complexity in economic theory: Inductive reasoning and bounded rationality," *The American Economic Review*, vol. 84, no. 2, pp. 406–411, May 1994.

[12] C. Boutilier, Y. Shoham, and M. P. Wellman, "Editorial: Economic principles of multi-agent systems," *Artificial Intelligence Journal*, vol. 94, pp. 1–6, 1997.

[13] N. I. Al-Najjar and R. Smorodinsky, "Large nonanonymous repeated games," *Game and Economic Behavior*, vol. 37, no. 26-39, 2001.

[14] A. Tversky and D. Kahneman, "Advances in prospect theory: Cumulative representation of uncertainty," *Journal of Risk and Uncertainty*, vol. 5, pp. 297–323, 1992.

[15] D. Kahneman, "A psychological perspective on economics," *American Economic Review (Proceedings)*, vol. 93:2, pp. 162–168, 2003.

[16] T. Cover and J. Thomas, *Elements of Information Theory.* New York: Wiley-Interscience, 1991.

[17] D. Mackay, *Information theory, inference, and learning algorithms.* Cambridge University Press, 2003.

[18] E. T. Jaynes and G. L. Bretthorst, *Probability Theory : The Logic of Science.* Cambridge University Press, 2003.

[19] D. H. Wolpert, "Factoring a canonical ensemble," 2003, preprint cond-mat/0307630.

[20] ——, "Generalizing mean field theory for distributed optimization and control," 2004, submitted.

[21] W. Macready, S. Bieniawski, and D. Wolpert, "Adaptive multi-agent systems for constrained optimization," 2004, technical report IC-04-123.

[22] C. F. Lee and D. H. Wolpert, "Product distribution theory and semi-coordinate transformations," 2004, submitted to AAMAS 04.

[23] S. Bieniawski and D. H. Wolpert, "Adaptive, distributed control of constrained multi-agent systems," in *Proceedings of AAMAS 04*, 2004.

[24] S. Bieniawski, D. H. Wolpert, and I. Kroo, "Discrete, continuous, and constrained optimization using collectives," in *Proceedings of 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, Albany, New York*, 2004, in press.

[25] E. T. Jaynes, "Information theory and statistical mechanics," *Physical Review*, vol. 106, p. 620, 1957.

[26] M. Opper and D. Saad, *Advanced Mean Field Methods: Theory and Practice (Neural Information Processing).* MIT Press, 2001.

[27] D. H. Wolpert and W. Macready, "Self-dissimilarity as a high dimensional complexity measure," in *Proceedings of the International Conference on Complex Systems, 2004*, 2004, in press.

[28] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2nd ed.).* Wiley and Sons, 2000.

[29] M. Mezard and G. Parisi, "The cavity method at zero temperature," 2002, cond-mat/0207121.

[30] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction.* Cambridge, MA: MIT Press, 1998.

[31] L. P. Kaelbing, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *Journal of Artificial Intelligence Research*, vol. 4, pp. 237–285, 1996.

[32] R. H. Crites and A. G. Barto, "Improving elevator performance using reinforcement learning," in *Advances in Neural Information Processing Systems - 8*, D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, Eds. MIT Press, 1996, pp. 1017–1023.

[33] J. Hu and M. P. Wellman, "Multiagent reinforcement learning: Theoretical framework and an algorithm," in *Proceedings of the Fifteenth International Conference on Machine Learning*, June 1998, pp. 242–250.

[34] D. H. Wolpert, K. Tumer, and J. Frank, "Using collective intelligence to route internet traffic," in *Advances in Neural Information Processing Systems - 11.* MIT Press, 1999, pp. 952–958.

[35] D. H. Wolpert and K. Tumer, "Optimal payoff functions for members of collectives," *Advances in Complex Systems*, vol. 4, no. 2/3, pp. 265–279, 2001.

[36] ——, "Collective intelligence, data routing and braess' paradox," *Journal of Artificial Intelligence Research*, 2002, to appear.

[37] D. H. Wolpert, "Theory of collective intelligence," in *Collectives and the Design of Complex Systems*, K. Tumer and D. H. Wolpert, Eds. New York: Springer, 2003.

[38] C. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, no. 3/4, pp. 279–292, 1992.

[39] J. Shamma and G. Arslan, "Dynamic fictitious play, dynamic gradient play, and distributed convergence to nash equilibria," 2004, submitted.

[40] D. Fudenberg and D. Kreps, "Learning mixed equilibria," *Game and Economic Behavior*, vol. 5, pp. 320–367, 1993.

[41] S. Durlauf, "How can statistical mechanics contribute to social science?" *Proc. Natl. Acad. Sci. USA*, vol. 96, pp. 10 582–10 584, 1999.

[42] S. Airiau and D. H. Wolpert, "Product distribution theory and semi-coordinate transformations," 2004, submitted to AAMAS 04.

[43] D. H. Wolpert and S. Bieniawski, "Distributed control by lagrangian steepest descent," in *Proceedings of CDC 04*, 2004.

[44] N. Antoine, S. Bieniawski, I. Kroo, and D. H. Wolpert, "Fleet assignment using collective intelligence," in *Proceedings of 42nd Aerospace Sciences Meeting*, 2004, aIAA-2004-0622.

[45] D. H. Wolpert and C. F. Lee, "Adaptive metropolis hastings sampling using product distributions," in *Proceedings of ICCS 04*, 2004.

[46] D. H. Wolpert, "What information theory says about best response, binding contracts, and collective intelligence," in *Proceedings of WEHIA04*, A. N. et al, Ed. Springer Verlag, 2004.

[47] D. H. Wolpert and S. Bieniawski, "Adaptive distributed control: beyond single-instant categorical variables," in *Proceedings of MSRAS04*, A. S. et al, Ed. Springer Verlag, 2004.

[48] A. Greif, "Economic history and game theory: A survey," in *Handbook of Game Theory with Economic Applications*, R. J. Aumann and S. Hart, Eds. Amsterdam: North Holland, 1999, vol. 3.

[49] M. Mesterton-Gibbons and E. S. Adams, "Animal contests as evolutionary games," *American Scientist*, vol. 86, pp. 334–341, 1998.

[50] G. Zlotkin and J. S. Rosenschein, "Coalition, cryptography, and stability: Mechanisms for coalition formation in task oriented domains," 1999, (preprint).

[51] N. Nisan and A. Ronen, "Algorithmic mechanism design," *Games and Economic Behavior*, vol. 35, pp. 166–196, 2001.

[52] D. Wolpert and K. Tumer, "Beyond mechanism design," in *International Congress of Mathematicians 2002 Proceedings*, H. G. et al., Ed. Qingdao Publishing, 2002.

[53] R. J. Aumann, "Correlated equilibrium as an expression of Bayesian rationality," *Econometrica*, vol. 55, no. 1, pp. 1–18, 1987.

[54] J. Bernardo and A. Smith, *Bayesian Theory.* Wiley and Sons, 2000.