ARE NASH EQUILIBRIA RATIONAL?

AYDIN MOHSENI, SIMON HUTTEGGER

Abstract. A long-standing open question in the theory of games is the extent and condition under which we may expect rational agents to play the Nash equilibria of games. Results from Kalai and Lehrer (1993) show that Bayesian agents whose priors satisfy the 'absolute continuity' condition invariably learn to play a Nash equilibrium of a game. In response to this, Foster and Young (2001) provide an impossibility theorem demonstrating the 'non-robustness' of Bayesian learning. They show that, in near-zero-sum games of uncertain information, agents cannot be simultaneously rational and also learn to accurately predict the strategies of other players. These results may appear contradictory. They are not. However, their implications bare teasing out. We explain the relation between the two results, and demonstrate that Foster & Young's result is not generic—it is not robust under minute perturbations of the assumptions regarding perfect implementations of agents' plans—and thus fully rational but less-than-perfect Bayesian agents may once again learn to play Nash.

1. Introduction

The Nash equilibrium concept has come to be central to game theory, economics, and other social sciences. A perennial open question is the extent and condition under which we can expect rational agents to play according to the Nash equilibria of games. Rational agents knowing the full structure of the interaction of which they are a part, including their own and others' strategies, outcomes, preferences, and the mutual awareness of all agents—that is, satisfying the common knowledge assumption—will play one of the Nash equilibria of a game.[1] But relaxing any of these requirements undermines the guarantee of Nash equilibrium play in one-shot games. Given this, there has been an incentive to look for extensions of the Nash concept to weaker assumptions and broader classes of strategic interactions. Indeed, this has been a central driver in the formulation of epistemic and evolutionary theories of games. Nearer to the origins of rational choice game theory, justification of the rationality of Nash equilibria has been sought in the context of learning in repeated games—where agents may learn through repeated interaction to converge on an equilibrium of the game. In fact, Nash himself

Date: January 23, 2025
E-mail: amohseni@uci.edu
UC, Irvine, Irvine, CA 92697, USA .
[1]Though which Nash equilibrium will be played, when there is more than one, is its own question, involving further considerations, cf. (1988).

seems to indicate that he had something akin to this sort of dynamical picture in mind when formulating his equilibrium concept (Nash 1951).

It is in this last tradition that, in "Rational Agents Learn to Play Nash Equilibrium" (1993), Kalai & Lehrer demonstrate that rational agents who update via Bayesian conditionalization and whose priors satisfy the 'absolute continuity' condition can learn to play the Nash equilibria of a broad class of games of incomplete information. Contra this result, in "The Impossibility of Prediction the Behavior of Rational Agents" (2001), Foster and Young provide an impossibility theorem that they argue demonstrates the non-robustness of Bayesian learning in justifying Nash play. Foster & Young show that, in games containing near-zero-sum subgames, agents cannot be both rational, in the sense of expected utility maximization, and also learn to accurately predict the strategies of other players, and so cannot play the Nash equilibria of such games.

These two results by Kalai & Lehrer and Foster & Young appear contradictory. They are not. But their implications bare teasing out. In particular, Young argues they have demonstrated that "Bayesian rational learning does not lead to Nash equilibrium in a robust sense (Young 2011, 106)." Leading to Nash in a robust sense would be to converge, a al Kalai & Lehrer, to the Nash of the game. They demonstrate a class of games—near-zero-sum games of uncertain information—where priors satisfying the absolute continuity condition cannot exist. In such games, the demands of rationality and of the accurate prediction of opponents' play required for Nash seem to pull apart.

We demonstrate that Foster & Young's result is not generic—in particular, it is not robust under minute perturbations of the assumptions regarding perfection implementations of agents' plan. Thus, fully rational but less-than-perfect Bayesian agents may once again learn to play Nash. Our conclusion is that Bayesian learning provides a more robust justification of the rationality of Nash than Foster & Young's result might suggest.

In §2, we present Kalai & Leher's proof of Nash play for Bayesian rational agents, and Foster & Young's impossibility theorem for rationality and prediction in near-zero-sum games. In §3, we introduce the technical preliminaries needed to understand our results. In §4, we introduce the formal notions of rationality, prediction, and the standards of convergence employed in the two results. §4, we presently our own computational and analytic results drawing demonstrating how, even in near-zero-sum games of uncertain information, agents may reconcile rationality and prediction, and learn the Nash equilibria of games. In §5, we discuss the implications of these results, and conclude.

2. Kalai & Lehrer and Foster & Young's Results

2.1. **Rational Bayesian Agents Learn to Play Nash.** In "Rational Agents Learn to Play Nash Equilibrium" (1993), Kalai & Lehrer demonstrate that, in infinitely repeated games of uncertain information, if agents are rational, know their own payoffs,

have beliefs regarding opponents' strategies that are "compatible with the truly chosen strategies," and update via Bayesian conditionalization, then they must eventually play a Nash equilibrium of the game as if they had complete information. Moreover, agents' posterior beliefs regarding future play of the game will become accurate with time. In other words, individuals will learn to predict the future play of the game, and to play arbitrarily close to a Nash equilibrium that game.

Kalai & Lehrer's result is notable in that it does not require players to be certain about either the strategies or payoffs of their opponents. These assumptions are replaced by a weaker one: that individuals' prior beliefs satisfy a "grain of truth" condition. The grain of truth condition requires that players' subjective prior beliefs do not assign zero probability to events that can occur in the play of the game. This means that, on future play paths of the game, the probability distribution induced by the possible strategies of all players must be *absolutely continuous*[2] with respect to the probability distributions induced by the private beliefs of the players. That is, any positive probability set of paths must be assigned some positive probability by each player.

Under these condition, Kalai & Lehrer demonstrate that players will converge to play arbitrarily close the Nash equilibrium of the game. This result demonstrates how Bayesian learning can support the Nash concept—in cases of imperfect information, individuals may not be able to play Nash in one-shot interactions, but they will learn to do so in repeated games, as long as their initial beliefs are compatible with possible realizations of future play and they learn via Bayesian conditionalization.

2.2. **An Impossibility Theorem.** In "The Impossibility of Predicting the Behavior of Rational Agents" (2001), Foster & Young provide an impossibility theorem demonstrating that, in infinitely repeated games of incomplete information which contain near-zero-sum subgames, individuals cannot be both rational and also learn to predict the strategies of their opponents, and hence they cannot play Nash.

At face value, this result seems to directly contract Kalai & Lehrer (1993). It does not. The reason, as we will see, is that the two results employ different standards of convergence to play Nash. Kalai & Lehrer use a weaker standard, which requires, roughly, that the behavior of individuals approach Nash. Foster & Young use a stronger standard, which requires that not only the behavior of individuals approach, but also their predictions as to one another's strategies be accurate.

The difference between the two results reflects the fact that optimal strategic behavior can decouple from accurate prediction. Individuals can learn to best respond to one another's distribution of play without learning one another's next-period strategies. This fact has been evinced in other results, as in the case of fictitious play.[3]

[2]text

[3]Fictitious play can itself be thought of as a naive form of Bayesian learning, one where each individual (wrongly) assumes that her opponent is playing a fixed mixed strategy about which she has a diffuse prior, and so she best-responds to the empirical distribution of her opponent's past play (Young 2011).

3. Notation and Technical Preliminaries

To analyze our results, we introduce a framework for learning in indefinitely repeated games, following Young (2011). Consider an n-person game G with a finite strategy space $X = \prod X_i$ and utility functions $u_i : X \to \mathbb{R}$. Payoffs take the form $u_i(x) = u_i^0(x) + \omega_i(x)$, where the $u_i^0(x)$ are the payoffs from the initial benchmark game, and the $\omega_i(x)$ are the i.i.d. perturbations drawn from a continuous random variable $\nu(\omega)$, with support in the interval $[-\lambda/2, \lambda/2]$ where $\lambda > 0$ is the uncertainty parameter. The distribution is common knowledge, but each player i knows only her own payoff realization $u_i(x)$. Perturbations are drawn just once before play, and the game is played indefinitely.

Players act once, simultaneously, in each period $t = 1, 2, ...$ The *outcome* in period t is an n-tuple of actions $x^t \in X$, where x_i^t is the action taken by agent i in period t. A *state* of the process at time t is the history of play up to t, that is, the sequence of outcomes $h = (x^1, x^2, ..., x^t)$. H^t denotes the set of histories of length t, and $H = \cup_t H^t$ the set of all finite histories. A realization of the process is denoted by h. Histories are observed publicly.

The discounted payoff to player i from a realization of play $h = (x^1, x^2, ..., x^t, ...)$ is

$$U_i(h) = (1 - \delta_i) \sum_{t=1}^{\infty} \delta_i^{t-1} u_i(x^t),$$

where $\delta_i \in (0, 1)$ denotes i's discount factor.[4] Δ_i denote the set of mixed strategies over X_i, $\Delta = \prod \Delta_i$ denote the product set of mixed strategies of all players, and $\Delta_{-i} = \prod_{j \neq i} \Delta_j$ denote the product set of mixed strategies of player i's opponents.

A behavioral strategy for player i specifies a conditional probability distribution over i's actions in each period, conditional on the state in the previous period. Represent player i's *behavioral strategy* by a function $q_i^t = g_i(h^{t-1}) \in \Delta_i$, where $q_i^t(x_i)$ is the probability that player i plays x_i, and h^{t-1} is the history of play up to period $t - 1$.

A forecast for player i specifies a conditional probability distribution over player i's opponents' actions in each period, conditional on the state in the previous period. Represent player i's *forecast* $p_{-}^t i = f_i(h^{t-1}) \in \Delta_{-i}$, where $f_i^t(x_{-i})$ is the probability that i assigns to her opponents playing the combination x_{-i} in period t.

A *learning process* then can be defined as an ordered pair $(f, g) = (f_1, ..., f_n, g_1, ...g_n)$, where $f : H \to \Delta_{-i}$ and $g_i : H \to \Delta_i$, for each player i.

4. Rationality, Prediction, and Convergence

To understand Kalai & Lehrer and Foster & Young's results, we need to consider their shared definition of rationality, and their different definitions of what it means to play

[4]Following standard practice, payoffs here are discounted so that we can compute the payoffs for infinite sequences of play.

a Nash equilibrium of game. For both, rationality is the familiar concept of expected utility maximization, given one's subjective beliefs. We can formulate this precisely as follows.

Definition 1 (Rationality)**.** *An agent i is rational if, for every finite history h^{t-1}, i's conditional strategy $g(\cdot|h^{t-1})$ optimizes i's expected utility from time t on, given i's conditional forecast $f(\cdot|h^{t-1})$. That is, for every alternative choice of strategy $g'(\cdot|h^{t-1})$,*

$$\int U_i(h)d\mu_i(f_i, g_i|h^{t-1}) \geq U_i(h)d\mu_i(f_i, g_i'|h^{t-1}).$$

This far, both K&L and F&Y are on the same page. However, their standards of convergence for Nash play differ in substantive ways. For K&L it is convergence to ϵ-Nash, and for F&Y it is the playing of Nash predicated on a specific concept of accurate prediction of the opponent's strategy. This is what makes their results compatible.

Definition 2 (ϵ-Nash)**.** *Given a game and a real non-negative parameter ϵ, a strategy profile is said to be an ϵ-equilibrium if it is not possible for any player to gain more than ϵ in expected payoff by unilaterally deviating from her ϵ-Nash equilibrium. A vector of strategies $\sigma \in \Delta$ is an ϵ-Nash equilibrium for G if*

$$u_i(\sigma) \geq u_i(\sigma_i^{'}, \sigma_(-i))) - \varepsilon \text{ for all } \sigma_i^{'} \in \Delta_i, i \in N.$$

Definition 3 (ϵ-close)**.** *Let $\epsilon > 0$ and let μ and μ' be two probability measures defined on the same space. We say μ is ϵ-close to μ' if there is a measurable set Q satisfying:*
(i) $\mu(Q)$ and $\mu'(Q)$ are greater than $1 - \epsilon$; and
(ii) for every measurable set $A \subseteq Q$, $(1 - \epsilon)\mu'(A) \leq \mu(A) \leq (1 + \epsilon)\mu'(A)$.

Definition 4 (Playing ϵ-close to ϵ-Nash)**.** *We say a learning process (f, g) learns to play ϵ-Nash of a game G, if for μ-almost all play paths there is a time T such that for every $t \geq T$ the play is ϵ-close to an ϵ-Nash of the repeated game.*

This is the standard employed by K&L. Note that playing ϵ-Nash does not require that players learn to predict the intended strategies of their opponents—only that they play as though their opponents were playing their portion of the Nash equilibrium. This should be familiar from convergence results in simple dynamics, such as Fictitious play, where players do not learn the actual strategies of opponents, but can converge in terms of proportions of play to the Nash equilibria of games (Fudenberg and Levine 1996). This stronger standard, that of accurate prediction of one's opponents' intended strategies, is what is required by Foster & Young.

Definition 5 (Prediction)**.** *We say that player i learns to predict if the mean squared error of i's next-period prediction goes to zero in almost all histories of play. That is, for $\mu(g)$-almost all realizations h,*

$$\lim_{T\to\infty} \sum_{t=1}^{T} |p_{-i}^t(h) - q_{-i}^t(h)^2|/T = 0.$$

Given a learning process (f, g) and specific history h, if the behavioral strategies come close to Nash equilibrium on h, then we expect the following condition to hold,

$$\lim_{T\to\infty}\left[\sum_{t=1}^{T} d(q^t(h), Q^N)^2\right]/T = 0. \tag{1}$$

This implies that, for every $\epsilon > 0$, play is within some ϵ of the Nash equilibrium at each point in time except possibly for a sparse set of times.

5. A Response to Foster & Young

Foster & Young present their impossibility theorem as a demonstration of the non-robustness of Bayesian learning as a justification of the Nash equilibrium concept. In (Young 2011), young states " — ". First, we should observe the limits of this result: for games lacking near-zero-sum subgames, it appears that Bayesian agents will learn to play Nash. Moreover, even in such games, while players do not learn to play Nash because they do not learn to predict the mixed strategies of their opponents, their actual proportions of play do correspond to Nash play (cf. fictitious play ()). That is, they play as though they had though had learned the Nash equilibrium. What we will show is that, even under a setup very similar to that of Foster & Young, rational agents can and learn Nash.

5.1. **Learning Mixed Strategies and Remaining Rational.**

K&L have shown us that rational agents will often play the Nash of games, F&Y have shown us that, in a significant class of games, Nash play comes apart from Nash prediction—that is, that rationality and prediction can conflict. In particular, this occurs under conditions of strong conflict of interest (near-zero-sum games) where only 'defensive' mixed strategy equilibria are viable, and where, moreover, there is uncertainty as to the payoffs of other players.

We propose that this can be undone by introducing modestly *more* uncertainty. Let us take up the exact same set of assumptions as F&Y, but allow that, instead of a single perturbation of payoffs before the first round of our game, we have continued perturbations—noise—in each round of play. What then? We will show that players will once again learn to predict the strategies of their opponents.

But, before we consider the results, let's consider the standard interpretations of perturbations, and how they are compatible with agent rationality. There are three relevant interpretations of such perturbations that we consider: un-observed player-specific payoff variations, imperfect implementation of strategies, and experimentation (Blume 2003). Each of these may provide a justificatory story for the introduction of noise to the actions of rational agents, and we may choose any we like.

1. Round-by-round payoff variations. Where F&Y consider a single set of perturbations to player payoffs prior to the first period of play, we can consider continued,

period-by-period perturbations. These perturbations can be thought of as either slight changes in the outcomes of player actions or minuscule oscillations in player preferences over time. These perturbations may be arbitrariliy small. As such, it seems plausible to posit that neither the world, nor players will remain perfectly static, and neither fact need conflict with the status of agents as rational actors.

2. Imperfect implementation. Noise in player actions may can also be caused by imperfections in the implementation of player strategies. This is the familiar model of where players may, through the slip-of-the-hand or 'tremble', play unintended strategies, albeit with negligible probabilities (Harsanyi and Selten 1988). An agent may be fully rational, in the sense of planning to act in accordance with a strategy that would maximize her expected utility, but she may also be less than absolutely perfect implementation of her strategy. That is, agents are subject to errors of implementation and not of reasoning. The result, formally, can be thought of as cost-dependent noise (Blume 2003).

3. Experimentation. Noisy play may also be justified as a form of hedging against unconceived alternatives. This approach has precedent in literature on the exploration-exploitation trade-off (Sutton and Barto 2012), where it can be beneficial to introduce strategic noise into one's actions. The idea here is that players may anticipate their ignorance of beneficial alternative paths of action, or the possibility of cycles or dead-end sequences of deterministic play, and so intentionally induce a modicum of noise in their actions.

Each of these interpretations is compatible with rationality. Yet only one is needed to justify the introduction of small cost-dependent noise into player strategies. Importantly, these can each be expressed in a formally equivalent way. We consider the case of period-by-period payoff perturbations.

5.2. **Rational Play and Learning Under Payoff Perturbations.** Here we use results from stochastic approximation theory to model the effect of period-by-period payoff perturbations. When each player's payoffs are perturbed in each period by random shocks a la Harsanyi (Harsanyi 1973), the limit behavior of a stochastic processes can be related to the limit behavior of a differential equation defined by the expected motion of the process. What this gives us is that each player's anticipated behavior in each period is a genuine mixed strategy.

The best known example of a choice probability function which can be generated from a random utility model is the logit choice function (McKelvey and Palfrey 1995a)(McKelvey and Palfrey 1995b),

$$P_{ij} = \frac{e^{\eta^{-1} EU_{ij}(P_{-i})}}{\sum_k e^{\eta^{-1} EU_{ik}(P_{-i})}} \tag{2}$$

where P_{ij} is the probability of player i choosing strategy j, and $EU_{ij}(P_{-i})$ is the expected utility to player i of choosing strategy j given other players are playing according to

	H	T
H	$1+\omega_{11}, -1+\omega'_{11}$	$-1+\omega_{12}, 1+\omega'_{12}$
T	$-1+\omega_{21}, 1++\omega'_{21}$	$1+\omega_{22}, -1+\omega_{22}$

the probability distribution P_{-i}. The logit choice function is generated by a random utility model whose random utility terms ϵ_i are i.i.d. with the double exponential distribution $F(x) = e^{-e^{-\eta^{-1}x-\gamma}}$, where γ is Euler's constant. However, we know that such a deterministic representation can be obtained for choice probabilities regardless of the distribution of the perturbations (Hofbauer and Sandholm 2002).

5.3. **Simulation.** We have that the behavior rational agents whose payoffs are perturbed period-by-period can be modeled by a deterministic approximation. Where the perturbations are drawn from the double exponential function, this approximation will take the form of the logit choice function. We can demonstrate, computationally, that rational agents of this sort will learn to play the Nash equilibrium of a near-zero-sum game a al Foster & Young. All assumptions are as in (Foster and Young 2001), with the exception of extending the payoff perturbation from once prior to play to every period of play.

Consider the two-player near-zero-sum game with the payoff matrix, where the ω_{ij}, ω'_{ij} are i.i.d. payoff perturbations in period t to players 1 and 2 respectively, distributed according to the double-exponential distribution F. This induces the logit best response function (4) as the behavioral strategy for our rational agent. Learning is Bayesian, and play is repeated indefinitely. Simulations of the game (Figure 1) reveal that rational Bayesian agents do indeed learn both to play the ϵ-Nash equilibria as K&L (Kalai and Lehrer 1993), and also to predict the strategies of their opponents, as F&Y would require(Foster and Young 2001).

5.4. **Proof.** Where we articulate a general, analytic result re. convergence of rational, Bayesian agents to Nash equilibrium play & prediction for near-zero sum games with period-by-period payoff perturbations (represented by the logic choice function).

6. Conclusion

References

Blume, L. E. (2003). How noise matters. *Games and Economic Behavior 44*(2), 251–271.

Foster, D. P. and H. P. Young (2001). On the impossibility of predicting the behavior of rational agents. *Proceedings of the National Academy of Sciences of the United States of America 98*(22), 12848–12853.

Fudenberg, D. and D. K. Levine (1996). *The Theory Of Learning In Games*.

Harsanyi, J. C. (1973). Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points. *International Journal of Game Theory 2*(1), 1–23.

Harsanyi, J. C. and R. Selten (1988). A general theory of equilibrium selection in games. Technical report.

Hofbauer, J. and W. H. Sandholm (2002). On the global convergence of stochastic fictitious play. *Econometrica 70*(6), 2265–2294.

Kalai, E. and E. Lehrer (1993). Rational Learning Leads to Nash Equilibrium. *Econometrica: Journal of the Econometric Society 61*(5), 1019–1045.

McKelvey, R. D. and T. R. Palfrey (1995a). Quantal Response Equilibria for Normal-Form Games. *Games and Economic Behavior 10*(1), 6–38.

McKelvey, R. D. and T. R. Palfrey (1995b). Quantal Response Equilibria for Normal-Form Games. *Games and Economic Behavior 10*(1), 6–38.

Nash, J. (1951). Non-Cooperative Games. *Annals of Mathematics 54*(2), 286–295.

Sutton, R. S. and A. G. Barto (2012). Reinforcement learning. *Learning 3*(9), 322.

Young, H. P. (2011). *Strategic Learning and its Limits*, Volume 9780199269.