HIER

Harvard Institute of Economic Research

Discussion Paper Number 1926

Learning To Play Bayesian Games

By

Eddie Dekel, Drew Fudenberg and David K. Levine

July 2001

Harvard University
Cambridge, Massachusetts

Learning to Play Bayesian Games[1]

Eddie Dekel, Drew Fudenberg and David K. Levine[2]

First draft: December 23, 1996

Current revision: July 20, 2001

Abstract

This paper discusses the implications of learning theory for the analysis of Bayesian games. One goal is to illuminate the issues that arise when modeling situations where players are learning about the distribution of Nature's move as well as learning about the opponents' strategies. A second goal is to argue that quite restrictive assumptions are necessary to justify the concept of Nash equilibrium without a common prior as a steady state of a learning process.

---

[1] This work is supported by the National Science Foundation under Grants 99-86170, 97-30181, and 97-30493.

[2] Departments of Economics: Northwestern and Tel Aviv Universities, Harvard University and UCLA.

1. Introduction

This paper discusses the implications of learning theory for the analysis of Bayesian games. One of our goals is to illuminate some of the modeling issues involved in thinking about learning about opponents' strategies when the distribution of Nature's moves is also unknown. A more specific goal is to investigate the concept of Nash equilibrium without a common prior, a solution concept that has been applied in a number of recent economic models. The status of this solution concept is important, given the recent popularity of papers that apply it, such as Banerjee and Somanathan [2001], Piketty [1995], and Spector [2000].[3] We argue that this concept is difficult to justify as the long-run result of a learning process. The intuition for this claim is simple: In order for repeated observations to lead players to learn the distribution of opponents' strategies, the signals observed at the end of each round of play must be sufficiently informative. Such information will tend to lead players to also have correct and hence identical beliefs about the distribution of Nature's moves. While our basic argument is straightforward, our examples highlight some less obvious points.

It is known that in simultaneous-move complete-information games, if players observe the profiles of actions played in each round, a wide range of learning processes have the property that the set of steady states coincide with the set of Nash equilibria of the game. By contrast, we show that with incomplete information, if players begin with inconsistent priors, there are games in which the Nash equilibria are not steady states of any plausible learning processes. Moreover, in many games of incomplete information without a common prior, the assumptions that imply that all Nash equilibria are steady states imply that many other outcomes are steady states as well, questioning the focus on only the Nash equilibrium. At the same time, we do identify some environments and games where Nash equilibrium without a common prior does have a learning justification.

This learning-theoretic critique is related to two other problems of Nash equilibrium without a common prior. One is internal consistency: a Nash equilibrium when players have different priors in general is not a Nash equilibrium when Nature is replaced with a player who is indifferent among all her choices and who behaves exactly as did Nature.[4] A related problem (Dekel and Gul [1997]) is that the epistemic foundations of Nash equilibrium without a common prior are unappealing. The epistemic foundation for Nash equilibrium relies on a common prior about strategies, and it is not obvious why we should impose this on the states of Nature underlying the strategic uncertainty and not on those corresponding to the incomplete information.

In Section 2 we present our basic solution concept, self-confirming equilibrium, which we motivate as a steady state of a learning process in which there is only one person in each player role and Nature's choice of the state of the world is iid over time. The relationship between this notion of self-confirming equilibrium and Nash equilibrium in games of incomplete information with diverse priors is discussed in Section 3. Finally, in Section 4 we discuss how the definition of self-confirming equilibrium needs to be revised if Nature's choice is not iid, or if there are many learning agents who are randomly placed in each player role.

2. The Model

We consider a static simultaneous-move game with I player roles.[5] (All parameters of the game, including the number of players, and their possible actions and types, are assumed to be finite.) In the static game, Nature moves first, determining players' types, which we denote $\theta_i \in \Theta_i$. To model cases where the types alone do not determine the realized payoffs, we also allow Nature to pick $\theta_0 \in \Theta_0$; we call this "Nature's type." Players observe their types, and then simultaneously choose actions $a_i \in A_i$ as a function of their type, so that a strategy σ_i for player i is a map from her

---

[3] We do not explore these applications in detail, so in particular we do not claim that their use of Nash equilibrium is inappropriate. We only want to argue that in the context of incorrect priors, the use of Nash equilibrium requires more careful justification than is typically given. In fact, Spector [2000] assumes that actions are observed while payoffs are not, noting that, while these are fairly extreme assumptions, if payoffs were observed then the players would learn the true distribution of Nature's move.

[4] This is because in a Nash equilibrium the strategy of the player replacing Nature is known.

[5] Fudenberg and Kreps [1988, 1995] and Fudenberg and Levine [1993] examined learning and steady states in games with non-trivial extensive forms; we do not consider such games so as to focus on the role of incomplete information and incorrect priors.

types to mixed actions. Player i's utility $u_i(a,\theta)$ depends on the profile $a=(a_1,...,a_I)\in A$ of realized actions, and on the realization $\theta=(\theta_0,\theta_1,...,\theta_I)\in\Theta$ of Nature's move. When $u_i(a,\theta)=u_i(a,\theta_i)$ we refer to the game as having *private values*. For any finite set X**,** we let $\Delta(X)$ denote the space of probability distributions over X. Thus player i's prior about Nature's move is denoted $\mu^i\in\Delta(\Theta)$; $\mu=\{\mu^1,...\mu^I\}$ is the profile of priors. When $\mu^i=\mu^j$ for all i and j, the game has a *common prior*; in the complementary case where $\mu^i\neq\mu^j$ for some i and j we say that the priors are *diverse*.

Our solution concept is motivated by thinking about a learning environment in which the game given above is played repeatedly. We suppose that players know their own payoff functions and the sets of possible moves by Nature (Θ) and players (A); but they know neither the strategies used by other players nor the distribution of Nature's move; the players learn about these latter variables from their observations after each period of play. We also suppose that each period the types are drawn independently over time from a fixed distribution p. Thus p corresponds to the true distribution of Nature's move in the stage game, and when $\mu^i=p$ for all players i we say that *the priors are correct*.[6] For the time being, we also assume that there is a single agent in each player role. Section 4 discusses the case where there is a large population of agents in each role who are matched together to play the game; we also discuss there the possibility that types are generated by a more general stochastic process.

Of course, what players might learn from repeated play depends on what they observe at the end of each round of play. To model this, we suppose that after each play of the game, players receive private signals $y_i(a,\theta)$ which is their only information about Nature's and their opponents' moves. It is natural to assume that players observe their own actions and types**,** but whether or not they observe others' actions, or their own and others' payoffs, depends on the observation structure and will affect which outcomes can arise in a steady state. We assume that each player observes her own private signal y_i, along with her own action and own type.[7]

---

[6] Note that if players are Bayesians they will have a "prior" about the state of the overall learning process, and this prior need not be the fixed "prior" μ^i that is taken as data in the specification of the stage game. We call the latter objects "priors" to conform to past usage, but the language is inaccurate once we set the stage game in a repeated learning setting.

[7] We consider the case in which knowledge of opponents' play comes *only* from learning by observation and updating, and not from deduction based on opponents' rationality, so we do not require that players know their opponents' utility functions or beliefs. Rubinstein and Wolinsky [1994] explore steady states in

We will not formally model the dynamics of learning, but will appeal informally to the idea that a steady state of a belief-based learning process must be a unitary self-confirming equilibrium (Fudenberg and Levine [1993]). Thus, our focus is on how the information that players observe at the end of each round of play determines the set of self-confirming equilibria, and how these equilibria relate to the Nash equilibria of the game.

The key components of self-confirming (and Nash) equilibrium are each player *i*'s *beliefs* about Nature's move, her *strategy,* and her *conjecture* about the strategies used by her opponents. Player *i*'s beliefs, denoted by $\hat{\mu}^i$, are a point in the space $\Delta(\Theta)$ of distributions over Nature's move, and her strategy is a map $\sigma_i : \Theta_i \to \Delta(A_i)$. The space of all such strategies is denoted Σ_i, and the player's conjectures about opponents' play are assumed to be a $\hat{\sigma}_{-i} \in \times_{-i} \Sigma_{-i}$, that is, a strategy profile of *i*'s opponents. The notation $\hat{\mu}^i(\cdot \mid \theta_i)$ refers to the conditional distribution corresponding to $\hat{\mu}^i$ and θ_i, while $\sigma_i(a_i \mid \theta_i)$ denotes the probability that $\sigma_i(\theta_i)$ assigns to a_i.

Definition: *A strategy profile* σ *is a* self-confirming equilibrium with conjecture $\hat{\sigma}_{-i}$ *and* beliefs $\hat{\mu}_i$ *if for each player* i*,*

(i) $p(\theta_i) > 0$ *implies* $\hat{\mu}_i(\theta_i) > 0$*,*

and for any pair $\theta_i, \hat{a}_i$ *such that* $\hat{\mu}^i(\theta_i) \cdot \sigma_i(\hat{a}_i \mid \theta_i) > 0$ *both the following conditions are satisfied*

(ii) $\hat{a}_i \in \arg\max_{a_i} \sum_{a_{-i}, \theta_{-i}} u_i(a_i, a_{-i}, \theta_i, \theta_{-i}) \hat{\mu}^i(\theta_{-i} \mid \theta_i) \hat{\sigma}_{-i}(a_{-i} \mid \theta_{-i})$,

and

(iii) $$\sum_{\{a_{-i}, \theta_{-i} : y_i(\hat{a}_i, a_{-i}, \theta_i, \theta_{-i}) = \bar{y}_i\}} \hat{\mu}^i(\theta_{-i} \mid \theta_i) \hat{\sigma}_{-i}(a_{-i} \mid \theta_{-i})$$
$$= \sum_{\{a_{-i}, \theta_{-i} : y_i(\hat{a}_i, a_{-i}, \theta_i, \theta_{-i}) = \bar{y}_i\}} p(\theta_{-i} \mid \theta_i) \sigma_{-i}(a_{-i} \mid \theta_{-i}).$$

---

learning processes where there is common knowledge of rationality, and Dekel, Fudenberg and Levine [1999] consider the case of almost common knowledge of rationality.

We say that σ is a *self-confirming equilibrium* if there is some profile $(\hat{\mu},\hat{\sigma})$ such that (i), (ii) and (iii) are satisfied.[8]

Condition (i) is a consequence of the assumption that players observe their own types. Condition (ii) says that any action played by a type of player *i* that has positive probability is a best response to her conjecture about opponents' play and beliefs about Nature's move. Condition (iii) says that the distribution of signals (conditional on type) that the player expects to see equals the actual distribution. This captures the least amount of information that we would expect to arise as the steady state of a learning process.

We will sometimes consider the restriction of self-confirming equilibria to the case where players' beliefs about Nature satisfy certain restrictions. In particular, we say that a self-confirming equilibrium has "independent beliefs" if for all players *i* the beliefs $\hat{\mu}^i$ are a product measure. Because the domain of $\hat{\mu}^i$ is all of $\Theta_0 \times \Theta_1 \times ... \Theta_I$, independence implies that player *i*'s beliefs about the types of her opponents do not depend on her own type. This restriction is most easily motivated in games where the true distribution *p* is a product measure, that is, players' types are in fact independent, as in this case assuming independent beliefs amounts to saying that players understand this particular fact about the structure of the game. The following game demonstrates the effect of assuming independent beliefs.

Example 1: Independent Beliefs

Consider a one-person two-type two-action game with two different states in Θ_0. The actions are labeled *In* and *Out*; the types are labeled "*Timid*" (*T*) and "*Brave*" (*B*), the Θ_0 states are labeled *L* and *R*. Both types get a payoff of 0 from *Out*. Payoffs from *In* are given in the table below.

[8] Battigalli [1988] and Kalai and Lehrer [1993] defined similar concepts. This is the "unitary" version of self-confirming equilibria because it supposes that there is a single $\hat{\sigma}_{-i}$ for each player *i*. The unitary version of the concept is appropriate here because of our assumption that there is a single agent in each player role; when we discuss large populations and matching in Section 4 we will need to allow for heterogeneous beliefs. Note that *i*'s beliefs about opponents' play take the form of a strategy profile as opposed to a probability distribution over strategy profiles. The complications that arise due to correlations in conjectures are discussed in Fudenberg and Kreps [1988] and Fudenberg and Levine [1993]; we simplify by ignoring them here.

	L	*R*
Brave	1	2
Timid	2	-1

Notice that *In* is a dominant strategy for the Brave type. Suppose the player does not observe Nature's move but does observe her own payoff.[9] Suppose also that the objective distribution *p* on Nature's move assigns equal probability to the four states (*B*, *L*), (*B*, *R*), (*T*, *L*) and (*T*, *R*). The Brave type has *In* as a dominant strategy, and so *Brave* will go *In* in every self-confirming equilibrium. Thus, since the player observes her payoff, the *Brave* type learns the distribution of Nature's move conditional on *Brave*, so the only self-confirming equilibrium with independent beliefs has $\hat{\mu} = p$ and both types playing *In*. However, there is also a self-confirming equilibrium without independent beliefs where the *Timid* type stays *Out* because the player believes that Nature plays *R* whenever the player is *Timid*, that is $\hat{\mu}^i(R \mid T) = 1$. ■

We are interested in the relationship between the set of self-confirming equilibria and the set of Nash equilibria. In a *Nash equilibrium*, each player's strategy must maximize her expected payoff given her prior about the distribution of θ and correct conjectures about the play of the opponents.

Definition: *A strategy profile* σ *is a* Nash equilibrium *with conjecture* $\hat{\sigma}_{-i}$ *and beliefs* $\hat{\mu}_i$ *if for each player* i, *and for any pair* $\theta_i, \hat{a}_i$ *such that* $\hat{\mu}^i(\theta_i) \cdot \sigma_i(\hat{a}_i) > 0$

(ii) $\hat{a}_i \in \arg\max_{a_i} \sum_{a_{-i},\theta_{-i}} u_i(a_i, a_{-i}, \theta_i, \theta_{-i}) \hat{\mu}^i(\theta_{-i} \mid \theta_i) \hat{\sigma}_{-i}(a_{-i} \mid \theta_{-i})$,

and

(iii') $\hat{\sigma}_{-i} = \sigma_{-i}$, $\hat{\mu}^i = \mu^i$.

[9] Note that even though player 1 and Nature move simultaneously, when Nature's move is not observed the problem has the structure of a one-armed bandit, which is normally thought of as a game with a non-trivial extensive form.

Note in particular that (iii') has the further implication that $\mu^i(\theta_i) > 0$ implies $\hat{\mu}^i(\theta_i) > 0$.[10]

When the priors are diverse, we say that the Nash equilibrium has *diverse priors*. Finally, to distinguish the case where the beliefs are correct, that is $\mu^i = p$ for all *i*, we say this is *a Nash equilibrium with correct priors*.

Note that the set of self-confirming equilibria does not depend on the exogenous stage-game priors μ. To see why, note that a complete belief-based learning model would specify priors over both Nature's probability distribution and opponents' strategies. These priors would be updated over time, so that the steady state belief-conjecture pair ($\hat{\mu}$, $\hat{\sigma}_{-i}$) need not be the same as the priors. In the learning process, different priors can lead to a different distribution over steady states; in our definition the set of self-confirming equilibria corresponds to the set of possible steady states for all initial conditions of the learning process.

Note also that there need not exist a self-confirming equilibrium where beliefs coincide with the exogenous prior μ. For example, if players observe Nature's move at the end of each period, then self-confirming equilibrium requires that the beliefs equal the objective distribution *p*. Conversely, there is always a self-confirming equilibrium with beliefs μ if players observe nothing at all, so that the set of outcomes *Y* has only a single element, but in that case the set of self-confirming equilibria includes all profiles of *ex-ante* undominated strategies.[11] (Since we consider different classes of games and different assumptions about the signals observed at the end of each round, we end up making several points. Since these points do not require lengthy proofs, we do not display them formally; instead, we underline them to make them easier to locate.)

---

[10] This definition of Nash equilibrium allows for a player to believe that an opponent is not optimizing, since *j* can assign strictly positive probability to a type of *i* to which *i* assigns zero probability. To deal with this issue we could state the primitives of the game as conditional probabilities $\mu^i(\theta_{-i} \mid \theta_i)$ and impose interim optimality even for own types to which one assigns zero probability. We chose to avoid this extra complexity in the notation.

[11] The strategies are *ex-ante* undominated because there is only one agent in each player role, so that an agent's conjectures about the other players' strategies must be the same regardless of that agent's action and type, and the belief about Nature must also be conditionally independent of the action chosen given the type.

3. The Relationship Between Self-Confirming Equilibria and Nash Equilibria

In this section we focus on the relationship between self-confirming equilibria and Nash equilibria. Specifically we explore the assumptions about observability under which the set of self-confirming equilibrium profiles with beliefs $\hat{\mu} = \mu$ coincide with the set of Nash equilibrium profiles of the game where players' priors regarding Nature are μ. We abbreviate this by saying (imprecisely) that the sets of Nash and self-confirming equilibria with beliefs μ coincide.

3.1 The tension between Nash and self-confirming equilibria

As mentioned above, if players cannot observe or deduce their opponents' actions at the end of each period, then in general there can be self-confirming equilibria that are not Nash equilibria. So we begin by considering the case in which players either directly observe, or indirectly deduce from other observations, the realized actions of their opponents after each play of the game.

Suppose that $y_i = u_i$, that is, players observe their own utility. With generic payoffs the map $u_i(a,\theta)$ is 1-1, and both the actions of other players, a_{-i}, and Nature's move, θ, can be uniquely determined from y_i. Consequently, the only beliefs and conjectures that are self-confirming are the correct ones. <u>We conclude that with generic and observed payoffs, the set of self-confirming equilibria coincides with the set of Nash equilibria of the game with the correct (hence common) prior.</u> In particular, if the priors in a given game of incomplete information are not common, and in addition, if the set of Nash equilibria of that game differs from the set of Nash equilibria with the correct prior (that is if the presumption of diverse priors has any significance), then the Nash equilibria of the game with diverse priors will not coincide with the self-confirming equilibria.

Next suppose that $y_i = a_{-i}$, that is, players observe their opponents' actions. Then in a self-confirming equilibrium players must know the conditional distribution of opponents' actions given their own type. Suppose in addition that the game is a game of private values, that is, $u_i(a,\theta) = u_i(a,\theta_i)$. Since players do not care about their opponents' types, this implies that <u>with private values and observed actions every self-</u>

confirming equilibrium has the same strategies as a Nash equilibrium of the game with the correct and hence common priors.[12,13] Once again, the fact that self-confirming equilibria are Nash equilibria of the game with the correct prior implies in particular that when such games have diverse priors, then any Nash equilibria that are not Nash in the game with correct priors are not self -confirming. That is, if the diverse priors create any additional Nash equilibria, those equilibria are not self-confirming.[14] This is demonstrated in Example 2 below.

Example 2: Nash Equilibria that are Not Self-Confirming Equilibria

We consider a game with a column player, *C*, and two row players, *R1* and *R2*. Nature chooses *L* or *R*, with equal probability; the column player observes Nature's choice of *L* or *R*, while the two other players do not. Thus players *R1* and *R2* each have a single type, player *C* has two types, *L* and ***R*****,** and the set Θ_0 of Nature's types is empty.

In this game, *C*'s payoff depends only on her own action and type, but not on the actions taken by the row players: specifically, *C*'s actions are labeled *l* and *r*, and *C* gets 1 for choosing the same as Nature, and 0 for choosing the opposite. The row players' payoffs each depend on the column player's action and their own action, as shown in the following two matrices.

---

[12] Note that the beliefs about opponents' types and strategies may not correspond to the Nash equilibrium; we return to this question in other contexts below.

[13] In this case the specification of assumptions about player's prior knowledge of the game is irrelevant for long-run learning. It may however make a difference for predictions of the model about play in the first few periods, before much learning has taken place, and so can also influence the long-run behavior of the system in cases where a steady state is not reached.

[14] Note that in two-person games with private values (in which by implication θ_0 is irrelevant) the set of Nash equilibria of a game with diverse priors depends only on the support of the priors and equals the set of Nash equilibria of the same game with common priors (and the same support). The key point is that a player's prior about her own type does not matter to him, and so there is no harm in modifying it to reflect her opponent's belief. In a similar vein, neither player cares about Nature's type θ_0. Formally, if σ is a Nash equilibrium where player 1's prior on $\Theta_1 \times \Theta_2$ is $\mu^1 = \mu_1^1 \times \mu_2^1$, and two's prior is $\mu^2 = \mu_1^2 \times \mu_2^2$, then it is also an equilibrium when the priors are both $\mu_1^2 \times \mu_2^1$.

R1	l	r
U	0	1
M	¾	¾
D	1	0

R2	l	r
U	1	0
M	¾	¾
D	0	1

This is a game with private values, because the row players' payoffs depend only on the column player's action, not her type. In the learning environment, everyone observes the column player's action after the game is played. Clearly the column player has a dominant strategy of playing *l* when type *L* and *r* when type *R*, so in the self-confirming equilibrium, the column player plays *l* on *L* and *r* on *R*, so plays each half the time. The row players observe this, so must play *M*.

Now suppose that *R1*'s prior assigns probability .9 to type *L* and .1 to *R*, while *R2*'s prior is the reverse, with .1 probability of type *L* and .9 to *R*. In a Nash equilibrium, *C* plays *l* upon observing *L* and *r* upon observing *R*, and the row players know this. Given the priors, this implies that *R1* and *R2* believe that they will face the actions *l* and *r* respectively .9 of the time. Consequently, in this Nash equilibrium with diverse priors, *R1* and *R2* will both choose *D*. However, this is not a Nash equilibrium for any common prior, and so it is not a self-confirming equilibrium for any *p* when the column player's action is observed.[15]

We see in this example that, when players observe actions, the self-confirming equilibria in which beliefs are equal to the priors is unique, and is different from the Nash equilibrium. When players observe nothing at all, the set of self-confirming equilibria with beliefs equal to the priors includes the Nash equilibrium, but in fact imposes no restrictions at all on the play by *R1* and *R2* since the row players will not know anything about columns choice. ■

[15] One way of summarizing this example is to say that two players who agree about an opponent's *strategy* can have different forecasts about the distribution of that opponent's actions if they have different beliefs about the distribution of that player's type. In contrast, with observed actions players correctly forecast the distribution of opponent's actions in any self-confirming equilibrium, but they can have different beliefs about the distribution of Nature's move and about the opponent's strategy.

<u>Without private values, if neither types nor payoffs are observed, but actions are, there can be self-confirming equilibria with correct beliefs about Nature that are not Nash even with correct priors</u>, as in the next example.

Example 3: Self-confirming equilibria that are not Nash

Player *R* and player *C* each choose either –1 or 1. Player *R*'s type is either +1 (with probability 2/3) or –1 (with probability 1/3), and player *R*'s payoff is her action times her type, so player *R* plays +1 when type 1 and –1 when type –1. Player *C*'s payoff is the product of player 1's type and the two actions, so the unique Nash equilibrium with the correct prior has player *C* play +1. If all that player *C* observes is player *R*'s action, then player *C* can have correct beliefs about Nature's move and conjecture that player *R* plays +1 when type –1 and mixes ½- ½ when type +1. In this case the best response is for player *C* to play –1. Consequently, player *C* plays –1 in all self-confirming equilibria. ■

By assuming that players observe very little (or that payoffs are not generic) it is easier for a Nash equilibrium with diverse priors to also be a self-confirming equilibrium. However, the less the players observe when the game is played, the less they learn about opponents' strategies, so the bigger the set of self-confirming equilibria, which makes it more likely that the set of self-confirming equilibria contains equilibria that are not Nash. This tension is noted following Example 2 above; to further illustrate it we consider the following team problem, based on the model of Banerjee and Somanathan [2001].

Example 4: Nash Equilibria and Self-confirming equilibria in a Team Problem

Nature chooses either 0 or 1. The team as a group needs to choose a single action in {0, 1}; each player's utility is 1 if the action matches Nature's choice and 0 otherwise. There is a continuum of players on the unit interval, $i \in [0,1]$. The game does not have common prior; instead, each player *i* believes that Nature plays "1" with probability *i.* Nature's actual probability is some fixed *p* not too far from ½. One of the players is chosen as the decision-maker. All other players observe a noisy signal that takes value 1 or 0, where for each player the probability that the signal value equals Nature's choice is $\alpha > 1/2$. After observing these signals, one player is chosen at random to send the decision-maker a message in $\{0, 1\}$. The decision-maker does not observe the identity of the sender, and then chooses the group's action.

This game has several Nash equilibria. One of these, which seems analogous to those on which Banerjee and Somanathan focused, has an interval of players of measure t around the decision maker send the message that corresponds to their signal. In the two intervals of measure $(1-t)/2$ consisting of players with more extreme priors players ignore their signals and always send 0 and 1, depending on whether they are at the "right" or "left" extreme. We now explore whether this Nash equilibrium (with diverse priors) is a self-confirming equilibrium under different assumptions regarding what players observe when the game is played.

Obviously if players observe the true state at the end of each period they will learn the actual distribution of Nature's move, so diverse priors would not persist. If players observe their own payoff and the decision-maker's action they would also learn the distribution of Nature's move. On the other hand, if the signal senders observe only their payoffs, and the decision maker observes nothing but her action, the Nash equilibrium is a self-confirming equilibrium with the given beliefs about Nature and the correct conjectures about players' strategies because in this equilibrium the distribution over payoffs is $(1-t)/2+\alpha t$, independent of the state.[16] The Nash equilibrium is also a self-confirming equilibrium if the players observe actions but not payoffs; but once again in this case there are other self-confirming equilibria as well. Finally, if the decision-maker observes her payoff, she can deduce the state and thereby learn the true distribution; in this case the Nash equilibrium is a self-confirming equilibrium only if the decision-maker's prior is correct. ■

We have seen that observing actions is not sufficient for Nash and self-confirming equilibria with given beliefs μ to coincide. The next example sharpens this point. Even if the set of strategy profiles in self confirming equilibria with beliefs $\mu = \hat{\mu}$ coincides with the set of Nash equilibria, conjectures about opponents' play may fail to be correct,

[16] There may be other SCE, but as the original paper does not analyze the whole set of Nash equilibria, we only emphasize here the stringent assumptions necessary to make *the particular* Nash equilibrium they consider correspond to a self-confirming equilibrium with the same beliefs.

and so the profile can fail to be self-confirming once actions are added to the available information.

Example 5: A game where when payoffs are observed, Nash equilibrium and self-confirming equilibrium are equivalent iff actions are not observed.

Consider a two-player game in which Nature chooses the left or right matrix. Neither player has private information; both players think the left matrix is chosen with probability $1-\varepsilon$, while it is actually chosen with probability ε.

	A	*B*
A	1, 1	0, 0
B	0, 0	0, 0

	A	*B*
A	0, 0	1, 1
B	1, 1	0, 0

The strategic form for this game given the common beliefs μ is

	A	*B*
A	$1\text{-}\varepsilon$, $1\text{-}\varepsilon$	ε, ε
B	ε, ε	0, 0

The unique Nash equilibrium is (*A*, *A*).

If players only observe their payoffs, then (*A*, *A*) is a self-confirming equilibrium with beliefs $(1-\varepsilon,\varepsilon)$ and conjecture that the opponent is playing *B*: in this case each player believes that playing *A* yields 1 with probability ε, and *B* yields 0. However, if players observe actions, the Nash equilibrium (*A*, *A*) is not self confirming. ■

3.2 Examples where Nash equilibria and self-confirming equilibria coincide

Example 2 showed that when players observe everyone's actions there could be Nash equilibria that are not self-confirming with respect to any beliefs. Our next example shows a sort of converse: for some games with diverse priors, the Nash equilibria and self-confirming equilibria do coincide, whether or not players observe actions. In this example the diverse priors are significant: the set of Nash (and self-confirming) equilibria with the diverse priors differs from the set of Nash (and self-confirming) equilibria with a common prior. The example demonstrates this point using *ex-ante* dominating strategies, in which it is irrelevant what players observe; in an example in the appendix the players do care about their opponents' actions, and in that example players must observe either their own payoffs, or opponents' actions, or both.

Example 6: A game where Nash equilibrium and self-confirming equilibrium coincide even with diverse priors

	L	*R*
U	1, 1	0, 0
D	0, 0	-1, -1

	L	*R*
U	-1, -1	0,0
D	0, 0	1, 1

This is a two-player game in which Nature chooses the left (*l*) or right (*r*) payoffs, and neither player observes Nature's move. The row player believes the left payoffs are chosen, the column player believes the opposite: $\mu^1(l) = \mu^2(r) = 1$. So the unique Nash equilibrium is for the row player to play *U* and the column player *R*, with payoffs (0, 0). Whether or not players observe their opponent's actions or their own utility, this profile is self-confirming with beliefs equal to the given priors. However, the subset of self-confirming equilibria with beliefs in which $\hat{\mu}^1 = \hat{\mu}^2$ is either (*U*, *L*), (*D*, *R*), or the entire strategy space. ■

To summarize, we have seen that observing actions is neither necessary nor sufficient for self confirming and Nash equilibria to coincide, and that when payoffs are observable, observing actions in addition may make a Nash equilibrium that was self-confirming no longer so. Thus, we see that the "best" case for Nash and self-confirming equilibria to coincide is when actions are observable and payoffs are not. Moreover, we saw that this assumption can only be useful when the game is not one of private values. Then, if $u_i(a,\theta) = u_i(a,\theta_i,\theta_0)$, and if players observe actions and not payoffs, they can correctly infer opponents' distribution of actions as a function of their own type. However, while this forces them to agree about the distribution of private types, it does not force them to agree about θ_0, the portion of Nature's move that is unknown to everyone. So in a self-confirming equilibrium players may disagree about the distribution of θ_0 and how it is correlated with the private types. In this class of games, where each player's utility depends only on that player's type and the type of Nature, the Nash equilibria coincide with the self-confirming equilibria when actions are observed but payoffs are not. A practical example that has this flavor is the case of voting by juries. Assume that jurors all wish to convict the guilty and release the innocent, and have the same preferences of the cost of making a mistake. Here we would take θ_0 to be whether or not the defendant is guilty, and suppose that there are no private types. Since they never get to observe whether a defendant is actually guilty they do not observe their own payoffs. But if some jurors get to participate in many trials, or hear about the deliberations of many juries, they may learn the strategies of other players, even though they never acquire any information about the correctness or incorrectness of their beliefs about the relationship between evidence and guilt.

Next consider more general games, in which players do care directly about opponents' types. The Nash equilibria will coincide with the self-confirming equilibria with the same beliefs if players observe opponents' types and actions, but do not observe their own payoffs, and if additionally the priors over private types are correct. (That is, the marginal of μ on players' types coincides with the marginal of *p* on players' types.)

4. The Joint Distribution of Nature's Moves over Time, Agents, and Players

4.1. Correlation Over Time

The assumption we made in Section 2 to justify self-confirming equilibrium as the steady state of a learning process is that Nature makes independent draws from *p* each period. However, there are stochastic processes generating agents' types that, loosely speaking, "draw types according to *p*" but where *p* turns out not to be directly relevant for the long-run outcome. For example, Nature might make a single once-and-for-all draw $\hat{\theta}$ from *p*, so that the "real game" about which learning occurs corresponds to the particular realization of $\hat{\theta}$. In this case, the "appropriate" definition of self-confirming equilibrium replaces *p* in condition (ii) with the degenerate distribution that chooses $\hat{\theta}$

with probability one. This is "appropriate" in the sense that the players learn only about the particular draw, and so steady states of a learning procedure will coincide with the self-confirming equilibria when p is replaced in this manner. In the remainder of this section we point out, for various stochastic processes, what is the "appropriate" replacement of p.

Obviously the fact that the realized $\hat{\theta}$ differs from p can have significant implications for what is learned. For instance, in the case just mentioned, players may learn the true state. The next example illustrates another possible implication: the players might learn the state for some realizations of $\hat{\theta}$ and not for others.

Example 7: Learning about only some states

Consider again the one-player game of Example 1, where the payoffs for Out are 0 for both types and for *In* are

	L	*R*
Brave	1	2
Timid	2	-1

As before, suppose the player does not observe Nature's move but does observe her own payoff. But now suppose also that the objective distribution p assigns probability ½ to (B, L) and ½ (T, L) so that Nature's type is always L. As before, the only self-confirming equilibrium with independent beliefs is for both types to play *In*, but this solution concept is not appropriate in the case when $\hat{\theta}$ is drawn once from p and then fixed. The supposition underlying self-confirming equilibrium is that player i has a single belief $\hat{\mu}^i$ over Θ that is possibly updated based on θ_i. This makes sense when a given agent gets many observations of games with each possible realization of θ_i. With a single draw, however, we need to allow for each type θ_i to have separate beliefs. In this case, even when player i knows that the distribution of Nature's type is independent of her own, there is a self-confirming equilibrium where the *Timid* type stays *Out* (believing that Nature always plays R) while the *Brave* type plays *In* believing (correctly) that Nature always plays L. ■

Underlying our notion of a steady state is the idea that players repeatedly sample from a fixed distribution that does not change over time. Suppose we consider the more general class of exchangeable processes for types, which have a representation as a "prior" probability distribution over (conditionally) iid processes. Then we can think of Nature making a single once-and-for-all draw $\hat{p}$ from the class of iid processes, and the "appropriate distribution" to use in the definition of μ self-confirming equilibrium is the $\hat{p}$ drawn by Nature; the fact that players "could have" faced some different distribution and that the overall distribution was *p* is not relevant in the long-run steady state.[17]

Thus, we can extend the discussion of independent private values in Section 3, where we said that every self-confirming equilibrium has the same strategies as a Nash equilibrium of the game with the correct and hence common priors: <u>With independent private values and observed actions of opposing players, a self-confirming equilibrium is a Nash equilibrium in a game with priors equal to the "realized distribution" of types, $\hat{p}$.</u> One can also consider the more general class of ergodic processes instead of exchangeable ones. It is natural in that case to think of *p* as the invariant distribution. Notice in this case that players are not actually drawing from *p* each period, rather they are drawing from time-varying distributions which average out to *p*. <u>If players believe that the true process is exchangeable, then beliefs in steady states will still satisfy the self-confirming conditions of section 2 with respect to this ergodic distribution.</u>[18]

4.2 Matching in Large Populations

Next we want to focus on a class of games of special interest in learning theory: games in which players are randomly matched to play a smaller "stage" game. In this setting it is natural to think of *p* as the distribution of types for a given match, and not deal directly with the distribution of types over all matches, but we must also consider the relationship between the matching process and the *p* from which players draw their observations.

---

[17] Note that the exchangeable model nests both the case of single once-and-for-all draw $\hat{\theta}$ and the case where each period's θ is an independent draw from *p*. Note also that the distribution from which Nature chooses $\hat{p}$ does influence the ex-ante distribution over steady states.

[18] Of course, sophisticated players might realize that Nature's moves do not have an exchangeable distribution, in which case our model would break down.

Suppose that players in a given player role are independently matched in each period with opponents in other roles. We mention three ways in which types could be determined over time. (a) If Nature makes a once-and-for-all draw, the relevant distribution is generated by independent draws from the realized distribution of $\theta 's$ in the population. This distribution exhibits independence across player roles on account of the matching procedure, even if the underlying *p* exhibited correlation. For instance, if types were perfectly correlated according to *p*, then in any rematch after the first draw there is positive probability that a profile with zero probability according to *p* will nevertheless be matched. (b) If Nature draws independently from a correlated distribution *p* each period, and the draw is made prior to matching, a similar observation holds true: Because the matching is independent of type, the relevant distribution is the product of the marginals from *p*. (c) On the other hand, if Nature draws independently each period, and draws the types for the agents in each match from *p* after the match is formed, then *p* is the relevant distribution.

Now we turn to the question of how the appropriate definition of self-confirming equilibrium depends on the specification of the stochastic process governing types. We know from Fudenberg-Levine [1993] that when there are multiple agents in each player role, there can be "heterogeneous" self-confirming equilibria in which different agents in the same role play different strategies and have different conjectures. Thus, when types are chosen independently over time, and separately for each match, the appropriate definition replaces allows the beliefs and conjectures of the agents to vary with the strategy chosen.

Definition: *A strategy profile* σ *is a* heterogeneous self-confirming equilibrium *if for each player* i *there exists* $\{\sigma_i^k : k \in K\} \subset \Sigma_i$ *such that* σ_i *is in the convex hull of* $\{\sigma_i^k : k \in K\}$ *and such that for each* σ_i^k *there are conjectures* $\hat{\sigma}_{-i}$ *and beliefs* $\hat{\mu}_i$ *(both of which can depend on* σ_i^k*), such that*

(i) $p(\theta_i) > 0$ *implies* $\hat{\mu}_i(\theta_i) > 0$,

and for any pair $\theta_i, \hat{a}_i$ *such that* $\hat{\mu}^i(\theta_i) \cdot \sigma_i(\hat{a}_i \mid \theta_i) > 0$ *both the following two conditions are satisfied*

(ii) $\hat{a}_i \in \arg\max_{a_i} \sum_{a_{-i},\theta_{-i}} u_i(a_i, a_{-i}, \theta_i, \theta_{-i}) \hat{\mu}^i(\theta_{-i} \mid \theta_i) \hat{\sigma}_{-i}(a_{-i} \mid \theta_{-i})$, *and*

(iii) $\sum_{\{a_{-i},\theta_{-i}:y_i(\hat{a}_i,a_{-i},\theta_i,\theta_{-i})=\bar{y}_i\}} \hat{\mu}^i(\theta_{-i} \mid \theta_i)\hat{\sigma}_{-i}(a_{-i} \mid \theta_{-i})$

$= \sum_{\{a_{-i},\theta_{-i}:y_i(\hat{a}_i,a_{-i},\theta_i,\theta_{-i})=\bar{y}_i\}} p(\theta_{-i} \mid \theta_i)\sigma_{-i}(a_{-i} \mid \theta_{-i}).$

This definition allows for player *i*'s beliefs and conjecture to depend on the strategy she plays, but as in the unitary definition it requires all types of player *i* to form their beliefs by updating from a common distribution $\hat{\mu}_i$. To see the difference this heterogeneity can make, consider a simplified version of example 7 where the player is always *Timid*, and where the distribution of Nature's move is such that it is optimal for the player to always play *In***.** Then there is no self-confirming equilibrium where the player randomizes, but there are heterogeneous self-confirming equilibria where some players play *In* and others stay *Out*.

Heterogeneous self-confirming equilibrium is appropriate when Nature's move is iid over time (case (c) above), since a given agent eventually receives many observations of the distribution of signals corresponding to each possible type θ_i in the support of *p*. However, if types are fixed once and for all, then each agent is only in the role of a single type, and there is no reason that beliefs across types should be consistent with updating from a common prior.[19] Therefore, instead of imposing that restriction, we allow each type θ_i to have any "interim belief" $\tilde{\mu}^{\theta_i}$ that is consistent with that type's observations. Similarly, when types are fixed, conjectures may depend on types. The following notion of type heterogeneous self-confirming equilibrium captures the idea that types are fixed initially, but that players are subsequently matched with opponents whose types have been drawn from *p*.

Definition: *A strategy profile* σ *is a* type heterogeneous self-confirming equilibrium *if for each player* i*, and for each* $\hat{a}_i$ *and* θ_i *such that* $p(\theta_i)\cdot\sigma_i(\hat{a}_i \mid \theta_i) > 0$ *there are*

[19] If no restrictions are imposed on the prior, then *any* collection of interim beliefs $(\tilde{\mu}^{\theta_i})_{\theta_i\in\Theta_i}$ can be generated from a prior μ_i by setting $\mu_i(\theta_{-i},\theta_i) = \mu_i(\theta_i)\tilde{\mu}_i(\theta_{-i})$ for some marginals $\mu_i(\theta_i)$, but the interim definition allows for each type of player *i* to think all types are independently distributed while also allowing different types to have different beliefs.

conjectures $\hat{\sigma}_{-i}$ *and interim beliefs* $\tilde{\mu}^{\theta_i}$ *(both of which can depend on* $\hat{a}_i$ *and* θ_i*), such that both the following conditions are satisfied*

(ii) $\hat{a}_i \in \arg\max_{a_i} \sum_{a_{-i},\theta_{-i}} u_i(a_i, a_{-i}, \theta_i, \theta_{-i}) \tilde{\mu}^{\theta_i}(\theta_{-i}) \hat{\sigma}_{-i}(a_{-i} \mid \theta_{-i})$,

(iii)

$$\sum_{\{a_{-i},\theta_{-i}: y_i(\hat{a}_i, a_{-i}, \theta_i, \theta_{-i}) = \bar{y}_i\}} \tilde{\mu}^{\theta_i}(\theta_{-i}) \hat{\sigma}_{-i}(a_{-i} \mid \theta_{-i})$$
$$= \sum_{\{a_{-i},\theta_{-i}: y_i(\hat{a}_i, a_{-i}, \theta_i, \theta_{-i}) = \bar{y}_i\}} p(\theta_{-i} \mid \theta_i) \sigma_{-i}(a_{-i} \mid \theta_{-i}).$$

Notice that the "full support" condition (i) is no longer needed, since we no longer derive the interim (that is, type-dependent) beliefs by updating from an "*ex-ante"* prior.

Returning to example 7, recall that "*In* when *Brave*, *Out* when *Timid*" is a self-confirming equilibrium, since the player can believe that Nature's type is correlated with her own. This is not a self-confirming equilibrium with independent beliefs, and it is not a heterogeneous self-confirming equilibrium with independent beliefs, but it is a type-heterogeneous self-confirming equilibrium with independent beliefs. The same distinction appears in a modified version of example 7 where "Nature" is replaced by a player 2 who has *L* as a dominant strategy, and where player 1 observes her own payoff but not the action of player 2. The only self-confirming equilibrium is for both types of player 1 to play *In*; this is also the only heterogeneous self-confirming equilibrium, but "*Brave In*, *Timid Out*" is a type-heterogeneous self-confirming equilibrium. [20]

We conclude with a final example that also uses independence to create a distinction between heterogeneous and type-heterogeneous self-confirming equilibrium.

Example 8: Independent heterogeneous Self-Confirming Equilibria versus Independent type-heterogeneous Self-Confirming Equilibria

Consider a variation of the one-player game of Example 7, where the payoffs for *Out* remain 0 and those for *In* are

---

[20] The difference between the versions of the example with "Nature" and "player 2" comes from the fact that player 1 can think Nature's move is correlated with her type, but player 1's conjecture about player 2 must correspond to a strategy for player 2, and since player 2 does not observe player 1's type, player 2's strategy cannot depend on it.

	L	*R*
Brave	2	-1
Timid	-1	2

Here both types can stay *Out* only if they disagree about Nature's move: *Brave* must believe *R* and *Timid* must believe *L*. Suppose in fact that players observe nothing, so that behavior depends only on priors. If the players' types are drawn anew each period and beliefs are restricted to be independent, then in any self-confirming equilibrium $\sigma(In) \geq \sigma(Out)$ since the beliefs corresponding to any σ^k must lead them to play In either when they are *Brave* or when they are *Timid* (or both). On the other hand, if players' types are drawn once and for all, they can stay *Out* forever (each type can have constant beliefs justifying *Out*). ■

Appendix

Since Example 6 involves dominant strategies, it is not very interesting from a game-theoretic perspective. The next, more complicated, example, due to Phil Reny, shows that dominant strategies are not required for the property that the Nash and self-confirming equilibria coincide even when actions are not observed.

Example A

In this game there are three states of Nature $\theta_0', \theta_0'', \theta_0'''$ and no types. There are two players, a row and a column player; each chooses between three actions *T, M, B*. Payoffs in each of the states is given in the table below.

	θ_0'		
	T	*M*	*B*
T	0, 1	0, ½	1, -1
M	5, 0	5, 5	½, 5
B	0, 0	0, 5	-1, 0

	θ_0''		
	T	*M*	*B*
T	0, -1	0, ½	-1, 1
M	5, 0	5, 5	½, 0
B	0, 0	0, 5	1, 0

	θ_0'''		
	T	*M*	*B*
T	0, 1	0, 5	1, 1
M	5, 0	5, 5	5, 0
B	0, 0	0, 5	1, 0

Beliefs about and the actual distribution of Nature's move are given below

	θ_0'	θ_0''	θ_0'''
μ_1	1-2ε	ε	ε
μ_2	ε	1-2ε	ε
μ	ε	ε	1-2ε

To analyze the game, note that if 2 plays *T* or *M* it is a strict best response for 1 to play *M*; if 1 plays *M* or *B* it is a strict best response for 2 to play *M*. Hence the relevant portion of the game involves 2 playing *B* or 1 playing *T*. Payoffs in these cases are summarized below.

	θ_0'	θ_0''	θ'
$u_1(T, B, \theta_0)=u_2(T, T, \theta_0)=$	1	-1	1
$u_1(M, B, \theta_0)=u_2(T, M, \theta_0)=$	½	½	5
$u_1(B, B, \theta_0)=u_2(T, B, \theta_0)=$	-1	1	1

With the given priors the pure strategy Nash equilibria are (M, M) and (T, B). The latter is not a Nash equilibrium with a common prior. If players observe payoffs, then there are two (pure strategy) self-confirming equilibria with beliefs μ_1, μ_2: one with the strategy profiles (T, B) and the other with (M, M).

References

Banerjee, Abhijit and Rohini Somanathan [2001] "A Simple Model of Voice," *Quarterly Journal of Economics*; 116 (1), 189-228.

Dekel, Eddie, Drew Fudenberg and David K. Levine [1999] "Payoff Information and Self-Confirming Equilibrium," *Journal of Economic Theory*; 89(2), 165-85.

Battigalli, Pierpaolo [1988] "Implementable Strategies, Prior Information and the Problem of Credibility in Extensive Games," *Rivista Internazionale di Scienze Economiche e Commerciali*; 35(8), 705-33.

Dekel, Eddie and Faruk Gul [1997] "Rationality and Knowledge in Game Theory," in *Advances in economics and econometrics: Theory and applications, Seventh World Congress, Volume 1, Econometric Society Monographs, no. 26,* David M. Kreps and Kenneth F. Wallis, (eds.). Cambridge, New York and Melbourne: Cambridge University Press, 87-172.

Fudenberg, Drew and David M. Kreps [1995] "Learning in Extensive-Form Games. I. Self-Confirming Equilibria," *Games and Economic Behavior*; 8(1), 20-55.

Fudenberg, Drew and David K. Levine, [1993] "Self-Confirming Equilibrium," *Econometrica*; 61(3), 523-45.

Kalai, Ehud and Ehud Lehrer [1993] "Rational Learning Leads to Nash Equilibrium," *Econometrica*; 61(5), 1019-45.

Piketty, Thomas [1995] "Social Mobility and Redistributive Politics" *Quarterly Journal of Economics*; 110(3), 551-84.

Rubinstein, Ariel and Asher Wolinsky [1994] "Rationalizable Conjectural Equilibrium: Between Nash and Rationalizability," *Games and Economic Behavior*; 6(2), 299-311.

Spector, David [2000] "Rational Debate and One-Dimensional Conflict," *Quarterly Journal of Economics*; 115(1), 181-200.