Whither Game Theory? Towards a Theory of Learning in Games
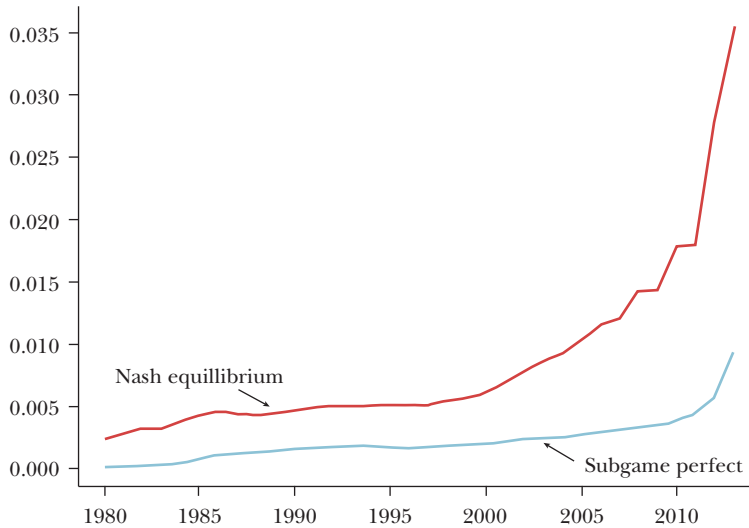
Drew Fudenberg and David K. Levine

When we were graduate students at MIT (1977–81), the only game theory mentioned in the first-year core was static Cournot (1838) oligopoly, although Eric Maskin taught an advanced elective on game theory and mechanism design. Just ten years later, game theory had invaded mainstream economic research and graduate education, and instructors had a choice of three graduate texts: Fudenberg and Tirole (1991), Osborne and Rubinstein (1994), and Myerson (1997).

Game theory has been successful in economics because it works empirically in many important circumstances. Many of the theoretical applications in the 1980s involved issues in industrial organization, like commitment and timing in patent races and preemptive investment. For example, Kreps and Wilson (1982) developed sequential equilibrium to analyze dynamic competition more effectively, Milgrom and Roberts (1982) showed how to apply game theory to limit pricing, and most of Tirole (1988)'s influential industrial organization textbook is applied game theory of one form or another. An array of other applications has followed, including problems ranging from auctions to market organization, and to monetary, fiscal, and international agreements involving trade and environmental issues. Many of these

■ *Drew Fudenberg is Paul Samuelson Professor of Economics, Massachusetts Institute of Technology, Cambridge, Massachusetts. David K. Levine holds a Joint Chair in the Department of Economics and Robert Schuman Center for Advanced Study, European University Institute, Florence, Italy. He is also John H. Biggs Distinguished Professor of Economics Emeritus, Washington University in St. Louis, St. Louis, Missouri. Their email addresses are drew.fudenberg@gmail.com and david@dklevine.com.*

Figure 1

Google Scholar Results for "Nash equilibrium" and "subgame perfect" Relative to Results for "economics" for 1980 to the Present

Source: Authors using Google Scholar.

theoretical applications have had successful empirical application: for example, a substantial fraction of the roughly 4,000 papers that cite the Berry, Levinsohn, and Pakes (1995) game-theoretic empirical study of oligopoly in automobile markets are themselves empirical game-theoretic studies. Moreover, feedback from theory to the laboratory and back has become an important way to refine and improve game theoretic predictions, as in the literatures on ultimatum-game experiments following Güth, Schmittberger, and Schwarze (1982), on inequity-averse preferences following Fehr and Schmidt (1999), and on the winner's curse following Kagel and Levin (1986).

Now that the basic tools of Nash equilibrium analysis are well understood and have been widely applied, work by abstract game theorists is no longer as tightly focused and has diffused into a number of themes. This hardly means that interest in or use of game theory has declined, as illustrated by Figure 1, which compares Google Scholar hits for "Nash equilibrium" and "subgame perfect" to those for "economics" from 1980 to the present. The role of game theory in economics does not seem to be in much doubt. Moreover, game theory continues to draw top young scholars into the field, as indicated by the recent award in 2016 of the John Bates Clark medal to Yuliy Sannikov, a game theorist.

There continue to be important theoretical developments in game theory: for example, over the last decade we have made tremendous progress in understanding the role of information in dynamic games. As theoretical understanding improves,

it is important also to improve our understanding of a wider range of field and lab settings. Theoretical research has helped explain the circumstances in which game theory does a good job empirically. It has also helped explain the circumstances in which game theory does a poor job empirically: for example, when equilibria are not robust, the environment is complex, or when circumstances are unfamiliar, standard game theory is less likely to perform well in the sense of making too wide a range of predictions (for example, Levine and Zheng 2015). Moreover, we know from laboratory studies—for example, those involving "beauty contest" games in which the players must estimate what everyone else is going to choose—that equilibrium theory does a poor job when people have inadequate time to learn about the game and the behavior of their opponents. As we have argued in the past, learning theory offers a way to improve and widen game theory's predictive power by shedding light on what sorts of equilibria to expect in various contexts (Fudenberg and Levine 1998; Fudenberg and Levine 2009). However, not enough is known about the speed of learning, an issue that we feel deserves much more attention.

Game Theory and Learning

Nash equilibrium has proven itself to be a remarkably powerful tool for understanding human interactions. As one commonplace example, the rush-hour traffic game is played on weekday mornings in major cities in the world. It lasts several hours. The millions of players are the commuters traveling by car from home to school or work. The actions are the many different routes that they can take: main highway, side streets, a North or a South route, and so forth. Commuters may also have some leeway over their departure times. To the first approximation, their preferences are to reach their destination as quickly as possible (although an expanded game could include additional preferences about outcomes like time of arrival and safety during the trip). Nash equilibrium holds when no commuter can save any time by taking a different route. For this equilibrium to be approximately true, for example, it must be true that if you get off the congested main highway and dodge through side streets you find that just enough other people are doing the same thing that you derive no advantage from doing so. One of us has considerable experience with a particular version of the rush-hour traffic game and, after extensive experimentation, has concluded that Nash equilibrium fits well. Less anecdotally, traffic engineers usually assume that traffic flow is well described by Nash equilibrium, although they call it "user equilibrium" (for example, Florian 1976).

If we believe that Nash equilibrium is applicable in many practical situations, as the large body of empirical work using it suggests, then it is natural to inquire as to why this should be so. For several reasons, most economists have come to think of Nash equilibrium and its variations as arising not from introspection and calculation, but rather from some nonequilibrium adaptive process of learning, imitation, or evolution. First, as the rush-hour traffic game makes clear, in many practical

instances, it seems infeasible to find Nash equilibrium by rationality alone: after all, commuters do not and could not choose routes by contemplating the rational play of all other commuters and working out how they should best respond. Instead, commuters approximate an optimal choice through by trial and error learning. Second, initial choices in laboratory games do not usually resemble Nash equilibrium (except in some special cases); instead, there is abundant experimental evidence that play in many games moves toward equilibrium as subjects play the game repeatedly and receive feedback (some classic citations include Smith 1962; Selten and Stoecker 1986; Van Huyck, Battallio, and Beil 1990, pp. 240–41). Finally, in many coordination problems with multiple equilibria, it is difficult to believe that reasoning from first principles could lead to Nash equilibrium. For example, the rush-hour traffic game can work equally well if the players all drive on the right or all on the left. But that coordination problem is not likely to be solved by introspection alone.

There are several reasons for our interest in the process by which Nash equilibrium is reached. First, "how we get there" may provide clues as to which of multiple equilibria are likely to be chosen. Second, as we shall discuss, there are cases where convergence to equilibrium does not occur quickly enough to be empirically observed. Third, even when a Nash equilibrium is reached and persists for a long time, regime changes may still occur. For example, in Sweden, the coordination equilibrium changed from driving on the left to driving on the right in 1967. We next offer two examples of significant regime change, one that took decades and one that took minutes.

Learning about Monetary Policy

Views about the tradeoff between inflation and unemployment have changed over time. The view of the Federal Reserve in the 1960s was summarized in this way by Ben Bernanke several decades later (as quoted in Domitrovic 2012): "[E]conomic theory and practice in the '50s and early '60s suggested that there was a permanent tradeoff between inflation and employment, the notion being that if we could just keep inflation a little bit above normal that we could get permanent increases in employment, permanent reductions in unemployment." This view is now widely regarded as wrong. It took the Fed four decades to figure this out, which we would describe as "slow" learning (for example, see Sargent 1999).

To help explain how it is possible to become stuck with a wrong theory, we use a simplified version of an idea from Sargent, Williams, and Zha (2006) that we described in Fudenberg and Levine (2009), which we will refer to as the Phillips curve game. There are two players: the Fed and a representative consumer. The Fed chooses a monetary policy, which we take to be either high or low inflation; the consumer observes the chosen policy and chooses either high or low unemployment. The Fed prefers low inflation, but is willing to choose high inflation if this leads to lower unemployment. Regardless of what inflation policy is chosen, the representative consumer always prefers low unemployment. For illustrative purposes we will suppose that the policymaker's payoff is the sum of an unemployment term and an

inflation term, and that the policymaker gets 2 for low unemployment, 0 for high unemployment, 1 for low inflation, and 0 for high inflation. The representative consumer gets 1 for low unemployment and 0 for high unemployment.

The Phillips curve game has two different types of Nash equilibria. It is an equilibrium for the representative consumer to always choose low unemployment and for the Fed to choose low inflation. However, it is also a Nash equilibrium for the consumer to follow a strategy of "respond to low inflation by choosing high unemployment and high inflation by choosing low unemployment, and for the Fed to choose high inflation in response. Why should the consumer choose high unemployment when we said low unemployment is preferred? The rationale behind this second outcome lies in a technicality in the definition of Nash equilibrium. Because the Fed is choosing *high inflation* in response to the consumer strategy, the way in which the representative consumer responds to low inflation is purely hypothetical—it is "off the equilibrium path." In effect, Nash equilibrium asks: would you choose high or low unemployment in response to low inflation? And the representative consumer answers: it does not really matter to me because I don't expect to see low inflation in my lifetime. And the representative consumer is correct as the Fed is choosing high inflation. As a technical matter Nash equilibrium allows this possibility. Game theorists have developed two responses.

1) Fix the technicality. The requirement of subgame perfect equilibrium strengthens Nash equilibrium by requiring that players should choose best responses to all contingencies, whether or not these contingencies are expected to occur. In the example, subgame perfection requires that the consumer makes the optimal choice in response to low inflation even if the consumer never expects to see low inflation. Nash equilibrium makes no such requirement. However, there are issues with subgame perfection. In more complex games than the example used here, subgame perfection requires that players have extensive knowledge about opponents' responses to hypothetical contingencies. For example, it can require that a player who has never seen opponents play a particular coordination subgame believe that the opponents will coordinate on one of the equilibria of that subgame and that the particular equilibrium that is coordinated on can be correctly forecast. Moreover, subgame perfection is often not robust to even small amounts of payoff uncertainty (Fudenberg, Kreps, and Levine 1988; Myerson 1978).

2) Take learning more seriously. An alternative story we can tell about the Phillips curve game is that the representative consumer always chooses low unemployment. However, the Fed chooses high inflation and incorrectly believes that if it were to choose low inflation, the representative consumer would choose high unemployment. If the Fed does not have detailed knowledge of consumer behavior and preferences, then when the Fed chooses high inflation it receives no feedback about the consequences of low inflation, and so has no basis on which to discover that its beliefs are incorrect. This leads to the concept of self-confirming equilibrium, which weakens Nash equilibrium by only requiring that player's beliefs about other player's strategies are consistent with what they observe when the game is played, and so allows players to have incorrect beliefs about how opponents would play.

The story then that explains the misguided Fed beliefs of the 1960s is that it was stuck in a non-Nash self-confirming equilibrium in which the consumers would choose low unemployment in response to low inflation but the Fed mistakenly thinks they would choose high employment instead. Believing that low inflation would lead to high unemployment, the Fed in the 1960s and into the 1970s chose high inflation, and it took time to learn that its belief was wrong. However, while self-confirming equilibrium may be a good short-run description, it may not be a good long-run description, because players may eventually get enough evidence about off-path play to correct their mistaken beliefs.

The Hijacking Game

We turn now to a case where Levine (2012) points out that learning was breathtakingly fast. In response to changed circumstances, an equilibrium that involved millions of air passengers and had lasted for decades changed to an opposite equilibrium, and the switch took place spontaneously and in less than half an hour.

We start with a game-theoretic model of air hijackings with two players: hijackers and passengers. There are two types of hijackers: mild and severe. The hijackers' type is private information, but the probabilities are common knowledge. The game proceeds sequentially. First, the hijackers decide to stay out or to hijack the plane. If they stay out, the game ends and everyone gets utility of 0. If the plane is hijacked, then the passengers must decide whether to fight or acquiesce. If the passengers fight, the hijackers are defeated with severe hijackers getting −1 and mild hijackers getting −2; because fighting generates a chance that the plane crashes, the passengers get an expected payoff of −2. If the passengers acquiesce, then the hijackers get +1 and the amount that the passengers get depends on the type of hijacker. If the hijackers are of the mild type, the passengers are eventually released and get −1. If the hijackers are of the severe type the passengers may be killed, and we set payoff of the passengers to be −3.

Suppose initially all hijackers are mild, so there is no private information. Then there are two Nash equilibrium outcomes: either the hijackers stay out or they enter and the passengers acquiesce. The latter is the unique subgame perfect equilibrium.

What if an exogenous change in circumstances drops the probability of mild hijackers to 25 percent? Then hijack/acquiesce is no longer a Nash or even a self-confirming equilibrium, and the only Nash equilibrium is for the hijackers to stay out and for the passengers to fight with at least 50 percent probability. Over time, the passengers learn that it is now better to fight, and when the hijackers in turn learn that passengers are choosing to fight, hijacking would diminish and we would reach the stay-out equilibrium. Formal learning models, as well as experimental studies involving unraveling in the finitely repeated prisoner's dilemma, suggest that this process would take some time. That, however, is not how it happened.

The 1990s saw roughly 18 aircraft hijackings a year. Flight crews were trained in the rational and government-approved "Common Strategy." Hijackers' demands should be complied with, the plane should be landed safely as soon as possible, and security forces should be allowed to handle the situation. Passengers should

sit quietly, and nobody should play hero. This advice was well established, rational, successful, and validated by decades of experience. This was the hijack/acquiesce equilibrium. Most hijackings ended peacefully, and the longer a hijacking persisted, the more often there was a peaceful ending.

Circumstances changed on September 11, 2001, when hijackers used the aircraft for suicide attacks on ground targets. The theory predicts that the equilibrium should shift from hijack/acquiesce to stay out/fight. This indeed has been the case: there have been very few hijackings since September 11, 2001, and in those very few cases the passengers have always resisted. But in this example, learning was extremely fast (National Commission on Terrorist Attacks on the United States 2004). American Airlines Flight 11 crashed into the North Tower of the World Trade Center at 8:42 am on September 11, 2011. Forty-two minutes later, United Airlines Flight 93 was hijacked. Only 29 minutes later, passengers and the flight crew on United Airlines Flight 93 assaulted their hijackers. A dramatic regime change happened on a plane already in the air based on limited information obtained through a few dozen telephone calls that relayed second-hand experience.

Of course, some ingredients are missing in the simple model. For example, one thing that was clear to the passengers, although it might not have been to a computer program implementing simple learning rules, that the other severe hijackings of the day made it virtually certain that their hijackers were of the severe type, too: after all, it would be a cosmic coincidence if severe hijackers had seized two planes and the third plane was seized by an independent group of mild hijackers. This highlights one problem with learning theory, namely that people are smart and sophisticated in ways it is hard to model.

Passive Learning

The basic concept of Nash equilibrium describes a situation where further learning about opponents' strategies is not needed and cannot occur. That is, since everyone is doing the best they can given the play of the others, nobody will ever discover a better action than the one they are using. As noted earlier, players in a game observe only what actually happens ("the equilibrium path") and not generally all the strategic choices of the other players.

However, in static simultaneous move games such as Cournot or Bertrand duopoly, the strategies are simply choices of actions. In this situation, observing the realized actions of others is enough for a player to determine what the outcome would have been had that player used a different strategy, so the question of "what would the opponent have done if I'd done something else" does not arise. To put it another way, all learning is "passive" in this situation, meaning that because what players do has no impact on what they see, players have no incentive to change their actions to gain additional information.

There are two main models of passive learning. The first, called *fictitious play*, was introduced by Brown (1951) and analyzed by Shapley (1964), Fudenberg and

Kreps (1993), and Monderer, Samet, and Selta (1997), among others. In the first period, players make an arbitrary choice; no data has been received, no learning has taken place. Subsequently, players keep track of the frequency with which their opponent has played different actions. It is straightforward to show that if the actions of both players converge, they must converge to a Nash equilibrium, and only a bit harder to show that the same is true if the empirical marginal distributions of actions converge to a pair of mixed strategies (Fudenberg and Kreps 1993). The exact rule for forming beliefs does not matter for this result; all that matters is that (asymptotically) players choose actions that are a best response to the empirical distribution of play. Economists tend to think of learning in terms of Bayes' law, and fictitious play can be interpreted in this way.

However, fictitious play raises two difficulties as a model of learning. First, fictitious play involves a deterministic best response based on the information collected, which can open a player up to exploitation by a clever opponent (Blackwell 1956; Fudenberg and Kreps 1993). Second, from a purely descriptive point of view, the exact best response of fictitious play implies that a small change in beliefs can lead to a discontinuous change in response probabilities, which seems implausible. Indeed, even when we see convergence to Nash equilibrium in experimental data, the play of individual players is quite noisy and is better described by a random response.

In response to these concerns, Fudenberg and Kreps (1993) replaced the exact best response of fictitious play with the assumption that payoffs are subject to independently distributed, privately observed payoff shocks. Introducing a degree of randomness means that the strategy is no longer deterministic and therefore is harder to exploit. More generally, this approach replaces the exact best response function of fictitious play with a smooth approximation to it, called "smooth fictitious play," which is both a better approximation of how players act and also has other analytical advantages. The resulting "Nash distribution"—the distribution over actions induced by the introduction of randomness—has over time become known as a "quantal response equilibrium" (McKelvey and Palfrey 1995).

The second class of passive learning models, *reinforcement learning models*, were drawn from the psychology literature and applied to learning in games by Roth and Erev (1995). These models do not deal with beliefs but rather directly update a measure of the utility of each action—called a "propensity"—and derive probabilities so that actions with higher propensities are more likely to be played. For certain models of choice and certain parameters for converting expected utility into probabilities of play, this model can be formally equivalent to smooth fictitious play.

In the original Roth and Erev (1995) formulation, utility weights are updated only for the action that was chosen. This is the reinforcement learning idea: the action chosen is reinforced according to how well it did. After many years of study of experimental data using variations of this model (for example, Cheung and Freedman 1997; Camerer and Ho 1999; Salmon 2001), Ho, Camerer, and Chong (2007) proposed "self-tuning experience-weighted attraction," which they report does a good job of fitting a variety of experimental data. In this model, weights are updated for every action that would have done at least as well as the action that was

chosen, although the utility of actions that are not used is depreciated. (Learners may depreciate the utility of unused actions by less when the data indicates a high degree of certainty about the environment, as in Fudenberg and Levine 2014 and Block, Fudenberg, and Levine 2016.) This method of updating has the property that if players are playing bad strategies they behave much as they would in fictitious play, while if they are playing a good strategy they play much as they would in a pure reinforcement learning model. Ho, Camerer, and Chong view this as a model of limited attention: people are more "likely to focus on strategies that would have given higher payoffs than ... [those] actually received, because these strategies present missed opportunities."[1]

The Cobweb and Recency

The discussion of passive learning models up to this point has assumed that current and past observations are weighted the same. *Recency* is the idea that recent observations might get more weight than older observations.
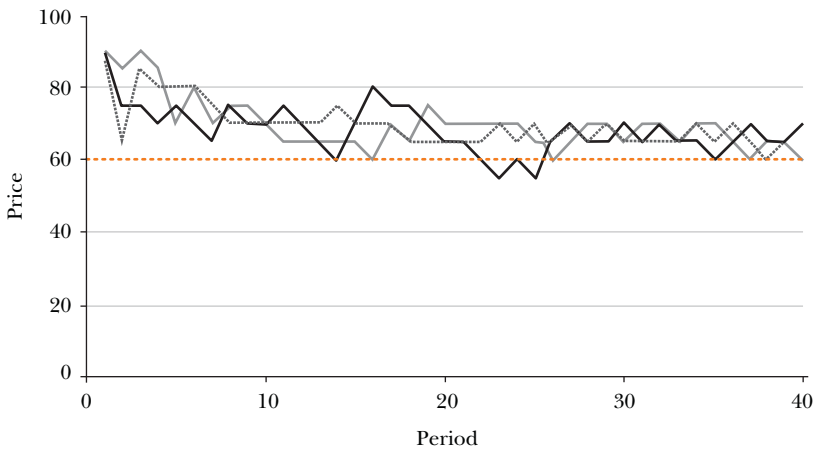
An extreme example of recency is to update using only data from the most recent period. Depending on the parameters, this can give rise to cycles, as in the famous "cobweb" first described in Kaldor (1934). Sutan and Willinger (2004) investigated this possibility. They studied a Cournot oligopoly market with five suppliers with increasing marginal costs, where the Cournot equilibrium is at a price of 65. The effects of recency can be demonstrated in such a setting by using the simplifying assumptions that players assume today's price will be the same as yesterday's and that they ignore their own market power. This leads to the conclusion that if producers expect a low price, they produce 0, which causes the price to rise to 100, so the next period there is overproduction and a low price, leading to 0 output the period after that, and so on and so on. However, the laboratory investigation of this Cournot oligopoly market found that experimental participants converged rather quickly to about 65, as shown in Figure 2.

This experiment (along with many others such as Hommes, Sonnemans, Tunisstra, and Van De Velden 2007) shows that extreme recency is not a good model. Interestingly our calculations show that if participants were to use an average of the prices in the last two periods, rather than just the last period, prices do converge, and the rate of convergence matches that in the experimental data rather well, as things settle down nicely after ten periods.

When and why should we expect people to give recent observations more weight? If the process that is generating observations undergoes unobserved regime shifts, then older observations may indeed be less informative than recent ones. Psychologists talk of "recency bias" because sometimes people do this even when they are told the environment they face is stationary (for example, Erev and Haruvy 2016).

---

[1] In the Ho, Camerer, and Chong (2007) approach, agents are homogeneous. Wilcox (2006) argues that agents are heterogeneous, and that assuming homogeneity biases the estimates of the learning parameters against belief learning and in favor of reinforcement learning.

Figure 2
Experimental Results from a Cournot Oligopoly Game

Notes: Each solid line shows results from a group of five players playing a Cournot oligopoly game. The competitive price in this market is at a price of 60. The Cournot equilibrium is at a price of 65.

There are two approaches to modeling recency. One is to develop explicit Bayesian models of changing environment. This approach is not widely used because models of this type are quite complex.[2] Instead, most models of recency have instead focused on simple rules of thumb. A starting point is to modify the fictitious play/reinforcement learning specification to specify that older observations receive exponentially less weight (as, for example, in the econometric specification of Cheung and Freedman 1997; see also the theoretical work of Benaïm, Hofbauer, and Hopkins 2009; Fudenberg and Levine 2014).

In lab experiments, a recency model often does well in describing some learning paths. But in general, the discount rate placed on recency varies from experiment to experiment, and may well vary from subject to subject. However, when subjects get stochastic feedback, recency can lead to a distribution of play that is very different than any Nash equilibrium. Fudenberg and Peysakovich (2014) provide evidence on this point: in their experiment, subjects face a "lemons" problem in which the computer chooses a value v uniformly distributed on the interval $[0,10]$ and the value of the object to the buyer is $v + k$, where k is a known parameter. The buyer makes a take-it-or-leave-it offer, which the computer accepts exactly when its value is below the offer. In one treatment (when $k = 3$) the Nash equilibrium is 3, while the actual mean bids converged to 5.18. However, in another treatment, the

---

[2] An example is hidden Markov models. Chen, Chen, and Levine (2015) develop techniques for simulating learning in models of this sort—but the algorithms used to solve the hidden Markov model cannot be used to generate analytic results.

Nash equilibrium bid is 6 while the actual mean bids converged to a bit less than that value. These outcomes can be matched by simulations with a high degree of recency, and in the data, even after some 20 rounds of play, participants appeared to place most of their weight on the most recent two observations.

Unfortunately, we do not yet have satisfactory models of recency. Here's an example of one of the complications. It may seem intuitive that if things have been stable for a while, but then seem to have changed, then you use greater recency. This insight is similar in spirit to work that assumes people act "as if" the world is stationary unless the data looks "too non-stationary" as in Fudenberg and Kreps (1994), Sargent (1999), and Cho and Kasa (2015). However, this intuition is incomplete. When we are surprised, we do not just put weight on more recent data, but we also re-evaluate our previous interpretations of old data. When the economy experienced an unusually severe recession as in 2007–2009, many analysts stopped looking for comparisons in post–World War II recessions and instead went back further to look for relevant lessons in the experience of the Great Depression.[3]

A thought experiment about the coordination game of driving on a certain side of the road helps to illustrate the difficulties. Imagine that you are driving on a deserted road and you encounter a driver who is on the wrong side. The next driver you encounter is also on the wrong side, and so is the next. Even so, you probably still would not conclude that the equilibrium had changed, but would instead contemplate the possibility that there was something unusual about this road at this particular time. The answer to the question of how many drivers in a row would you have to encounter before you concluded that the equilibrium had changed and all drivers everywhere were driving on the wrong side of the road is not obvious. Moreover, this example also highlights that real-world players are not operating in a vacuum. After encountering two or three wrong-side drivers in a row, you might well start listening to the radio or call someone on the phone, and ask if they had heard about anything funny going on.

From the Laboratory to the Field

There is a concern that the laboratory evidence on recency may not capture how real-world decisions are made. For example, Fudenberg and Peysakovitch (2014) find that when they give players a group of ten observations, the players display recency and the specific game does not converge to Nash equilibrium. However, if they give players a summary of those ten observations in the form of an

---

[3] One promising approach is the self-tuning experience-weighted attraction model of Ho, Camerer, and Chong (2007), which proposes a way to endogenize the "window" of time that players use in looking back. In this model, the window is taken to be the number of actions in the support of Nash equilibrium. The concept of the window seems an essential conceptual building block. However, it seems questionable to base a learning equilibrium on computations of Nash equilibrium by players. The approach is not consistent with evidence about recency, which sometimes seems to take into account longer or shorter windows. It is unclear how the idea of a window takes the re-evaluation of older experience into account. Also, it is unclear how this approach is to be applied when there is more than one Nash equilibrium.

average, then the players no longer exhibit recency, and the game does converge to Nash equilibrium.

Are real-world decision makers more likely to be seeing a series of data or some summary of that data? Federal Reserve decisions, for example, are taken by a committee that looks a bit at specific data from last week or last month, but mostly uses sophisticated econometric models using data going back a long way. Investment banks do sophisticated analysis of the term structure of interest rates and look for arbitrage opportunities using historical data series. Some part of how people learn to play games should be conceived of as choosing between model *A*, model *B*, model *C*, and others because people learn about which model should be used in addition to learning the parameters of a specific model.

Active Learning

Learning about Off-Path Play

In some settings, a passive learning approach will not gather information about the outcomes of alternative strategies because much of the game is "off the equilibrium path." Yet players must still infer causality: if I were to cooperate, would my opponent reward me? If so, for how long? For this reason, players in a dynamic game may choose to experiment with actions that they think might prove to be suboptimal, just to learn more about their consequences.

This perspective sheds additional light on the Sargent, Williams, and Zha (2006) analysis of the Federal Reserve and the Phillips curve game discussed earlier. If the game is fully understood by all players, then backward induction leads to the low inflation/low unemployment outcome. However, there is a self-confirming equilibrium in which the policymaker chooses high inflation due to a mistaken belief that low inflation leads to high unemployment. Sargent, Williams, and Zha argue that self-confirming equilibrium cannot adequately explain either the accelerating inflation of the 1970s nor the dramatic fall in inflation in the 1980s. They provide a more detailed model of active Bayesian learning that takes into account that some relevant data is revealed even in the high inflation regime, and argue that this learning model can explain many details of US monetary policy and inflation during 1970s and 1980s.

In incorporating learning about off-path play into learning models, several key issues must be addressed. First, the patience of the players matters—that is, time preference or discounting—because a patient player will be more willing to risk short-term mistakes in pursuit of better performance in the long run. Second, there seems to be a role here for random play. Remember that even in the passive models of fictitious play, random play helped in avoiding a situation where strategy choices became predictably exploitable by other players or unrealistically discontinuous. Here, randomness can also serve as a mechanism for learning about off-path play. Third, if the potential risks from experimentation are large and negative, then less of it will occur. Finally, some games may include many information sets, and thus it

would potentially require a lot of experimentation to figure out what was going on. A crucial case in point is that of repeated games, to which we turn our attention next.

Cooperation in Infinitely Repeated Games

The relations of consumers and businesses usually involve an important element of repetition, as do employment relationships, family relationships, and others. When games are repeated over time, the possibility of creating incentives through rewards and punishments arises. However, learning in these games is complicated by the need to infer causality "off the equilibrium path." To focus thoughts, we begin with the classic repeated prisoner's dilemma game. In each period, two players play each other in a *stage game* and decide whether to cooperate *C* or defect *D*. Typical payoffs are given by the payoff matrix:

	C	*D*
C	2,2	0,3
D	3,0	1,1

Each player has the defect choice *D* as the unique dominant strategy when the game is played only once. However, the infinitely repeated case, which can be considered as an idealized representation of settings without a commonly known ending date, is quite different. Consider the "grim trigger strategy" of cooperating in the first period and up until the other player defects, and then defecting forever afterwards. In this situation, if players do not have too high a discount rate on future payoffs, they will find it optimal to cooperate. **This is a special case of the celebrated** "Folk Theorem" as developed by Auman and Shapley (1992), Friedman (1971), Fudenberg and Maskin (1986), and Fudenberg, Levine, and Maskin (1994), among others: any payoff vector that is individually rational for the players is a subgame perfect equilibrium provided that the players are sufficiently patient (and assuming some technical conditions are satisfied). Although repeated games allow the possibility of cooperation in which incentives are established through future rewards, they allow many other possibilities as well.

What predictions should we make about repeated games with patient players? A common assumption in applied theory and theoretical industrial organization is that people cooperate whenever there is a cooperative equilibrium. Moreover, that prediction can be derived from various evolutionary game theory models, as in Axelrod and Hamilton (1981), Fudenberg and Maskin (1990), Binmore and Samuelson (1992), and Dal Bó and Pujals (2015). Unfortunately, there is little hard empirical support for this prediction, and laboratory evidence leans against it. Moreover, the laboratory provides a way to better understand what does happen in repeated games, and suggests that learning plays a key role.

Dal Bó (2005) relaunched the experimental study of repeated games by having participants play 7–10 iterations of the repeated game, with a different partner each time, as opposed to past work such as Murnighan and Roth (1983) in which

each subject played just once. There have now been a great many experimental studies of infinitely repeated games; there are 11 of them with 28 conditions in the meta-analysis of Dal Bó and Fréchette (2015). One main takeaway is that the discount factor matters much more once participants have played the game a few times. Another is that laboratory participants do respond to game parameters such as continuation probability and the payoff matrix in the direction suggested by theory, with the tendency to cooperate increasing in the gains from doing so, but it is not the case that players always succeed in cooperating whenever there is an equilibrium that supports cooperation.

Yet another take-away is that in contrast to predictions based on equilibrium analysis, the loss a cooperator incurs when the other player defects does matter. That is, whether an equilibrium exists in which players cooperate every period is theoretically independent of the loss a player incurs when that player cooperates and the other defects, because in an equilibrium where players always cooperate, no player ever expects that his opponent would be the first to play the defect choice D. In contrast, the loss that arises when one player cooperates and the other defects matters in practice because participants cannot be sure what strategies their partners are using, and even in treatments where most participants eventually cooperate, some of them do play the defect choice D. To try to capture the effect of this "strategic uncertainty," Blonski, Ockenfels, and Spagnolo (2011) look at risk dominance in an associated game where the only strategies are "Tit for Tat," which is repeating the choice just made by the other player, and "Always Defect" (Axelrod and Hamilton 1981). The payoff matrix is:

	Tit-for-Tat	Always Defect
Tit-for-Tat	2, 2	$\delta, 3(1 - \delta) + \delta$
Always Defect	$3(1 - \delta) + \delta, \delta$	1, 1

They propose that players will only cooperate if the risk of a Tit-for-Tat choice dominates Always Defect choice, which in our example requires $\delta \geq 2/3$, so that for $2/3 > \delta \geq 1/2$, they would predict little or no cooperation. This criterion does a fairly good job of matching the data in the Dal Bó and Fréchette (2015) meta-study. To get more precise predictions, Dal Bó and Fréchette relate cooperation to the size of the "basin" of Tit-for-Tat, which is the probability of the opponent playing Always Defect that leaves the player just indifferent between Tit-for-Tat and Always Defect. However, even in "cooperative treatments"—where the payoff matrix and discount factor should lead to a high rate of cooperation—Dal Bó and Fréchette estimate that about 10–20 percent of the participants use Always Defect. There are similarly large shares of Always Defect in the "cooperative treatments" of the noisy prisoner's dilemma in Fudenberg, Rand, and Dreber (2012), who also compute the payoffs of the estimated strategies against the estimated distribution and find that the Always Defect players do substantially worse than the conditional cooperators who use Tit-for-Tat. This heterogeneity of play does not seem to reflect social preferences,

and is uncorrelated with how the same participants play in a dictator game (Dreber, Fudenberg, and Rand 2014), so we interpret it as showing that people find it hard to learn which strategies will perform well, both because of the size of their own strategy space and because of the many possible strategies their opponents might be using.

What Constitutes a Good Theory?

Game theory has become a basic working tool for economists—and not just for theoreticians, but for empirical investigation. So where do we go from here? We have argued here that improved models of learning can lead to a better theory. To see why, let us review what we would like to see in a good theory.

First, we would like precise and valid predictions. Our impression is that existing game theoretic methods often do pretty well on validity, at least when researchers use appropriately robust versions of the theory. However, precision means that the theory generates precise statements about what will happen: for example, specifying which of several possible equilibria will occur. In some settings, for example in repeated games, game theory does not do well on this score. In other settings, equilibrium models have proven useful in structural estimation; see Bajari, Hong, and Nekipelov (2013) for a survey of recent work. In settings where equilibrium theory does not yield precise predictions, theoretical analysis of learning models has a great deal of promise both in explaining which equilibria are more likely to be observed and how initial conditions may matter. The extensive literature on evolutionary games with random shocks is a step in this direction. Early work both in discrete time models (Kandori, Mailath and Rob 1993; Young 1993) and in continuous time (Foster and Young 1990; Fudenberg and Harris 1992) gave conditions under which risk dominant equilibria are likely to be observed, and subsequent work such as Johnson, Levine, and Pesendorfer (2001) extended the idea that random shocks generate equilibrium selection to the case of dynamic games. These results are likely to expand, as there has been a recent resurgence of interest both in traditional evolutionary models, such as Levine and Modica (2016) and in fast learning, such as Kriendler and Young (2013) and Ellison, Fudenberg, and Imhof (2016). Applications of some of these recent results can be found in Block, Fudenberg, and Levine (2016).

Another way to use learning models to generate better predictions is by the use of simulations. For example, Brandts and Holt (1996) use this approach to argue that a form of myopic fictitious play fits experimental play of signaling games; Dal Bó and Fréchette (2015) used simulations of a belief-based learning model to predict how play in their repeated game experiment would have changed if subjects had played more games; and Fudenberg and Peysakovich (2014) argued that simulations of a model of learning with recency fit their data better than either Nash or the "cursed equilibrium" of Eyster and Rabin (2005). More generally, while theory is useful for building insight and intuition into how models work, and crucial for understanding the generality and robustness of various conclusions, simulations

have proven to be useful for generating conjectures, and can be essential for developing quantitative results.

Second, we need theoretical simplicity. As time goes on, game theory may be able to deal with more sophisticated models of human behavior, but it is likely always to be the case that simple models will be helpful for providing understanding and intuition about how more elaborate models work. Of course, theories of learning are more complicated than static theories of equilibrium. However, existing models such as variations on fictitious play are quite tractable. We may hope that the analysis remains tractable after adding important features such as events that trigger re-evaluation of models.

Third, we need breadth. A single unified theory is subject to a great deal of possible testing and falsification. On the other hand, a theory specifically tailored to explain only one fact cannot be tested at all. Game theory, both static and the theory of learning dynamics, fares well on this score.

Finally, we would argue for the methodological value of conservatism. There is no point in introducing, say, a new theory of social preference and fairness that explains exactly one experiment and is inconsistent with many other facts for which we already have common and coherent explanations. The theory of Nash equilibrium has proven useful and accurate. Adding learning theory to the mix preserves the basic results and insights of Nash equilibrium, while also adding greater precision to prediction in many cases and offering new predictions about the speed of learning.

References

Aumann, Robert J., and Lloyd S. Shapley. 1992. "Long Term Competition: A Game Theoretic Analysis." http://www.econ.ucla.edu/workingpapers/wp676.pdf.

Axelrod, Robert, and William D. Hamilton. 1981. "The Evolution of Cooperation." *Science*, March 27, 211(4489): 1390–96.

Bajari, Patrick, Han Hong, and Denis Nekipelov. 2013. "Game Theory and Econometrics: A Survey of Some Recent Research." Chap. 1 in *Advances in Economics and Econometrics: Tenth World Congress*, vol. 3, edited by Daron Acemoglu, Manuel Arellano, and Eddie Dekel. Cambridge University Press.

Benaïm, Michel, Josef Hofbauer, and Ed Hopkins. 2009. "Learning in Games with Unstable Equilibria." *Journal of Economic Theory* 144(4): 1694–1709.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market

Equilibrium." *Econometrica* 63(4): 841–90.

Binmore, Kenneth G., and Larry Samuelson. 1992. "Evolutionary Stability in Repeated Games Played by Finite Automata." *Journal of Economic Theory* 57(2): 278–305.

Blackwell, David. 1956. "An Analog of the Minimax Theorem for Vector Payoffs." *Pacific Journal of Mathematics* 6(1): 1–8.

Block, Juan, Drew Fudenberg, and David K. Levin. 2016. "Social Comparisons Leading to Nash Equilibrium." Unpublished paper.

Blonski, Matthias, Peter Ockenfels, and Giancarlo Spagnolo. 2011. "Equilibrium Selection in the Repeated Prisoner's Dilemma: Axiomatic Approach and Experimental Evidence." *American Economic Journal: Microeconomics* 3(3): 164–92.

Brandts, Jordi, and Charles A. Holt. 1996. "Naïve Bayesian Learning and Adjustment to Equilibrium in Signaling Games." *Journal of Economic Behavior and Organization* 3(3): 164–92.

Brown, George W. 1951. "Iterative Solution of Games by Fictitious Play." Chap. 24 in *Activity Analysis of Production and Allocation,* edited by Tjalling C. Koopmans in cooperation with Armen Alchaian et al. John Wiley & Sons.

Camerer, Colin, and Teck-Hua Ho. 1999. "Experienced-Weighted Attraction Learning in Normal Form Games." *Econometrica* 67(4): 827–74.

Cason, Timothy, Daniel Friedman, and Florian Wagener. 2005. "The Dynamics of Price Dispersion or Edgeworth Variations." *Journal of Economic Dynamics and Control* 29(4): 801–22.

Chen, Wenlin, Yixen Chen, and David K. Levine. 2015. "A Unifying Learning Framework for Building Artificial Game-Playing Agents." *Annals of Mathematics and Artificial Intelligence* 73(3): 335–58.

Cheung, Yin-Wong, and Daniel Friedman. 1997. "Individual Learning in Normal Form Games: Some Laboratory Results." *Games and Economic Behavior* 19(1): 46–76.

Cho, In-Koo, and Kenneth Kasa. 2015. "Learning and Model Validation." *Review of Economic Studies* 82(1): 45–82.

Cournot, Augustin. 1838. *Recherches sur les Principes Mathematiques de la Theorie de la Richesses.*

Dal Bó, Pedro. 2005. "Cooperation under the Shadow of the Future: Experimental Evidence from Infinitely Repeated Games." *American Economic Review* 95(5): 1591–1604.

Dal Bó, Pedro, and Guillaume R. Fréchette. 2011. "The Evolution of Cooperation in Infinitely Repeated Games: Experimental Evidence." *American Economic Review* 101(1): 411–29.

Dal Bó, Pedro, and Guillaume R. Fréchette. 2015. "Strategy Choice in the Infinitely Repeated Prisoners' Dilemma." http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2292390.

Dal Bó, Pedro, and Enrique R. Pujals. 2015. "The Evolutionary Robustness of Forgiveness and Cooperation." Unpublished paper.

Domitrovic, Brian. 2012. "Chairman Bernanke Buries the Phillips Curve: Bravo!" *Forbes*, April 3.

Dreber, Anna, Drew Fudenberg, and David G. Rand. 2014. "Who Cooperates in Repeated Games: The Role of Altruism, Inequity Aversion, and Demographics." *Journal of Economic Behavior and Organization* 98: 41–55.

Ellison, Glenn, Drew Fudenberg, and Lorens A. Imhof. 2016. "Fast Convergence in Evolutionary Models: A Lyapunov Approach." *Journal of Economic Theory* 161: 1–36.

Erev, Ido, and Ernan Haruvy. 2016. "Learning and the Economics of Small Decisions." Chap. 10 in *Handbook of Experimental Economics*, vol. 2, edited by John H. Kagel and Alvin E. Roth. Princeton University Press.

Eyster, Erik, and Matthew Rabin. 2005. "Cursed Equilibrium." *Econometrica* 73(5): 1623–72.

Fehr, Ernst, and Klaus M. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics* 114(3): 817–68.

Florian, Michael A., ed. 1976. *Traffic Equilibrium Methods.* Lecture Notes in Economics and Mathematical Systems, vol. 118. Proceedings of the International Symposium held at the Université de Montréal, November 21–23, 1974. Springer-Verlag.

Foster, Dean, and Peyton Young. 1990. "Stochastic Evolutionary Game Dynamics." *Theoretical Population Biology* 38: 219–32.

Friedman, James W. 1971. "A Non-Cooperative Equilibrium for Supergames." *Review of Economic Studies* 38(1): 1–12.

Fudenberg, Drew, and Christopher Harris. 1992. "Evolutionary Dynamics with Aggregate Shocks." *Journal of Economic Theory* 57(2): 420–41.

Fudenberg, Drew, and David M. Kreps. 1993. "Learning Mixed Equilibria." *Games and Economic Behavior* 5(3): 320–67.

Fudenberg, Drew, and David M. Kreps. 1994. "Learning in Extensive Form Games, II: Experimentation and Nash Equilibrium." http://scholar.harvard.edu/files/fudenberg/files/learning_in_extensive-form_games_ii.pdf.

Fudenberg, Drew, David M. Kreps, and David K. Levine. 1988. "On the Robustness of Equilibrium Refinements." *Journal of Economic Theory* 44(2): 354–80.

Fudenberg, Drew, and David K. Levine. 1998. *Theory of Learning in Games*. MIT Press.

Fudenberg, Drew, and David K. Levine. 2009. "Learning and Equilibrium." *Annual Review of Economics* 1: 385–420.

Fudenberg, Drew, and David K. Levine. 2014. "Recency, Consistent Learning, and Nash

Equilibrium." *Proceedings of the National Academy of Sciences* 111(Supplement 3): 10826–29.

Fudenberg, Drew, and David K. Levine. 2016. "Whither Game Theory?" http://www.dklevine.com/archive/refs4786969000000001307.pdf.

Fudenberg, Drew, David Levine, and Eric Maskin. 1994. "The Folk Theorem with Imperfect Public Information." *Econometrica* 62(5): 997–1039.

Fudenberg, Drew, and Eric Maskin. 1986. "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information." *Econometrica* 54(3): 533–54.

Fudenberg, Drew, and Eric Maskin. 1990. "Evolution and Cooperation in Noisy Repeated Games." *American Economic Review* 80(2): 274–79.

Fudenberg, Drew, and Alexander Peysakhovich. 2014. "Recency, Records and Recaps: Learning and Non-equilibrium Behavior in a Simple Decision Problem." *Proceedings of the Fifteenth ACM Conference on Economics and Computation.*

Fudenberg, Drew, David G. Rand, and Anna Dreber. 2012. "Slow to Anger and Fast to Forgive: Cooperation in an Uncertain World." *American Economic Review* 102(2): 720–49.

Fudenberg, Drew, and Jean Tirole. 1991. *Game Theory.* MIT Press.

Güth, Werner, Rolf Schmittberger, and Bernd Schwarze. 1982. "An Experimental Analysis of Ultimatum Bargaining." *Journal of Economic Behavior and Organization* 3(4): 367–88.

Ho, Teck H., Colin F. Camerer, and Juin-Kuan Chong. 2007. "Self-tuning Experience Weighted Attraction Learning in Games." *Journal of Economic Theory* 133(1): 177–98.

Hommes, Cars, Joep Sonnemans, Jan Tuinstra, and Henk Van De Velden. 2007. "Learning in Cobweb Experiments." *Macroeconomic Dynamics* 11(S1): 8–33.

Johnson, Philip, David K. Levine, and Wolfgang Pesendorfer. 2001. "Evolution and Information in a Gift-Giving Game." *Journal of Economic Theory* 100(1): 1–21.

Kagel, John H., and Dan Levin. 1986. "The Winner's Curse and Public Information in Common Value Auctions." *American Economic Review* 76(5): 894–920.

Kaldor, Nicholas. 1934. "A Classificatory Note on the Determination of Equilibrium." *Review of Economic Studies* 1: 122–36.

Kandori, Michihiro, George J. Mailath, and Rafael Rob. 1993. "Learning, Mutation, and Long Run Equilibria in Games." *Econometrica* 61(1): 29–56.

Kreindler, Gabriel E., and H. Peyton Young. 2013. "Fast Convergence in Evolutionary Equilibrium Selection." *Games and Economic Behavior* 80: 39–67.

Kreps, David M., and Robert Wilson. 1982. "Sequential Equilibria." *Econometrica* 50(4): 863–94.

Levine, David K. 2012. *Is Behavioral Economics Doomed? The Ordinary versus the Extraordinary.* Open-book Publishers.

Levine, David K., and Salvatore Modica. 2016. "Dynamics in Stochastic Evolutionary Models." *Theoretical Economics* 11(1): 89–131.

Levine, David K., and Jie Zheng. 2015. "The Relationship of Economic Theory to Experiments." Chap. 2 in *Handbook of Experimental Economic Methodology,* edited by Guillame Fréchette and Andrew Schotter. Oxford University Press.

McDermott, Joseph M. 1967. "Operational Effects of Automatic Ramp Control on Network Traffic." *Highway Research Record* 202.

McFadden, Daniel. 1973. "Conditional Logit Analysis of Qualitative Choice Behavior." *Frontiers in Econometrics,* edited by P. Zarembka, 105–142. Academic Press.

McKelvey, Richard D., and Thomas R. Palfrey. 1995. "Quantal Response Equilibria for Normal Form Games." *Games and Economic Behavior* 10(1): 6–38.

Milgrom, Paul, and John Roberts. 1982. "Predation, Reputation, and Entry Deterrence." *Journal of Economic Theory* 27(2): 280–312.

Monderer, Dov, Dov Samet, and Aner Sela. 1997. "Belief Affirming in Learning Processes." *Journal of Economic Theory* 73(2): 438–52.

Murnighan, J. Keith, and Alvin E. Roth. 1983. "Expecting Continued Play in Prisoner's Dilemma Games: A Test of Several Models." *Journal of Conflict Resolution* 27(2): 279–300.

Myerson, Roger B. 1978. "Refinements of the Nash Equilibrium Concept." *International Journal of Game Theory* 7(2): 73–80.

Myerson, Roger B. 1997. *Game Theory: Analysis of Conflict.* Harvard University Press.

National Commission on Terrorist Attacks on the United States. 2004. *9/11 Commission Report.* http://www.9-11commission.gov/report/.

Osborne, Martin J., and Ariel Rubinstein. 1994. *A Course in Game Theory.* MIT Press.

Roth, Alvin E., and Ido Erev. 1995. "Learning in Extensive-form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term." *Games and Economic Behavior* 8(1): 164–212.

Salmon, Timothy C. 2001. "An Evaluation of Econometric Models of Adaptive Learning." *Econometrica* 69(6): 1597–1628.

Sargent, Thomas J. 1999. *The Conquest of American Inflation.* Princeton University Press.

Sargent, T., N. Williams, and Tao Zhao. 2006. "Shocks and Government Beliefs: The Rise and Fall of American Inflation." *American Economic Review* 96(4): 1193–1224.

Selten, Reinhard, and Rolf Stoecker. 1986. "End Behavior in Sequences of Finite Prisoner's Dilemma Supergames: A Learning Theory Approach." *Journal of Economic Behavior and Organization* 7(1): 47–70.

Shapley, Lloyd S. 1964. "Some Topics in Two Person Games." In *Advances in Game Theory* Annals of Mathematical Studies vol. 52, edited by Melvin Dresher, Lloyd S. Shapley, and Albert William Tucker, 1–28. Princeton University Press.

Smith, Vernon L. 1962. "An Experimental Study of Competitive Market Behavior." *Journal of Political Economy* 70(2): 111–37.

Sutan, Angela, and Marc Willinger. 2004. "Coordination In Cobweb Experiments With(out) Belief Elicitation." https://www.researchgate.net/profile/Angela_Sutan/publication/228793824_Coordination_in_Cobweb_experiments_with_%28out%29_elicited_beliefs/links/00b7d51fb70db34616000000.pdf.

Tirole, Jean. 1988. *The Theory of Industrial Organization.* MIT Press.

Van Huyck, John B., Raymond C. Battalio, and Richard O. Beil. 1990. "Tacit Coordination Games, Strategic Uncertainty, and Coordination Failure." *American Economic Review* 80(1): 234–48.

Wilcox, Nathaniel T. 2006. "Theories of Learning in Games and Heterogeneity Bias." *Econometrica* 74(5): 1271–92.

Young, H. Peyton. 1993. "The Evolution of Conventions." *Econometrica* 61(1): 57–84.