

Learning and Equilibrium Refinements

Drew Fudenberg

22.1 INTRODUCTION

The literature on learning in games interprets equilibrium strategy profiles as the long-run average behavior of agents who are selected at random to play the game. As suggested by Nash (1950), in normal-form games we expect that as the agents accumulate evidence about play of the game they will develop accurate beliefs, so that the stationary points of the process correspond to the Nash equilibria.¹ The definition of Nash equilibrium applies unchanged to games in extensive form, but the learning foundation for it does change, because in games with a nontrivial extensive form simply playing the game repeatedly may not lead agents to know how their opponents would respond to deviations that the agents have not tried. Thus there is no reason to expect learning by myopic agents to lead to Nash equilibrium in general games, as agents may not experiment enough to learn the consequences of deviating from the equilibrium path. Instead, learning is consistent with *self-confirming equilibrium* (Fudenberg and Levine, 1993a,b).

The focus here is on settings where the agents are patient, so they do have an incentive to experiment.² In this case, as shown by Fudenberg and Levine (1993a), here Nash's "mass action" interpretation of equilibrium is again valid. But extensive-form games typically have many equilibria, and not all of them seem equally plausible. An advantage of the learning approach is that some actions that are off-path according to the limiting equilibrium distribution are not counterfactual, but will actually be played by young agents as "experiments," so that equilibrium refinements can be derived from properties of optimal experimentation.

I thank Daniel Clark, Kevin He, and Giacomo Lanzani for helpful comments, and each of them as well as David Kreps and David K. Levine for all their work on our collaborations. National Science Foundation grant SES 1951056 provided financial support.

¹ See Kuhn et al. (1996) for more on Nash's "mass action" interpretation.

² See Fudenberg and Levine (2009, 2016) for a broader discussion of learning in games.

22.2 BACKGROUND

Suppose that there are large populations of agents who play a given game many times, with anonymous random matching, so that they have no incentive to try to alter their current opponents' play in future matches. New agents are initially uncertain about the prevailing strategies. They believe they face a fixed but unknown steady-state distribution of opponent strategies and use their observations of play (the realized terminal nodes in their matches) to update their beliefs. If agents only get to play the game a few times, they may be unable to learn very much. If they play many times but are myopic, they will learn the equilibrium path but may not learn the consequences of deviating from it. So we will consider agents who are both long-lived and patient.

To rule out convergence to non-Nash outcomes, it is enough that agents have correct beliefs about play at any "relevant" information set – information sets that can be reached if just one player deviates from the equilibrium path. Suppose that as time $t \rightarrow \infty$, play converges to a limit profile. Fudenberg and Kreps (1988) noted that if along this limit agents use strategies that assign each action a minimum probability that goes to 0 at rate $1/t$, these relevant information sets are reached infinitely often, because $\sum_{t=1}^{\infty} 1/t = \infty$. Thus agents will have arbitrarily large samples of the consequences of deviating from their limit play. From the law of large numbers, these samples converge to the true limit distribution, so correctly specified Bayesian agents will eventually have correct beliefs at all relevant information sets. Thus the limit strategy profile must be a Nash equilibrium if the limit strategies are best responses to the limit beliefs. This game has a unique subgame-perfect equilibrium: All players pass. But suppose that player 1 thinks player 2 will probably drop, so 1 wants to drop, and player 2 thinks 3 will drop, so 2 wants to drop. If players 1 and 2 pass with probability $1/t$, then 1 will learn that 2 indeed drops, which justifies 1's desire to drop. Because player 3's information set is only reached with probability $1/t^2$, and $\sum_{t=1}^{\infty} 1/t^2 < \infty$, player 2 will only get a finite sample of player 3's play. Thus if player 2 was initially pretty sure that player 3 typically drops, their sample may not be large enough to change their mind, which raises the question of how much experimentation agents will actually do. In addition, the assumed behavior is not consistent with expected utility maximization, as rational Bayesians only randomize when indifferent, and even when patient they may never play some actions.

In response, a series of papers has assumed that agents maximize their discounted expected utility and derived the agents' experimentation behavior from discounted dynamic programming. These papers focus on the *patiently stable steady states* where the agents are both long-lived (and thus have lots of data) and patient (and thus have a reason to experiment). More specifically, patient stability is defined as the iterated limit where first lifetimes grow long and then discount factors tend to 1. As I explain throughout this chapter, it is not yet known if the results extend to a general double limit; this is one of the many open questions in this literature.

22.3 MODEL

Here is a brief sketch of one version of the learning model that underlies this talk. The stage game is a finite extensive-form game with players i , terminal nodes z , pure strategies s_i , behavior strategies π_i , and utility functions $u_i(z)$.

There is a unit mass population in each player role i . Every period, each agent is anonymously matched with opponents from the other populations uniformly at random to play the stage game. The agents believe the system is in a steady state. They don't know the distribution of strategies in the population; all agents in each population i have the same prior μ_i over the space Π_{-i} of strategy profiles of the opponents. The agents' prior beliefs are *nondoctrinaire*, meaning that they can be represented by a continuous density function that is strictly positive at interior points.³ At the end of each play, agents observe the realized terminal node in their match and use Bayes rule to update their beliefs about the distribution of opponents' play.⁴

The agents have geometrically distributed lifetimes: At the end of every period, each agent in population i exits the system with probability $0 < 1 - \gamma \leq 1$, and a mass of newcomers is added to each population to replace the departing agents.⁵ Agents maximize their expected discounted utility with discount factor $\delta < 1$. The agents face dynamic programming problems, and these problems have deterministic optimal policies that map histories to strategies. For simplicity we assume that all agents in the same population use the same optimal policy. This does not mean that they all play in the same way, as agents with the same policy may meet different opponents and so have different histories and play different strategies.

The *state* of the system is the mass of agents with each possible history: new with null history, played once and saw z , played twice and saw (z, z') , and so forth. Because there is a continuum of agents, the random-matching process is deterministic, so the deterministic policies of the agents generate a *deterministic* map from states today to states tomorrow: The current state determines the current distribution of strategies used, agents are randomly matched and observe their outcomes, and some agents leave and are replaced. *Steady states* are fixed points of this map; each steady state generates a corresponding steady-state strategy profile. Steady states exist even though individual agents learn,

³ Thus no possible observation has probability 0, and it follows from Diaconis and Freedman (1990) that the agents' beliefs converge to the observed distribution at a rate that is uniform over all realizations of the data. Note that *nondoctrinaire* priors can go to 0 on the boundary of the probability simplex, as is the case for Dirichlet priors.

⁴ Clark et al. (2022) relaxes this to the assumption that agent i observes the element of their *terminal node partition* $\mathcal{P}_i$ that contains the realized terminal node of the game. Fudenberg and Kamada (2015, 2018) investigated the implications of commonly known terminal node partitions in an epistemic model without providing an explicit learning foundation.

⁵ This is the model used by Fudenberg and He (2018, 2020, 2021). Fudenberg and Levine (1993a, 2006) assumed agents have fixed finite lifetimes, but that difference is inessential. Clark and Fudenberg (2021) allowed for each population i to have a different γ_i (which is discussed in more detail later in this chapter).

because departing agents take their information with them. Note that even with patient agents the steady-state strategy profiles depend on the prior: If everyone enters the system assigning high probability to a particular strict equilibrium, that's what will occur.

22.4 WHAT WE KNOW SO FAR

When the agents are short-lived they have little chance to learn and simply play a best response to their priors. When agents are long-lived but impatient, they do learn the steady-state path of play but need not learn how opponents respond to deviations, so any self-confirming equilibrium in strategies that is not weakly dominated could arise. To obtain refinements of Nash equilibrium, the literature focuses on steady states where agents are both long-lived and patient, and more specifically on strategy profiles that arise in steady states in the limit where agents first become long-lived ($\gamma \rightarrow 1$) and then arbitrarily patient ($\delta \rightarrow 1$). These are the *patiently stable* strategy profiles. The order of limits corresponds to an environment where agents are long-lived relative to their effective discount factors. This implies that people spend most of their lives myopically responding to their current beliefs.

22.4.1 Patiently Stable States Must Be Nash Equilibria

As first shown by Fudenberg and Levine (1993a), patiently stable profiles must be Nash equilibria. To see why, note that if a strategy s^i is played a positive fraction of the time when agents live a long time (e.g., as lifetime $T \rightarrow \infty$ or survival probability $\gamma \rightarrow 1$), then it must be played by a positive fraction of the population a positive fraction of their lifetime, and most agents who have played s^i many times have approximately correct beliefs about its consequences. Moreover, in the iterated limit that defines patiently stable profiles, agents eventually stop experimenting and play a myopic best response to their beliefs. That patiently stable states must be Nash equilibria is shown by contradiction: If a steady state is not a Nash equilibrium, agents in some player role can gain by deviating. Their prior is nondoctrinaire, so it assigns positive probability to the opponent strategies that make deviation worthwhile. The combination of the law of large numbers and the Diaconis and Freedman (1990) results then imply that most of these agents will find it worthwhile to continue experimenting with the potentially improving deviation when they are sufficiently patient, which contradicts the fact that they must stop experimenting in finite time.

22.4.2 Patiently Stable States Need Not Satisfy Backward Induction

Specifically, Fudenberg and Levine (2006) showed that (Drop, Drop, Pass) is patiently stable in the game in Figure 22.1. If player 1's prior belief is that

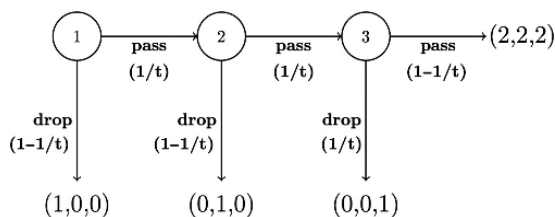


FIGURE 22.1 A three-player game.

player 2 probably plays drop, they will experiment with pass when young, and if player 2 does indeed mostly drop, then player 1 will decide that drop is optimal. Even if new player 2 agents initially believe that player 3 agents mostly drop, they may experiment with pass if they get a chance to play when they are very young. But when most player 1s are old, the player 2s will rarely get a chance to act. Once the player 2s learn this, they no longer view experimentation with pass as worthwhile. Fudenberg and Levine (2006) generalized this observation by identifying a class of games of perfect information where any *subgame confirmed* equilibrium is patiently stable: These are Nash equilibrium strategy profiles where in each subgame beginning one step off the path, the restriction of π to the subgame is self-confirming in that subgame.⁶ They argue that this may help explain the persistence of some superstitions, such as those apparently embodied in the second code of Hammurabi, which prescribes a “trial by river” for those accused of a crime, with the accuser being killed if the accused survives. This seems based on the superstition that the guilty are more likely to drown than the innocent, but if people will believe arbitrary superstitions, it would have been simpler to tell them that those who are guilty will be struck dead by lightning while the innocent will not be.

The difference between these two false beliefs is that young people can commit crimes as experiments and learn that the lightning strike superstition is false. In contrast, if older people believe the Hammurabi superstition, then when young players commit crimes, they are accused and punished so they correctly learn that (given the behavior of accusers) crime does not pay, and as they grow older they stop committing crimes. Accusers wrongly believe that the probability of the suspect drowning depends on whom they accuse. Since there are few crimes, accusers only get to play very infrequently, so it is not worthwhile experimenting with false accusations.

22.4.3 Learning in Signaling Games

In signaling games, the sender observes their type θ , which is drawn from a known prior distribution λ , and chooses a signal s . The receiver observes

⁶ More precisely, this has only been shown in equilibria where there is no randomization by the players and any random moves by nature only occur off-path.

the signal and chooses an action a without observing the sender's type. These games typically have many perfect Bayesian equilibria (PBE) because the Bayes rule doesn't pin down what the receiver believes about the sender's type at signals the equilibrium assigns probability 0. This has led to substantial interest in equilibrium refinements for signaling games, such as those proposed by Banks and Sobel (1987) and Cho and Kreps (1987). In the signaling game setting, learning theory partially validates the ideas behind traditional refinements and shows how they need to be modified.

Fudenberg and He (2018), Fudenberg and He (2020), and Clark and Fudenberg (2021) considered the implications of steady-state learning models where there are populations of size $\lambda(\theta)$ of each sender type.⁷ The arguments of Fudenberg and Levine (1993a) imply that patiently stable states must be Nash equilibria, and moreover no receiver agent will play a conditionally dominated response. However, different sender types can have different beliefs about receiver play, and aggregate sender play can be a mixture over undominated responses that would itself be dominated if played by a single agent. Thus the patiently stable states need not be perfect Bayesian equilibria, but they must be *perfect Bayesian equilibria with heterogeneous beliefs* (PBE-H).⁸ This conclusion, while too weak to be of much use, already illustrates a key implication of the learning perspective. Moreover, the learning models have additional implications.

Fudenberg and He (2018) and Fudenberg and He (2020) assumed that the sender agents believe the distributions of receiver responses to different signals are independent, so senders don't learn about the response to signal s from the response to other signals. This makes each sender's problem a multiarmed bandit whose solution is characterized by the Gittins index (Gittins, 1979). Moreover, different types of senders gain different amounts from experimenting with a given signal, which generates restrictions on what receivers observe, and hence on what the typical receiver believes when they see each signal. Thus belief restrictions can be derived instead of assumed.

Specifically, $\theta' \succsim_s \theta''$ (θ' is more compatible with s than θ'') if whenever s is a *weak* best response for θ'' against some receiver strategy, it is a *strict* best response for θ' against that strategy. It is easy to see that if $\theta' \succsim_s \theta''$, then in every perfect Bayesian equilibrium π^* , $\pi_1^*(s | \theta') \geq \pi_1^*(s | \theta'')$. Thus in a PBE-H if $\pi_1^*(s | \theta') > 0$ and $\theta' \succsim_s \theta''$, then seeing s cannot increase the odds ratio of the less compatible type: $p(\theta'' | s)/p(\theta' | s) \leq \lambda(\theta'')/\lambda(\theta)$. *Type compatible equilibrium* imposes this condition at off-path signals as well. Fudenberg and He (2018) showed that type compatibility is a consequence of patient stability by showing that at any history where a signal s has the highest Gittins index for

⁷ This “fixed types” model seems more appropriate for most applications than the alternative model where each agent's type is iid. See Dekel et al. (2004) for a discussion of the differences this can make, and Fudenberg and Vespa (2019) for an experiment that verifies that play in the two treatments varies in the way theory suggests.

⁸ Clark and Fudenberg (2021) showed there can be patiently stable profiles that are not PBE.

type θ , the signal is the unique Gittins maximizer for any more compatible type θ' . The conclusion then follows from a coupling argument.

Type-compatible equilibria satisfy the Cho and Kreps (1987) Intuitive Criterion (IC) when they are on-path strict; it is not known if this is true in general. Fudenberg and He (2020) developed a more demanding equilibrium by requiring that agents assign probability 1 to their opponents using undominated strategies and modifying the compatibility relation accordingly.⁹ The resulting *rationality-compatible equilibrium* (RCE) implies IC but is strictly stronger: The Intuitive Criterion does not impose any restrictions on the relative likelihood of two types after a signal that is not equilibrium dominated for either of them, but RCE can. Fudenberg and He (2020) also showed that a uniform version of RCE is sufficient for patient stability.

Clark and Fudenberg (2021) developed a still-stronger equilibrium refinement for signaling games.¹⁰ To do this they add explicit cheap-talk messages to the signaling games. This seems desirable on descriptive grounds, since cheap talk is available in many of the settings signaling games are intended to model and can generate different equilibrium outcomes. It also makes the informal “speeches” used by Cho and Kreps (1987) to motivate the Intuitive Criterion into actual moves in the game, so that learning theory can be used to help determine how various messages are interpreted. In contrast to Fudenberg and He (2018) and Fudenberg and He (2020), Clark and Fudenberg (2021) did not impose an independence assumption on the sender priors. Instead, Clark and Fudenberg (2021) assumed that the message space is sufficiently rich and that the receivers are “initially trusting,” which means their prior leads them to trust certain messages they have not previously observed to be lies. Finally, instead of assuming that the agents in each population have the same frequency of play, they assume that the typical sender experiences many more matches over their life than the typical receiver. This fits some signaling game settings where senders are institutions who interact with a very large number of individuals in the role of the receivers.¹¹

Given these assumptions, as we pass to the iterated limit, a signal s will only be sent by agents with “justified types” for that signal.¹² So when a sender playing s sends a message claiming to be a justified type, most receivers will believe

⁹ This restriction makes sense in signaling games. In general extensive-form games, though, agents may find it optimal to play strictly dominated strategies as experiments to learn about the consequences of their other strategies, so requiring prior beliefs to be supported on opponents’ undominated strategies could lead to agents observing play they had assigned probability 0.

¹⁰ Clark (2021) adapts this refinement and its learning foundation to mechanism design by an informed principal.

¹¹ The simplest way to model this is to have senders live longer than receivers, but the same results obtain when lifetimes are equal, senders play every period, and receivers play in a given period with probabilities less than 1.

¹² Given an equilibrium π and signal s , the set of *justified types* $\bar{\Theta}(s, \pi)$ is essentially the minimal set of types such that for any receiver response to s that gives some type in Θ a weakly higher payoff than they get in equilibrium, at least one type in $\bar{\Theta}(s, \pi)$ would get a strictly higher payoff.

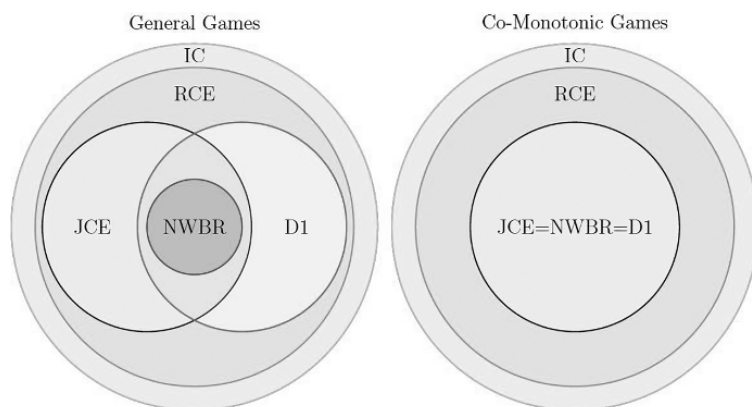


FIGURE 22.2 Restrictions of PBE-H.

this claim and choose their responses accordingly. *Justified communication equilibrium* (JCE) adds the corresponding restrictions to PBE-H.

As Figure 22.2 shows, IC is weaker than RCE, which is weaker than both JCE and the closely related $D1$ criterion of Banks and Sobel (1987), as well as the even stronger “NWBR in signaling games” introduced by Cho and Kreps (1987). Moreover, JCE, $D1$, and NWBR are identical (up to the difference between PBE and PBE-H) in “co-monotonic” signaling games such as the job market model proposed by Spence (1973).¹³

22.4.4 Patient Stability versus Classical Equilibrium Refinements

In simple games, patient stability does not imply backward induction.¹⁴ The implications of patient stability are much closer to those of classical refinements in signaling games. One part of the explanation is that signaling games have length 2, so the receivers have no experimentation incentive, and if the sender’s type were known, patient stability and backward induction would make the same predictions.

Moreover, in signaling games, the implications of patient stability have the flavor of forward induction as formalized in the requirement of invariance to the deletion of strategies that are not weakly optimal given equilibrium play.¹⁵

¹³ *Monotonic* signaling games (Cho and Sobel, 1990) require sender types have the same preference over receiver mixed best responses; co-monotonicity requires the same preference on the convex hull of those responses.

¹⁴ A classical explanation for backward induction is a common belief in rationality, as held by, for example, Reny (1993), Aumann (1995), Ben-Porath (1997), and Battigalli (1997). Learning theory suggests caution in predicting the backward induction solution when the common belief in rationality seems unlikely.

¹⁵ See, for example, Kohlberg and Mertens (1986), Hillas and Kohlberg (2002), and Govindan and Wilson (2009).

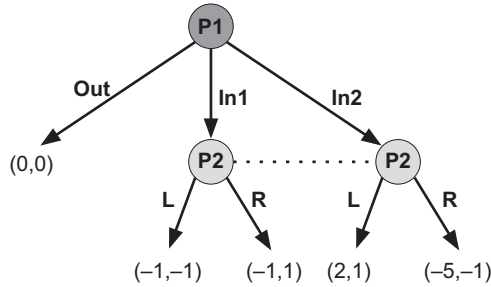


FIGURE 22.3 **In1** is strictly dominated by **Out** but provides the same information as **In2** and performs better than **In2** against some P2 strategies.

But in general games, patient stability is not even invariant to the weaker requirement of invariance to the deletion of weakly dominated strategies. The issue is that two actions that lead to the same information set provide the same information about play there, and this can lead agents to use strictly dominated actions as experiments.¹⁶ Figure 22.3 shows an example of this from Clark and colleagues (2022).

Here **In1** is strictly dominated by **Out** for P1, and (**In2**, L) is the only strategically stable equilibrium, but player 1's experimentation incentive makes (**Out**, R) patiently stable.

22.4.5 Invariance

The argument that a solution concept for games should only depend on the normal form is based on the claim that the differences between extensive forms with the same normal form are “irrelevant details” because they do not change the decision problem of a player who faces the *same fixed and known* strategies of the opponents. Because the normal form abstracts from many aspects of game play that are relevant for how people learn what strategies are used by others, there is no reason to expect learning to depend only on this very abstract representation of strategic interaction. Clark and colleagues (2022) show that the patiently stable profiles are only invariant to transformations that are both *decision invariant* (i.e., lead to the same best responses as a function of opponent strategies) and *information invariant* in the sense of providing the same feedback to the agents in their learning problems. Specifically, learning outcomes, unlike sequential equilibria, are invariant to the coalescing of consecutive moves by the same player. However, like sequential equilibria and unlike the various definitions of strategic stability, learning outcomes are not invariant to replacing an extensive form game with the corresponding game in

¹⁶ For this reason, it is not obvious what the implications of “rational play” are in a learning model, let alone how to modify Battigalli and Siniscalchi (2002)’s characterization of forward induction as the result of strong belief in rationality.

normal form: In the latter case there are no unreached information sets, and the terminal node that is reached reveals the strategy used by each player.

To capture what is essential for learning outcomes in the normal form, Clark and colleagues (2022) augment it with *terminal node partitions* and show that playing the extensive form game is equivalent to playing the normal form with the terminal node partitions that give players the same information as would be revealed by the terminal nodes in the extensive form, so that the two games are information invariant.

22.5 CONCLUSION

To conclude, here is a brief list of some open issues and future directions. On the technical side, the current results that refine Nash equilibria considered iterated limits where the agents who have experimentation motives become very long-lived before they become patient. It is not yet known if the conclusion extends to the general double limit, but Daniel Clark and I hope to show that it does. Also, some results only apply to equilibria without random play on the equilibrium path. There are no known counterexamples, so it may be possible to show this restriction is not needed using sharper bounds on the probability that agents observe outcomes that induce them to permanently switch to another strategy.

The analysis of patiently stable steady states raises the question of which steady states are dynamically stable. For static games there are local and global stability results for smooth fictitious play in models where all agents have the same age and thus the same amount of data (e.g., Fudenberg and Levine, 1993a; Benaïm and Hirsch, 1999; Hofbauer and Sandholm, 2002). The overlapping generation models used here permit the analysis of steady states, but perhaps models with a single cohort of infinite-lived agents would give some insight about dynamic stability.

Intuitively, we would expect agents to learn the path of play more quickly than they learn the consequences of deviating from it, with information coming in less and less frequently and information sets that are more steps off of the equilibrium path. Can this intuition be formalized, and can useful quantitative bounds be obtained? Clark and Fudenberg (2021) consider a setting where a natural difference in interaction rates between agents in two populations leads to strong and interesting results. Are there other such settings?

It would be interesting to extend the learning in games program to agents with misspecified (i.e., priors that rule out the true distribution of play).¹⁷ For example, agents might have incorrect beliefs about the extensive form of the game, as in analogy-based equilibrium (Jehiel, 2005). Or agents might know the possible actions of their opponents and the resulting payoffs but not know

¹⁷ Misspecified priors conflict with the ϕ -positivity assumption of Diaconis and Freedman (1990) but that result has recently been relaxed by Fudenberg et al. (2021). One potential complication is that a misspecified agent may perceive an option value for experimenting with an action they have already used many times, as shown by Fudenberg et al. (2017).

their opponents' information sets or whether the opponents have access to a mediator or correlating device. An extension to misspecified beliefs could lead to equilibrium refinements for "behavioral agents" who are overconfident in their own abilities, as in Heidhues and colleagues (2018), or who mistakenly perceive correlated variables as independent, as in, for example, Kroll and colleagues (1988) and Levy and colleagues (2018).

Maximizing expected discounted utility is a natural benchmark and starting point, but it would be interesting to extend the learning foundations program to consider heuristics like upper confidence bound policies and the other heuristics studied by Kaufmann et al. (2012) and Francetich and Kreps (2020a,b). It would also be natural to consider patient agents who maximize some form of ambiguity-averse preferences.¹⁸

Gonçalves (2020) studied a related model of learning to play a one-shot simultaneous-move game, where each agent lives a single period, and agents can buy random samples of past play at a cost. As the sampling cost decreases, the limits of the steady states converge to Nash equilibria in weakly undominated strategies. It would be interesting to extend this analysis to extensive-form games, but it is not clear how to introduce experimentation in this setting.

The learning in games literature has resulted in a greater understanding of which equilibria and refinements are likely to emerge in various classes of games such as signaling games. Hopefully this will lead to similar understanding in other classes of games. Of course, there are multiple ways to formulate models of nonequilibrium learning, just as there are many definitions of forward induction and several variants of the Kohlberg and Mertens (1986) axioms. However, it seems easier to judge the plausibility of learning models than that of axiomatic conditions on equilibrium concepts, especially axioms that are imposed without any reference to how equilibrium play might arise.

22.6 REFERENCES

- R. J. Aumann. Backward induction and common knowledge of rationality. *Games and Economic Behavior*, 8(1):6–19, 1995.
- J. S. Banks and J. Sobel. Equilibrium selection in signaling games. *Econometrica*, 55:647–661, 1987.
- P. Battigalli. On rationalizability in extensive games. *Journal of Economic Theory*, 74(1):40–61, 1997.
- P. Battigalli and M. Siniscalchi. Strong belief and forward induction reasoning. *Journal of Economic Theory*, 106(2):356–391, 2002.
- P. Battigalli, A. Francetich, G. Lanzani, and M. Marinacci. Learning and self-confirming long-run biases. *Journal of Economic Theory*, 183:740–785, 2019.
- E. Ben-Porath. Rationality, Nash equilibrium and backwards induction in perfect-information games. *Review of Economic Studies*, 64(1):23–46, 1997.

¹⁸ Battigalli et al. (2019) used a model of myopic ambiguity averse agents to provide a learning foundation for self-confirming equilibrium with model uncertainty.

- M. Benaïm and M. W. Hirsch. Mixed equilibria and dynamical systems arising from fictitious play in perturbed games. *Games and Economic Behavior*, 29(1–2):36–72, 1999.
- I.-K. Cho and D. M. Kreps. Signaling games and stable equilibria. *Quarterly Journal of Economics*, 102:179–221, 1987.
- I.-K. Cho and J. Sobel. Strategic stability and uniqueness in signaling games. *Journal of Economic Theory*, 50:381–413, 1990.
- D. Clark. The informed principal with agent moral hazard. Working paper, 2021.
- D. Clark and D. Fudenberg. Justified communication equilibrium. *American Economic Review*, 111(9):3004–3034, 2021.
- D. Clark, D. Fudenberg, and K. He. Observability, dominance, and induction in learning models. Has been published in *Journal of Economic Theory*, 206:105569, 2022.
- E. Dekel, D. Fudenberg, and D. K. Levine. Learning to play Bayesian games. *Games and Economic Behavior*, 46(2):282–303, 2004.
- P. Diaconis and D. Freedman. On the uniform consistency of Bayes estimates for multinomial probabilities. *Annals of Statistics*, 18:1317–1327, 1990.
- A. Francetich and D. Kreps. Choosing a good toolkit, I: Prior-free heuristics. *Journal of Economic Dynamics and Control*, 111:103813, 2020a.
- A. Francetich and D. Kreps. Choosing a good toolkit, II: Bayes-rule-based heuristics. *Journal of Economic Dynamics and Control*, 111:103814, 2020b.
- D. Fudenberg and K. He. Learning and type compatibility in signaling games. *Econometrica*, 86(4):1215–1255, 2018.
- D. Fudenberg and K. He. Payoff information and learning in signaling games. *Games and Economic Behavior*, 120:96–120, 2020.
- D. Fudenberg and K. He. Player-compatible learning and player-compatible equilibrium. *Journal of Economic Theory*, 194:105238, 2021.
- D. Fudenberg and Y. Kamada. Rationalizable partition-confirmed equilibrium. *Theoretical Economics*, 10(3):775–806, 2015.
- D. Fudenberg and Y. Kamada. Rationalizable partition-confirmed equilibrium with heterogeneous beliefs. *Games and Economic Behavior*, 109:364–381, 2018.
- D. Fudenberg and D. M. Kreps. A theory of learning, experimentation, and equilibrium in games. Working paper, 1988.
- D. Fudenberg and D. K. Levine. Steady state learning and Nash equilibrium. *Econometrica*, 61:547–573, 1993a.
- D. Fudenberg and D. K. Levine. Self-confirming equilibrium. *Econometrica*, 16(3):525–545, 1993b.
- D. Fudenberg and D. K. Levine. Superstition and rational learning. *American Economic Review*, 96:630–651, 2006.
- D. Fudenberg and D. K. Levine. Learning and equilibrium. *Annual Review of Economics*, 1(1):385–420, 2009.
- D. Fudenberg and D. K. Levine. Whither game theory? Towards a theory of learning in games. *Journal of Economic Perspectives*, 30(4):151–170, 2016.
- D. Fudenberg and E. Vespa. Learning theory and heterogeneous play in a signaling-game experiment. *American Economic Journal: Microeconomics*, 11(4):186–215, 2019.
- D. Fudenberg, G. Romanyuk, and P. Strack. Active learning with a misspecified prior. *Theoretical Economics*, 12(3):1155–1189, 2017.
- D. Fudenberg, G. Lanzani, and P. Strack. Limit points of endogenous misspecified learning. *Econometrica*, 89(3):1065–1098, 2021.

- J. C. Gittins. Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society: Series B*, 41(2):148–164, 1979.
- D. Gonçalves. Sequential sampling and equilibrium. *Job Market Paper*, 2020.
- S. Govindan and R. Wilson. On forward induction. *Econometrica*, 77(1):1–28, 2009.
- P. Heidhues, B. Kőszegi, and P. Strack. Unrealistic expectations and misguided learning. *Econometrica*, 86(4):1159–1214, 2018.
- J. Hillas and E. Kohlberg. Foundations of strategic equilibrium. In R. Aumann and S. Hart, eds., *Handbook of Game Theory with Economic Applications*, vol. 3, ch. 42, pp. 1597–1663. Amsterdam: Elsevier, 2002.
- J. Hofbauer and W. H. Sandholm. On the global convergence of stochastic fictitious play. *Econometrica*, 70(6):2265–2294, 2002.
- P. Jehiel. Analogy-based expectation equilibrium. *Journal of Economic Theory*, 123(2):81–104, 2005.
- E. Kaufmann, O. Cappé, and A. Garivier. On Bayesian upper confidence bounds for bandit problems. In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*, pp. 592–600, 2012.
- E. Kohlberg and J.-F. Mertens. On the strategic stability of equilibria. *Econometrica*, 54:1003–1037, 1986.
- Y. Kroll, H. Levy, and A. Rapoport. Experimental tests of the separation theorem and the capital asset pricing model. *American Economic Review*, 78(3):500–519, 1988.
- H. W. Kuhn, J. Harsanyi, R. Selten, J. Weibull, and E. van Damme. The work of John Nash in game theory. *Journal of Economic Theory*, 69(1):153–185, 1996.
- G. Levy, I. Moreno de Barreda, and R. Razin. Persuasion with correlation neglect: Media power via correlation of news content. Working paper, 2018.
- J. Nash. Non-cooperative games. PhD thesis, Mathematics Department, Princeton University, 1950.
- P. J. Reny. Common belief and the theory of games with perfect information. *Journal of Economic Theory*, 59(2):257–274, 1993.
- M. Spence. Job market signaling. *Quarterly Journal of Economics*, 87:355–374, 1973.

