Whither Game Theory? *

Drew Fudenberg[†] David K. Levine[‡]

January 31, 2016

Abstract

We examine the state of game theory. Many important questions have been answered, and game theoretic methods are now central to much economic investigation. We suggest areas where further advances are important, and argue that models of learning and of social preferences provide promising routes for improving and widening game theory's predictive power, while preserving the sucesses of existing theory where it works well. We emphasize in particular the need for better understanding of the speed with which learning takes place, and of the evolution of social norms.

JEL Codes: 0000, 0001

Keywords: game theory, learning, social preference

David K. Levine, Department of Economics, European University Institute, Villa San Paolo, Via della Piazzuola 43, 50133 Firenze - Italy

*We are grateful to Juan Block, Michele Boldrin, Marco Casari, Anna Dreber, Rohan Dutta, Ben Golub, Gabriel Kreindler, Salvatore Modica, Andrea Mattozzi, David Rand, and members of the EUI Theory Reading Group.

[†]Departments of Economics, Harvard University and Yonsei University. Email: drew.fudenberg@gmail.com

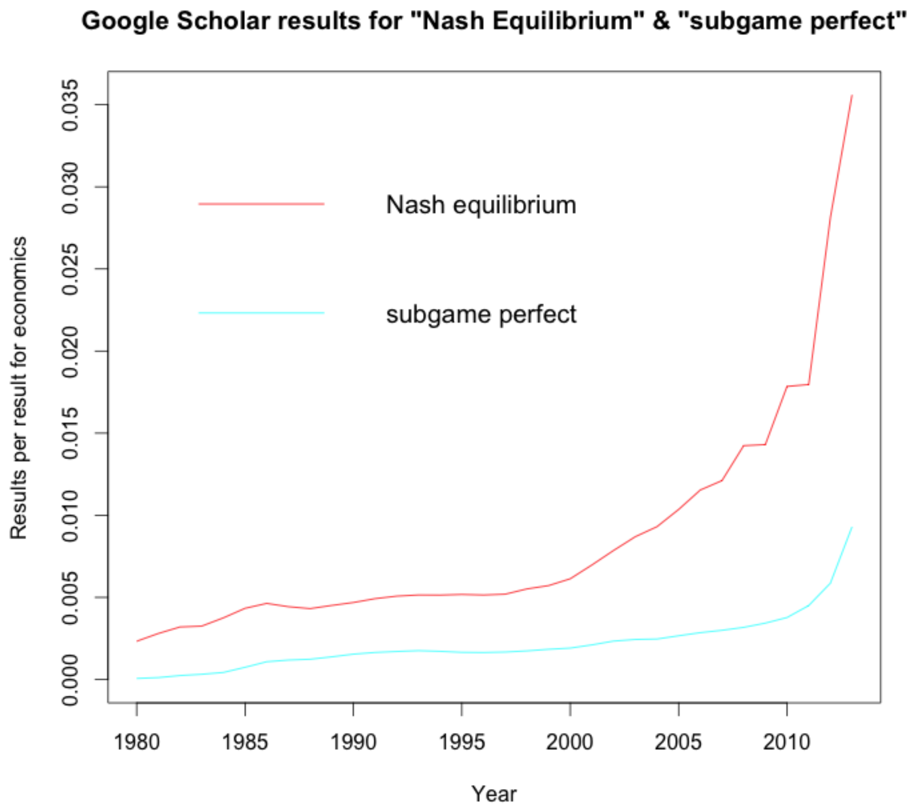
[‡]Departments of Economics, EUI and WUSTL. Email: david@dklevine.com

1 The Development of Game Theory in Economics

When we were graduate students at MIT (1977-81) the only game theory mentioned in the first-year core was static Cournot oligopoly, although Eric Maskin taught an advanced elective on game theory and mechanism design.[1] At the time, theorists such as Dasgupta, Dixit, Spence, and Stiglitz were exploring the implications of dynamic issues such as commitment and timing for such industrial organization topics as patent races and preemptive investment, Simultaneously, game theorists were developing tools that seemed natural for studying these problems, such as sequential equilibrium Kreps and Wilson (1982), and showing how to use formal game theoretic models to study dynamic competition, as in Milgrom and Roberts (1982) on limit pricing. The link between these tools and their applications expanded game theory's appeal to economists, and by the late 1980's, game theory had become a core part of many graduate programs. Ten years later, the "game theory takeover" was well established, and instructors had a choice of three graduate texts (Fudenberg and Tirole (1991), Osborne and Rubsinstein (1994) and Myerson (1991). Today game theory is integrated into mainstream economic reasoning. Mechanism design, and its many applications to market design, is an important and prominent spin off from core game theory, but game-theoretic ideas are at the core of modern industrial organization, political economy and experimental economics, and play a key role in the macroeconomic analysis of policy. A lot of progess has been made in applications to concrete problems ranging from auctions to market organization, and to monetary, fiscal and trade policy. At the same time game theorists have made significant progress in the more abstract parts of the theory, extending our understanding of dynamic games to account for informational imperfections, persistence and reputation. In the process great progress has been made in refining and extending game theoretic tools and work on the robustness of equilibrium, rationalizability, learning and evolution have changed our understanding of when and how game theory can be applied. In parallel experimental economics has developed as a field and the feedback from the theory to the laboratory and back again has become an important part of game theoretic research. Game theory has also spread to other fields, most notably computer science, where new computational methods have been developed and the theory has been applied to practical problems involving networked computers.

Now that the basic tools of equilibrium analysis are well understood, and have been widely applied, work by abstract theorists is no longer as tighly focused, and has diffused onto a number of themes. This hardly means that interest in or use of game theory has declined, as shown

[1] The syllabus for the class was the minmax theorem, the existence of Nash equilibrium, implementation theory, the core, Shapley value, and the bargaining set and other "cooperative" solution concepts.



by the following chart,
which compares Google scholar hits for “Nash equilibrium” and “subgame perfect equilibrium” to those for “economics” from 1980 to the present. So the role of game theory in economics does not seem to be in much doubt, but the role of game theory specialists is less clear. In this essay we argue that game theorists should focus their efforts on improving the ability of the theory to better predict actual behavior in a wider range of field and lab settings. To be clear, current game theory is successful because it works empirically in many important circumstances. To be equally clear: there are also circumstances in which it does a poor job empirically. The circumstances under which game theoretic equilibrium concepts do a good job can often be accurately predicted by the theory itself: when equilibria are not robust, the environment is complex, or when circumstances are unfamiliar, standard theory is less likely to perform well. In these circumstances there is greater scope for models of non-equilibrium play and behavioral factors to improve the theory. In thinking about where game theory should go next, we examine which of these forces are likely to be important, and the characteristics of a good theory that would explain them. Instead of trying to survey all of the promsing avenues for improving the theory,we focus on two that interest us and are close to our own current research, namely that models of boundedly rational learning and models of social preferences. We argue that these are two promising routes for improving and widening game theory’s predictive power, while preserving the sucesses of existing theory where it works well.

2 Game Theory and Learning

Nash equilibrium is a remarkably powerful tool for understanding human interactions. One example of a game with which we are all familiar is the rush-hour traffic game. This is played on weekdays in the morning in most major cities in the world and lasts several hours. There are millions of players, the commuters going by car from home to school or work. The actions are the many different routes that they can take: the main highway, short-cuts through the sides streets, the North route, the South route, and so forth, and the commuters may also have some some leeway over their departure times. Their preferences are to the first approximation get to the destination as quickly as possible[2] Nash equilibrium is the requirement that no commuter can save any time by taking a different route. For this to be approximately true, it must be that if you get off the congested main highway and dodge through side streets you should find that just enough other people are doing the same thing that you derive no advantage from doing so. [3]

Historically there are two views of Nash equilibrium. The first formal analysis is due to Cournot (1838), who viewed (what we now call) "Cournot equilibrium" as the steady state of a non-equilibrium adjustment process. This viewpoint was eclipsed for a time by the "ultra-rational" view of Von Neumann andMorgernstern (1944) who saw game theory's task as providing "mathematically complete principles which define 'rational behavior' in a social economy." They wanted to make unique unambiguous predictions. In two-player constant-sum games rational behavior implies Nash equilibrium and does indeed deliver a clear prediction. However most of the games economists care about are not two-player constant-sum games, and in general the assumption of rational behavior does little to restrict the set of possible outcomes; indeed, even stronger notions such as common knowledge of rationality not deliver a unique prediction when there are multiple equilibria. To see the problem consider the very simple example of choosing on which side of the road to drive on. This can be described by a payoff matrix where player 1, the row player chooses to drive either on the $L(eft)$ or the $R(ight)$ and similarly for the column player.

	L	R
L	$1,1$	$0,0$
R	$0,0$	$1,1$

If both drive on the same side of the road, both reach their destination and get 1. If they choose opposite sides of the road they collide, and neither reaches their destination and both get zero. This is the classical example of a coordination game. It has two pure strategy Nash equilibria, as well as a mixed equilibrium where they randomize 50-50. The key point is that being rational does

[2]If their time of arrival is not fixed, they may have preferences over this as well. They may also have preferences over things such as safety.

[3]Traffic engineers have made extensive use of the assumption of Nash equilibrium, where it is called a "user equilibrium," and have found some qualitative support for its predictions. See, for example, Florian (1976). One of us has extensive experience with a particular version of the rush-hour traffic game, and after extensive experimentation concluded that Nash equilibrium fit well. On the other hand, Larcom, Rauch, and Willems (2015) argue that a recent Tube strike shows that Nash equilibrium did not apply to pre-strike Tube commuters

not provide a clear answer to which equilibrium should be chosen - and indeed while we observe the right-hand side of the road equilibrium in most countries, in others such as England, Ireland, Australia and Hong Kong we observe the left-hand side of the road equilibrium.

Today, we, and we believe most economists, think of Nash equilibrium and its variations as arising from some sort of non-equilibrium adaptive process of learning, imitation, or evolution. There are several reasons this is the case. First, as the driving example makes clear, it is difficult to understand how reasoning alone can solve coordination problems. Second, as the rush hour game should make clear in many practical instances it is scarcely feasible to find Nash equilibrium by rationality alone: it seems certain that commuters do not make choices of routes by contemplating the rational play of all the other commuters and working out how they should best respond. Indeed, it is pretty clear that the optimal pay of commuters comes about by trial and error: Some people try different routes and settle on the ones that seem the quickest. That is, players have many opportunities to play and they learn from their experience. Finally, laboratory evidence shows little sign that - except in some special cases - initial play resembles Nash equilibrium when the equilibrium is not unique, while there is abundant experimental evidence that play in many games moves towards equilibrium as subjects play the game repeatedly and receive feedback.[4]

There are several reasons for our interest in the process by which Nash equilibrium is reached. One is that "how we get there" may have clues to where "there" is when there are multiple equilibria. 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 eqiulibrium is reached and persists for a long time, there may still be regime changes. In the driving game, for example, we observe that in 1967 in Sweden the equilibrium changed from driving on the left to driving on the right. Our next goal is to give two case studies of significant regime change. In one case it took decades to move from one equilibrium to another. In the other it took minutes. Our focus in this essay is on understanding how rapidly learning takes place with the goal of understanding why regime change is sometimes very slow and other times very fast.

Learning about Monetary Policy

At one time the most fraught issue that divided the economics profession was the proper use of monetary policy to combat recessions and the nature of the Phillips curve relating unemployment to inflation. This debate was of key interest to policy makers, especially the Federal Reserve Board, and Fed policy to provide long term stimulus and combat recession led eventually to the stagflation of the 1970s which managed to combine high unemployment with high inflation. The original Fed view, as summarized by Ben Bernake several decades later

> economic theory and practice in the '50s and early '60s suggested that there was a

[4] Smith (1962) is an early example. Selten and Stoecker (1986) document the unravelling of (non-equilibrium) cooperation in the finitely repeated prisoners dilemma as sujects gain feedback, while in their study of coordination games Van Huyck, Battallio and Beil (1990) say (pp 240-241) "Repeating the period game does cause actions to converge to a stable outcome." Note that this leaves open the question of how many repetitions are needed; we will say more about this below.

> 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.[5]

is now widely viewed as wrong. It took the Fed four decades to figure this out, what we would describe as "slow" learning.[6]

How this came about has been addressed by Sargent, Williams and Zha (2006) who used a game theoretic model of inflation and study how learning takes place. For illustrative purposes we adopt a simplification of the model from Fudenberg and Levine (2009). 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 policy maker prefers low inflation but is willing to chose high inflation if this leads to lower unemployment; for concreteness we will suppose that the policy maker's payoff is the sum of an unemployment term and an inflation term, and that the policy maker gets 2 for low unemployment, 0 for high unemployment, 1 for low inflation and 0 for high inflation. Regardless of what inflation policy is chosen, the representative consumer's payoffs are such that he will always choose low unemployment.

This 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 representative consumer to respond to low inflation by choosing high unemployment and high inflation by choosing low unemployment, and for the Fed to choose high inflation. On the face this is a bit puzzling - given the way the representative consumer is playing obviously the Fed should choose high inflation. But how can it be optimal for the representative consumer to choose high unemployment in response to low inflation? The answer lies in a kind of loophole in the definition of Nash equilibrium. Because the Fed is choosing high inflation the way in which the representative consumer responds to low inflation is purely hypothetical - it is "off the equilibrium path" and so it does not make any difference to the representative consumer whether she chooses high or low unemployment in response to a circumstance which never happens. Game theorists have developed two responses to this loophole.

1. Close the loophole. Subgame perfect equilibrium strengthens Nash equilibrium by requiring that players should choose best responses to all contingencies whether or not they are on the equilibrium path. In the Phillips curve game only the low inflation/low unemployment Nash equilibrium is subgame perfect. However, subgame perfection is not robust to even small amounts of payoff uncertainty (Fudenberg, Kreps and Levine (1988), Myerson (1978), and in more complex games than the illustrative example used here it requires that players have extensive knowledge about off-path play. For example, it requires that a player who has never seen his opponents play a particular 2x2 coordination subgame believes that they will coordinate on one of the equilibria of that subgame and moreover that the player correctly forecasts the way in which his opponents

[5] Domitrovic (2012)

[6] See, for example Sargent (1999).

will coordinate.

2. Take learning more seriously. An alternative story we can tell about high inflation is that the representative consumer always chooses low unemployment and that the Fed chooses high inflation and incorrectly believes that if they were to choose low inflation the representative consumer would choose high unemployment. The key point is that on the plausible assumption that the Fed does not have detailed knowledge of consumer behavior and preferences but learns from experience, 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 at information sets that have probability zero under the equilibrium strategies. [7]

The story then that explains Fed behavior is that they were stuck in a non-Nash self-confirming equilibrium. Believing that low inflation would lead to high unemployment, they chose high inflation, and never learned that their belief was wrong. While this story may have an element of truth in it, the self-confirmimg equilibrium concept cannot explain why - after a decades long delay - they eventually learned that low inflation would not lead to high unemployment. That is, while self-confirming equilibrium may be a good short-run description, it may not be a good long-run description, because people may eventually to get enough evidence about off path play to correct their mistaken beliefs in this example learning was very slow. We turn next to an example where learning was very fast.

The Hijacking Game

We turn now to a case where learning did not take decades but instead was breathtakingly fast. In response to changed circumstances, an equilibrium that involved millions of air passengers and had lasted for decades changed to a completely opposite equilibrium, and the switch took place spontaneously and in less than half an hour. We start with a simple 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 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 mild the passengers are eventually released and get -1. If the hijackers are severe the passengers will probably be killed; to make the calculation simple we set the expected utility

[7] Fudenberg and Levine (1993a) provide a definition for general extensive form games that allows for heterogeneous and correlated beliefs. See also Battigalli and Guatoli (1997) and Kalai and Lehrer (1993).

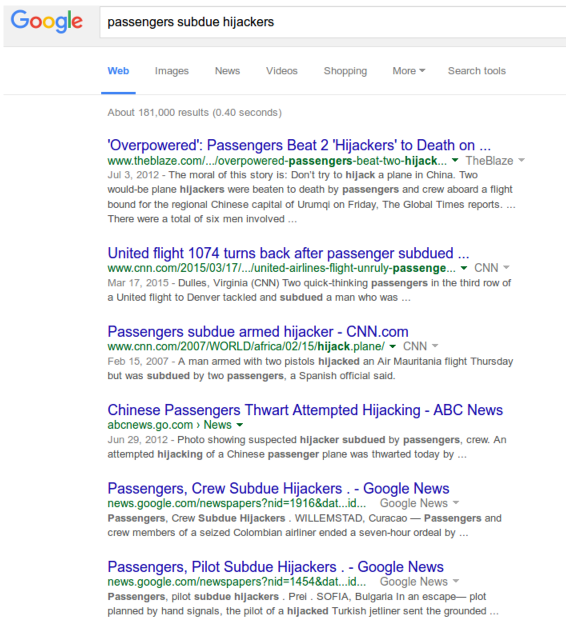
of the passengers to be -3.

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; suppose that this is the equilibrium that people play and expect to be played.

What if an exogenous change in circumstances drops the probability of mild hijackers to 25%? Then hijack/acquiesce is no longer a Nash or even a self-confirming equilibrium - the only Nash equilibrium is for the hijackers to stay out and for the passengers to fight at with at least 50% probability. Overtime we might would expect that the passengers would gradually learn that it is now better to fight. When the hijackers in turn learn that passengers are choosing to fight we would then expect that 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 would take some time. That, however, is not how it happened.

In the 1990's there were around 18 aircraft hijackings a year. Most ended peacefully, and the longer a hijacking persisted the more often there was a peaceful ending. Flight crews were trained in the rational and FAA-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 strongly validated by decades of experience. This was indeed the hijack/acquiesce equilibrium.

Circumstances changed on September 11, 2001 when hijackers, rather than landing planes and making demands, used the hijacked aircraft for suicide attacks on ground targets. The theory predicts that the equilibrium should shift from hijack/acquiesce to stay out/fight. And 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. The screen shot of a Google search page for "passengers subdue hijackers" below makes the point: Post 9/11, if you hijacked a plane you were lucky if you were only beaten half to death.



So far so good. However, while the Fed took decades to learn about monetary policy in this case-study learning was extremely fast.[8] At 8:42 a.m. on September 11, 2001, United Airlines Flight 93 took off. The first evidence of a change in hijacker type occurred four minutes later when American Airlines Flight 11 crashed into the North Tower of the World Trade Center. Forty-two minutes after this, at 9:28 a.m., United Airlines Flight 93 was itself hijacked. It took only another twenty-nine minutes for passengers and flight crews to adjust their behavior. At 9:57 a.m. the passengers and crew on United Airlines Flight 93 assaulted their hijackers. Only an hour and 11 minutes elapsed from the first evidence of a regime change until a rational response was determined and implemented. It happened on a plane already in the air based on limited information obtained through a few (37) telephone calls. The information available was that two previously hijacked planes had been crashed into the World Trade Center.[9] The passengers and crew of flight 93 did not have access to the dramatic video showing planes crashing into the World Trade Center, and the second hand reports provided by telephone was not unambiguous. A friend of one of the passengers on the phone with him at 9:49 a.m. is quoted as saying "Don't worry, they hijacked the plane, they're gonna take you for a ride, you go to their country, and you come back. You stay there for vacation". - This suggest that the hijack/acquiese equilibrium was still in effect. At 9:57 a.m. the passengers on the flight had at most three noisy observations to offset some four hundred well established observations of earlier hijackings. The response was not a minor adjustment. It was dangerous and dramatic. The passengers and crew of flight 93 risked – and sacrificed – their lives.

It should be clear that there are some missing ingredients in our model. One thing that would have been clear to the passengers although not to a computer program implementing some variation of the fictitious play learning rules described below is that the most recent observations were far

[8]See USA (2004).

[9]Although the third plane crashed into the Pentagon before the passengers started to fight it happened only thirteen minutes before so it is unclear if this information was available.

more salient than the earlier ones : It would be strange indeed if severe hijackers had seized two planes and the third plane was seized by some independent group of mild hijackers in some sort of cosmic coincidence. This highlight one of the problems with learning theory, namely that people are smart and sophisticated in ways it is hard to model.

It is worth noting that both in the case of the Fed and the hijackingdecisions were taken collectively, and were based on the reported experiences of others, while our learning models typically focus on individual decisions based on direct learning from personal experience. We will say more about the issue of collective decision makingwhen we examine game theoretic political economy models.

Games and Equilibria

We will need some notation and definitions to make our discussion of learning and equilibrium clear. The most general notion of a game is that of an extensive form game, which specifies the timing of moves and the information available to players when they make those moves. To keep things simple, we will only describe learning in one shot simultaneous move games, which in the two player case can be described by a payoff matrix. General extensive form games may be reduced to simultaneous move games by identifying the actions of players in the simultaneous move game with their strategies in the extensive form - complete contingent plans that specify actions in every situation in which the player may find himself. Note, however, that in the context of learning what is observed at the end of an extensive form game is not generally the entire strategic choice of the other players, but only what actually happened in the game ("the equilibrium path"). Our formal of learning models will gloss over this importamt point - formal analyses can be found, for example, in Fudenberg and Levine (1993b) or Fudenberg and Kreps (1988).

To give a formal description of a one shot simultaneous move game, it is a list of players $i = 1, \ldots, N$, a set A^i of the actions that each player might take, and the preferences of each player i over the actions taken by all players, which we will typically suppose tales the form of a utility function u^i , where $a = (a^1, ..., a^n)$ is an action profile that specifies the actions of each player i. A Nash equilibrium of such a game is an action profile where each player's action is optimal given the actions taken by the other players. To express that formally we generally write the vector $a = (a^i, a^{-i})$ where a^{-i} is the vector of actions of all players except for player i, and say that a is a *Nash equilibrium* if for every player i and every action $\tilde{a}^i \in A^i$ we have the equilibrium utility $u^i(a)$ at least as great as the utility $u^i(\tilde{a}^i, a^{-i})$ from the alternative $\tilde{a}^i$.

Another way of viewing Nash equilibrium is that it 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. On the other hand if a profile of actions is not a Nash equilibrium then at least one person is not doing the best they can given the play of the others.

3 Passive Learning

We now have two examples: one involving very slow learning (decades) and another involving very rapid learning (minutes). Motivated both by theoretical considerations and laboratory evidence, game theoretic models of learning have advanced significantly over the last decades. Our next step is to examine the state of the art in game-theoretic models of learning, and then discuss how these models do or do not inform us about the speed and nature of learning outside the laboratory.

The simplest setting is one in which players repeatedly meet other players on a one-time basis, observe the strategies of their opponents, and there are no moves by Nature, so that regardless of how a player acts he can compute the payoff he would have received to any of his choices. In this setting learning is purely passive: a player plays the game, observes what his opponent did and updates her beliefs about the strategic choices in the pool of opponents she faces. The issue of "off the equilibrium path" does not arise because there is no need to worry either about how the current opponent will play in the future (that opponent will not be seen again) and because the entire strategy of the opponent is observed, so the question of "what would the opponent have done if I'd done something else" does not arise. Note that this is not the case in the traffic game described earlier, as there players do not observe what their travel time would have been on other routes. In such cases "active learning" or "experimentation" is needed for players to know whether they are playing a best response; we defer until later the new issues that this generates..

For simplicity we will restrict our discussion of passive learning to two player simultaneous-move games. A classical model of learning is that of fictitious play 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, $t = 1$, 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. That is, player i keeps track of the statistic ϕ_t^i ,which is a probability distribution over A^{-i} with the property that $t\phi_t^i[a^{-i}]$ is the number of times that player $-i$ has played a^{-i} through the end of period t. In period $t + 1$ each player i plays a best response to ϕ_t^i, that is, chooses a^i to maximize $u^i(\tilde{a}^i, \phi_t^i)$ employing an arbitrary tie-breaking rule in case there is a tie. It is easy to see 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).[10]

Economists tend to think of learning in terms of Bayes law, and Fudenberg and Kreps pointed out that fictitious play can be interpreted in this way. If a player believes she faces a stationary distribution of opponent strategies, so that a^{-i} is like a coin flip then with a Dirichlet prior the posterior is also Dirichlet. The Dirichlet family is parameterized by numbers $\kappa_t^i(a^{-i})$ representing the number of times a^{-i} has been seen in the sample together with a "prior" number of observations.

[10] Moreover 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. This will be true for any Bayesian who thinks the world is stationary (that is that the sequence of opponent's actions is exchangeable) and has a non-doctrinaire prior.

That is $\kappa_t^i(a^{-i}) = \kappa_{t-1}^i(a^{-i})$ if a^{-i} was not observed at time $t-1$, that is $a^{-i} \neq a_t^{-i}$, and $\kappa_t^i(a^{-i}) = \kappa_{t-1}^i(a^{-i})+1$ when $a^{-i} = a_t^{-i}$. Initally, before play starts, the prior is described by given parameters $\kappa_1^i(a^{-i})$. Note that these prior parameters do not have to be integers nor even positive, but if they are positive integers they may be interpreted as an initial sample of observations acquired before the game begins. With the Dirichlet family, the probability of a^{-i} is given simply by $\kappa_t^i(a^{-i})/\sum_{\tilde{a}^{-i}} \kappa_t^i(\tilde{a}^{-i})$, that is, if we take the prior $\kappa_1^i(a^{-i}) = 0$, posterior probabalities are simply the empirical frequences $\phi_t^i[a^{-i}]$ the same as fictitious play, so that this model represents a slight generalization of fictitious play to allow for general priors.

Fictitious play, with or without the Bayesian interpretation, presupposes that players play a best response to the beliefs that they have learned. To allow play to converge to a mixed strategy Nash equilibrium, Fudenberg and Kreps (1993) replaced the exact best response with the assumption that payoffs are subject to independently distributed, privately observed payoff shocks, as in Harsanyi's [1973] purification theorem. The resulting "Nash distribution" - the distribution over actions induced by the intersection of the induced stochastic best response functions - is a steady state of the associated fictitious play; it is better known as a "quantal response equilibrium " (McKelvey and Palfrey (1995)).[11] More generally, replacing the exact best response function of fictitious play with a smooth approximation to it results in "smooth fictitious play", which has both theoretical and empirical advantages. First, using a deterministic best response opens a player up to exploitation by a clever opponent (Blackwell (1956), Fudenberg and Kreps (1993), while randomization allows various forms of non-exploitability . [12]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. This is implausible and indeed in experimental data even when we see convergence to Nash equilibrium the play of individual players is quite noisy, and is better described by a random response.

There is a second class of models used empirically to study learning, reinforcement learning models 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 the anticipated utility from an action.[13] A general class of reinforcement learning models can be written in the following

[11]The simplest and most commonly used random utility model is the logit, which assumes that the payoff shocks have an extreme-value distribution. Here player i's probability of choosing a^i is

$$\alpha_t^i(a^i) = \frac{e^{\lambda u^i(a^i,\phi_t^i)}}{\sum_{\tilde{a}^i} e^{\lambda u^i(\tilde{a}^i,\phi_t^i)}}.$$

When $\lambda = 0$ choice is completely random. As $\lambda \to \infty$ the probability a best response is played approaches 1. The same logit responses can also be generated by maximizing the sum of expected utility and λ times the entropy of the distribution of actions. The connection between the logit model and the extreme-value distsribution of random utility model has been know at least since the work of McFadden (1973).

[12]See for example Fudenberg and Levine (1995), Foster and Vohra (1999), Hart and Mas-Colell (2001). Instead of using a random utility model, Fudenberg and Levine generate stochastic responses by assuming that each player maximizes the sum of expected utility and a small strictly concave perturbation term; Fudenberg, Iijima, and Strzalecki (2015) gives a revealed preference characterization of a similar representation

[13]In the reinforcement learning literature the term "propensity" is used but since these propensities are in fact measured in units of utility it is easier to see the connection with fictitious play models if we think of them as anticipated utilities.

form

$$u_t^i(a^i) = \frac{t u_{t-1}^i(a^i) + \gamma_t(a^i) u^i(a^i, a_t^{-i})}{t+1}.$$

In modern work on reinforcement learning models, the logit choice model described above (often called "softmax" in this literature) is used to convert these expected utilities into probabilities of play. If $\gamma_t(a^i) = 1$ this model is exactly that of smooth fictitious play.

In the original Roth Erev formulation $\gamma_t(a^i) = 1$ if $a_t^i = a^i$ and 0 if $a_t^i \neq a^i$. In other words, the only action for which the utility weights are updated is 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[14] using variations of this model, Ho, Camerer and Chong (2007) proposed self-tuning EWA, which they report does a good job of fitting a variety of experimental data. In this model, the updating rule is $\gamma_t(a^i) = 1$ if $u^i(a^i, a_t^{-i}) \geq u^i(a_t)$ and 0 otherwise so that weights are updated for every action that would have done at least as well as the action that was chosen[15] This 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 justify 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."

Camerer, Ho and Chong argue that their model of $\gamma_t(a^i)$ fits experimental data better than belief learning models in which $\gamma_t(a^i) = 1$. Wilcox (2006) among others have contested this view. In particular, Camerer, Ho and Chong assume that agents are homogeneous. Wilcox (2006) argues that if players are are heterogeneous with different players having different values of the logit parameter λ, then assuming a common λ biases estimate of the learning parameters in against belief learning and in favor of reinforcement learning. We think that the assumption of homogeneous priors has the same effect. To see why, observe that in many experiments play converges to a pure strategy equilibrium. This means that there is little heterogeneity in play at the end of the experiment even though there is a great deal at the beginning. Explaining heterogeneity with a logit parameter λ that is time invariant (regardless of whether it is homogeneous or heterogeneous) and homogeneous priors yields the same degree of heterogeneity at the end as at the beginning. The way that the reinforcement learning model gets around this is by "depreciating" the utility of actions that are not used. By doing so it increases the utility differences between the action being converged to and the remaining actions, which reduces heterogeneity at the end of the game. We do not know whether allowing heterogeneous priors is sufficient to explain the difference between beginning and end heterogeneity when play converges to a pure strategy equilibrium. There are, however, alternative approaches to reinforcement learning for "depreciating" the utility of unused actions: for example learners may use larger values of λ when the data indicates a high degree of certainty about the environment. This approach is used both in Fudenberg and Levine (2014) and

[14]See for example, Cheung and Freedman (1997), Camerer and Ho (1999), Salmon (2001).

[15]They also include recency in the model, which we discuss subsequently.

Block, Fudenberg and Levine (2016).

All of these models weight current and past observations the same way. This is hard to reconcile with the hijacking example. With hundreds of past observations about hijackers, these models do not predict an abrupt change in behavior in response to only two new observations. Indeed there is little doubt that the two most recent observations weighed much more heavily in the minds of the passengers of flight 93 than the hundreds of previous observations. The idea that recent observations might get more weight than older observations is called *recency.* Before examining how game theorists have incorporated recency we look at an example to illustrate both how convergence to equilibrium takes place in the laboratory and how recency may - or may not - help us understand that convergence.

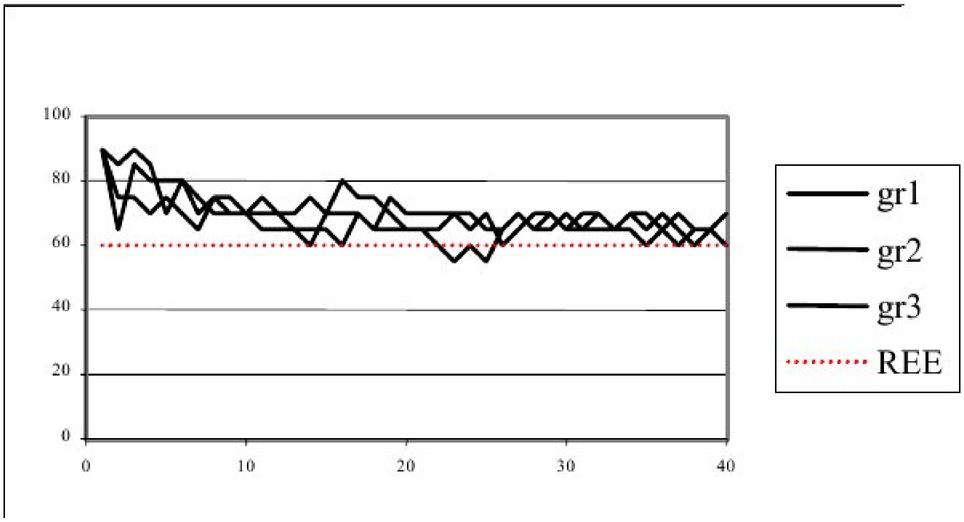
The Cobweb

The extreme example of recency is to use data only from the most recent period, as in the adjustment process described by Cournot. Depending on the parameters this can give rise to cycles, as in the famous cobweb first described in Kaldor (1933). This was studied in the laboratory by Sutan and Willinger (2004).[16] They consider a simple market with five suppliers. Each supplier i faces a marginal cost of producing output q^i of $5q^i/21 + 300/7$. Letting q be industry output and p industry price, industry demand is given by $q = 900 - 9p$. Suppose that the initial price was 90 as it typically was in the laboratory. Firms will wish to produce 990 units of output. That leads to a price of 0. Forecasting zero price next period firms produce nothing, causing price to rise to 100. The following period the industry produces 1200. The cycle then continues with the market alternating between overproduction leading to a zero price, then underproducing leading to a price of 100. Despite the extreme implausibility of this it has been used to explain everything from the business cycle to the collapse of capitalism.

The situation in the laboratory was rather different. As the graph below shows experimental subjects converged rather quickly.[17]

[16]Other authors have studied cobweb dynamics such as Hommes et al (2007), who examined a more complicated environment with demand uncertainty. They also find convergence to rational expectations equilibrium, albeit with some price volatility that can be related to the stability of the learning dynamics.

[17]The competitive price in this market is 60, and convergence is roughly 65 - this reflects the fact that with only five firms each firm does have some market power, and in fact the Cournot equilibrium is at a price of 65.



Vertical axis: price p; horizontal axis: number of periods played. gr1, gr2, gr3 are results of different sessions.

This experiment (along with many others) shows that extreme recency is not a good model. It is worth noting that our calculations show that if subjects were to use an average of the price 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 - things have settled down pretty well after ten periods.

Recency

In the hijacking example people placed a high degree of weight on the most recent observations, while in the cobweb example play is inconsistent with an extreme degree of recency. This raises the question of when and why we should expect people to give recent observations more weight. In many settings this sort of updating makes sense - if the signal-generating process undergoes unobserved regime shifts, then older observations may indeed be less informative than recent ones. Psychologists[18] talk of "recency bias" because sometimes people do this even when they are told the environment they face is stationary.[19] There are two approaches to modeling recency. One is to explicitly develop Bayesian models of changing environment such as hidden Markov models. This approach is not widely used because models of this type are quite complex.[20] As a result most models of recency have instead focused on simple rules of thumb. A simple and obvious starting point is to modify the fictitious play/reinforcement learning specification to specify that

[18] See for example Erev and Haruvy (2013).

[19] As with many psychology experiments, we need to consider whether the subjects believed what they were told about the experimental design.

[20] 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.

older observations receive exponentially less weight, so that the reinforcement rule becomes

$$u_t^i(a^i) = \frac{\theta_t^i u_{t-1}^i(a^i) + \gamma_t(a^i) u^i(a^i, a_t^{-i})}{\theta_t^i + 1}.$$

where θ_t^i is a weight on past data.[21] For example, if we take $\gamma_t(a^i) = 1$ and $\theta_t^i = t$ we have classical fictitious play. If $\gamma_t(a^i) = 1$ and $\theta_t^i = \theta$ we have

$$u_t^i(a^i) = \frac{\theta u_{t-1}^i(a^i) + u^i(a^i, a_t^{-i})}{\theta + 1}$$

so that when $\theta = 0$ all weight is placed on the most recent observations, while as $\theta \to \infty$ most weight is placed on the past and play becomes similar to classical fictitious play in a very large sample. If $\theta = 1/2$ then the weight on the most recent observation is $2/3$. In the cobweb example this is too high in the sense that it implies counterfactually that players converge to a cobweb cycle between a price of zero and 126.

The basic experimental fact about the recency model is that while it does well in describing some learning paths, in general the weight θ varies experiment to experiment, and may well vary from subject to subject. Moreover when subjects get stochastic feedback, recency can leads to an ergodic distribution of play that is very different than any Nash equilibrium. Fudenberg and Peysakovich (2014) provide evidence of this: in their experiment subjects face a simple lemons problem in which the computer chooses a value v uniformly distributed on $[0, 10]$,and the value of the object to the subjects is $v + k$, where k is a known parameter. The subjects make 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.[22] However, in another treatment the Nash equilibrium bid is 6 while the actual mean bids converged to about that value. [23] These facts can be matched by simulations with a high degree of recency, and in the data even after some 20 rounds of play subjects appeared to place most of their weight on the most recent two observations.

The self-tuning EWA model of Ho, Camerer and Chong (2007) proposes a way to endogenize the recency parameter θ_t^i.. This model has the specification

$$\theta_t^i = 0.5 + (t-1)\left[1 - (1/2)\sum_{a^{-i}}\left(\phi_t^i(a^{-i}) - \phi_{Wt}^i(a^{-i})\right)^2\right]$$

whereϕ_{Wt}^i is the frequency of opponent play over the most recent W periods only. We will discuss

[21] This is used in the econometric specification of Cheung and Freedman (1997) and the theoretical work of Benaim, Hofbauer and Hopkins (2009) and Fudenberg and Levine (2014). With this specification, beliefs do not asymptotically converge unless the agent sees exactly the same signal every day. This makes formal analysis difficult, except in the extreme case in which beliefs ignore all but the last outcome.

[22] In this treatment the cursed equilibrium of Eyster and Rabin (2005) predicts that each subject will consistently bid 4.

[23] Here the cursed equilibrium bid is higher than the data or the Nash equilibrium.

determination of the window W shortly, but for the moment regard it as a fixed parameter. If the frequency in the window $\phi^i_{Wt}(a^{-i})$ is similar to that $\phi^i_t(a^{-i})$ over the entire sample, then $\theta^i_t = 0.5 + (t-1)$ which means essentially no recency. If things do not seem to have changed, then behavior is as in classical fictitious play (or reinforcement learning). On the other hand $\sum_{a^{-i}} \left(\phi^i_t(a^{-i}) - \phi^i_{Wt}(a^{-i})\right)^2$ could be as large as 2 if the frequency in the window fluctuates quite a lot. In this case $\theta^i_t = 0.5$ which as we know is a very high degree of recency - enough that it does not occur in the cobweb example. All of this makes good sense. Indeed theoretical models of learning such as Fudenberg and Levine (2014) presuppose a limited degree of recency that is triggered by fluctuations, and show that these sorts of learning rules tend to perform well for the agents who use them, and have and that players using them converge globally to Nash equilibrium.

The key element of the self-tuning specification is the determination of the window W. This is taken to be the number of actions in the support of Nash equilibrium.[24] The reason for this is clear enough: If the window length W is less than the number of actions in the support of a Nash equilibrium then a mixed equilibrium will give rise to obvious fluctuations and lead to recency that prevents convergence, yet in many simple games with mixed strategy equilibrium (matching pennies, rock-paper-scissors) convergence does occur, so a model using a shorter window would be contradicted by the data. On the other hand when the equilibrium has many many outcomes, the requirement that W should be large could rule out any significant degree of recency. Hence while the window W is essential, it is not consistent with evidence about recency and is rather ad hoc, seeming to base a learning model on computations of Nash equilibrium by the players.

In its defense, though, self-tuning EWA does attempt to capture the important idea that recency is triggered by change. It says that if things have been stable for a while but seem to have changed then you use greater recency. This 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). What this misses is that when we are surprised we do not just put weight on more recent data, but we also re-evaluate old data. So if we have an unusually severe recession we go back and we look at 75 year old data and say things like "this is a bit like the Great Depression, we see that the solution to that seems to have been massive government spending on the war, so maybe we ought to try massive government spending to get out of this."

The bottom line is that we do not yet have satisfactory models of reconsideration. An introspective thought experiment involving the driving game helps illustrate the difficulties involved. Which side of the road to drive on is a long, well-established and highly stable equilibrium. If you are driving on a deserted road and you encounter a driver driving on the wrong side and have to swerve to get out of the way you would conclude that this was a crazy driver and would not conclude that the equilibrium had changed. If the next driver you encountered was also driving on the wrong side of the road you would conclude that something strange was going on - you still would not conclude that the equilibrium had changed, but would probably contemplate the possibility that there was

[24] It is unclear how this is to be applied when there is more than one Nash equilibrium.

something unusual about this road at this particular time. If the next driver you encountered was driving on the wrong side of the road you would surely conclude 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, but is presumably "many." By contrast, if after encountering the first driver everything was normal for a few days, and you then encountered a second driver on the wrong side of the road, your conclusion would probably be that the number of crazy drivers on the road had gone up. This example also highlights the fact that in practice there may be alternative sources of information - probably after encountering a second wrong-side drive in a row you would pull over, call someone on the phone, and ask if they had seen any crazy drivers or heard about anything funny going on.

From the Laboratory to the Field

There is a second problem with recency models in general and with self-tuning EWA in particular. Fudenberg and Peysakovich (2014) find that when they give players a cross-section of ten observations they make little use of the extra data. However, if they are given a summary of those ten observations in the form of an average they use that information effectively and no longer exhibit recency, converging instead to Nash equilibrium. Hence at least part of recency observed in the laboratory appears to be due to inability to process information effectively in laboratory conditions. It is worth noting that the Fed decisions are taken by a committee that has access to a wide range of data and data analysis tools. It hardly looks at data from last week or last year, but 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 opportunities using the Black-Scholes model again all estimated using pretty long data series. Even the guys who mucked up the bond ratings of mortgage backed securities used models of housing prices based on data going back a long way.

This is not to say that EWA or other models used to explain laboratory results are irrelevant for understanding what the Fed or the investment banks do. Rather, - we need to understand that the actions/strategies they are choosing between include "using econometric model A" or "model B" or "model C". Hence there is meta-learning (which econometric model to use to learn with) and the laboratory models may be good descriptions of the meta-learning process.

4 Active Learning

A crucial feature of meeting on a one time basis and observing the entire strategic choice of the opponent is that there is no need to speculate about counterfactuals. In a setting such as the simple game between the Fed and representative consumer the Fed cannot see the counterfactual "what would have happened if instead of setting high inflation I set low inflation" yet must somehow infer the causal relationship between inflation and unemployment. Similarly in the traffic game, players need to experiment to learn the travel times on each route. [25]This problem is in general pronounced in settings where players play the same opponents repeatedly since so much of the game is "off the equilibrium path." Yet players must infer causality: if I were to cooperate, would my opponent reward me? If so, for how long?

The simplest model of active learning is the one player bandit problem where only the payoff from the choice implemented is observed. A great deal is known about the mathematical structure of these problems Gittins (1988) (and it forms the underlying basis for the study of learning in games and self-confirming equilibrium. The basic insight from this literature is that a player may get stuck permanently on the wrong choice, corresponding to a self-confirming equilibrium, but this is less likely to happen if feedback is accurate and rapid.[26] In the game setting, as in a bandit problem, a player who never takes a given action never sees how others would respond to it. For this reason, we expect that patient players may choose to experiment with actions that they think are probably suboptimal, just to learn more about their consequences. In addition, a player who rarely if ever gets to move has little chance to accumulate information and fewer chances to use the information she does acquire, so even a patient player who rarely gets to move may choose not to experiment.

With these tools, we can now understand better the Sargent, Williams and Zha (2006) analysis of what happened to the Fed. As we have indicated, backwards induction leads the policy maker to chose low inflation and receive a payoff of 3. Also as we indicated, there is a self-confirming equilibrium in which the policy maker chooses high inflation due to a mistaken belief that low inflation leads to high unemployment; here the policy maker's payoff is only 1. Obviously this model is a highly simplified version of the Fed's learning process. 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 U.S. They provide a more detailed model of Bayesian learning that allows takes account of the fact that some relevant data is revealed even in the high inflation regime, so that posterior beliefs about the Phillips curve change as data accumulates and argue that this learning model can explain many of the details of U.S. monetary policy and inflation during 1970's and 80's.

The simple model can be used to highlight important issues about the speed of learning. Consider first the low inflation equilibrium. We may imagine that an econometrician comes to the policy maker with some evidence that a new policy ("high inflation") would have previously unsuspected benefits ("really low unemployment"). If the new "high inflation" regime is introduced, how long will it take before the policy maker discovers the error? By contrast, we can consider the high inflation self-confirming equilibrium - the Fed will have some incentive to investigate whether it is really true that low inflation will result in high unemployment. How long will it take them to learn

[25]Although in the traffic game, not all players need to experiment to ensure that the outcome is a Nash equilibrium.

[26]One complication is that self-confirming like behavior can persist for a long time even when enough information is generated on the equilibrium path - due to shocks and the like - that we should see Nash equilibrium. Of course if the information emerges slowly then the convergence to Nash can take a long time. See the discussion of Sargent, Williams and Zha (2006) below.

that this is not true? In both cases greater patience (a higher data arrival rate) increases learning effectiveness. However, the role of patience in the two cases is very different. In the case of the optimistic econometrician, a more patient player will spend more time playing the wrong action, but in a discounted sense will care less about this length of time. As the discount factor increases, learning effectiveness gradually improves, converging to one in the limit. In the self-confirming case, when the discount factor is small the player never escapes the self-confirming equilibrium, so learning effectiveness is zero. Once the discount factor passes a critical threshold, the agent starts to experiment with the risky action, and the chance of learning the truth jumps substantially. It turns out learning effectiveness is actually higher in the self-confirming case than in the optimistic econometrician case when the discount factor is close to one. This is because in the self-confirming case the informative action is also the best action.

In incorporating learning about off-path play into learning models there are several key issues that must be addressed. First: how much experimentation is there? Even in the models designed for simple simultaneous move environments some random play is induced by the random (for example, logistic) choice,- and this serves as well for learning about off-path play as intentional experimentation. On the other hand, it seems unlikely that the Fed is willing to do much experimenting with monetary policy, and indeed the Sargent, Williams and Zha (2006) analysis reveals that if the Fed attaches even a very small probability to the Keynesian model being correct they should not experiment with low inflation as the potential damage is very large.

In practice one of the problems with learning about off-path play is that when the game has many information sets a lot of experimentation may be needed to figure out what is going on. A crucial case in point is that of repeated games, to which we turn our attention next.

Cooperation in Repeated Games

One of the most important topics in game theory is that of repeated games. Few economic situations involve players who meet once, interact, and then never see each other again. The relations of consumers and businesses between each other and among themselves usually involve an important element of repetition, as do employment relationships, family relationships and so forth and so on. These games are quite importantly different than a game played a single time: When games take place over time, the possibility of creating incentives through rewards and punishments arises. Undeed the fact that our standard of living is orders of magnitude greater than it would be if we were unable to trade, borrow, lend, and generally cooperate with the people around us should make clear that dynamic incentives are a crucial part of modern (and ancient for that matter) economies. As we have observed, though, learning in these games is complicated by the need to infer causality "off the equilibrium path."

To focus thoughts we begin with the classical repeated prisoner's dilemma game. 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

$$\begin{array}{ccc} & C & D \\ C & 2,2 & 0,3 \\ D & 3,0 & 1,1 \end{array}$$

Each player has D as the unique dominant strategy when the game is played only once, and with a fixed and known terminal period a well-known backwards induction argument implies that the unique subgame perfect equilibrium is for each player to defect in every period regardless of previous play. The infinitely repeated case is quite different. Consider the "grim trigger strategies" of cooperating in the first period and as long as the other player has never defected, and otherwise defecting.. Faced with an opponent who plays this way, the choice is to cooperate and get a payoff of 2 every period, or to make a one time gain of 1 by defecting, followed by a loss of 1 in every subsequent period when cooperation breaks down. Assuming a discount factor of $0 \leq \delta < 1$ we may easily compute the increase in average present value from defecting as $(1-\delta)-\delta$ so that it is optimal to cooperate provided $\delta \geq 1/2$. This is a special case of the celebrated folk theorem as developed by Auman and Shapley (1976), Friedman (1971) Fudenberg and Maskin (1986) and Fudenberg, Levine and Maskin (1994) among others: If some technical conditions are satisfied then any payoff vector that is individually rational for the players is a subgame-perfect equilibrium provided that the players are sufficiently patient.

What we see is that repeated games allow the possibility of cooperation in which incentives are established through future rewards and punishments, but they allow many other possibilities as well. Moreover, Fudenberg, Levine and Maskin (1994) show that the folk theorem payoffs can be approximated arbitrarily closely by equilibrium in totally mixed strategies that satisfy all "standard" equilibrium refinements. In this sense the folk theorem poses a problem for game theory : What predictions should we make about repeated games with patient players? A common intuition is that people cooperate whenever there is a cooperative equilibrium and this is often assumed in applied theory and theoretical industrial organization. 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), Dal Bo and Pujals (2015). Unfortunately there is little hard empirical support for this prediction, and laboratory evidence leans against it. At the same time, though, the laboratory does provide a way to better understand what does happen in repeated games, and suggests that learning plays a key role.

Laboratory Play of Infinitely Repeated Games

How successful are players in establishing cooperation in infinite horizon repeated games? Laboratory subjects do respond to game parameters such as continuation probability and the payoff matrix, but it is not the case that players always succeed in cooperating whenever there is an equilibrium that supports cooperation.[27] Indeed, early work on repeated games by Roth and Murnighan (1978) and Murnighan and Roth (1983) found that cooperation rates in the infinitely repeated prisoner's dilemma are very insensitive to the discount factor and in particular to whether the discount factor was high enough to allow a cooperative equilibrium. However, these papers only had the subjects play a single repetition of the repeated game, so the subjects had no chance to learn from experience. Dal Bo (2005) re-launched the experimental study of repeated games by letting subjects play 7-10 iterations of the repeated game, with a different partner each time; he found that this led to much more effect of the discount factor, with higher discount factors generatig more cooperation, as a first cut application of the theory would suggest. There have now been a great many papers on this topic; there are 11 of them with 28 conditions in the meta-analysis of Dal Bo and Frechette (2015). One main take-away from their meta-analysis is that the discount factor matters much more once subjects have played the game a few times.

The second main take-away is that in contrast to predictions based on the set of equilibria, the loss when the other player defects matters too. That is, whether there is an equilibrium in which players cooperate every period is independent of the loss a player incurs when he cooperates and the other defects, because in such an equilibrium players assign probability 1 to their opponent playing C, and the continuation play after any defection can be taken to be (D,D). In contrast, the loss to (C,D) matters in practice, because subjects cannot be sure what strategies their partners are using, and even in treatments where most eventually cooperate some subjects do play 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" and "Always Defect" (AllD)Axelrod and Hamilton (1981) , so that the payoff matrix is

$$\begin{array}{ccc} & TfT & AllD \\ TFT & 2,2 & \delta, 3(1-\delta)+\delta \\ AllD & 3(1-\delta)+\delta, \delta & 1,1 \end{array}$$

They propose that players will only cooperate if TfT risk dominates AllD , 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 Bo and Frechette meta-study; to get more precise predictions Dal Bo and Frechette relate cooperation to the size of the "basin" of TfT, which is the probability of the opponent playing AllD that leaves the player just indifferent between TfT and AllD. However, even in "cooperative treatments" - where the payoff matrix and discount factor lead to a high rate of cooperation - Dal Bo and Frechette estimate that about 10-20% of the subjects seem to be using the strategy "always defect." There are similarly large shares of AllD in the "cooperative treatments" of the noisy prisoners's dilemma in Fudenberg, Rand and Dreber (2012), who also compute the payoffs of the estimated strategies against the

[27] In some cases subjects also cooperate when the prisoner's dilemma is played with a known finite horizon, but this cooperatopn typically unravels from the end period backwatrs when the subjects play the finitely repeated game multiple times with anonymous re-matching; see Embrey, Frechette, and Yuksel (2015).

estimated distribution, and find that the "always defect" players do substantially worse than the conditional cooperators. This heterogeneity of play does not seem to reflect social preferences, and is uncorrelated with how subjects play in a dictator game Dreber, Fudenberg and Rand (2014), so we interpret it as showing that subjects 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.

5 Social Preferences and Political Contests

Collective decisions, such as decisions about who will determine public policies such as taxes and transfers, are often made through political contests such as lobbying, voting, or even warfare. Moreover, the passengers of flight 93 apparently held a vote on whether to attack the hijackers.[28] Formal models of political contests are inevitably game-theoretic, but this raises the question of the identities of the players. Although the contests are often between between groups, if we want games based on individual behavior, we must model the behavior of those individuals. Traditionally political economy has not taken this route, rather modelling each group as a single player much as we often model a business firm or family as a single decision maker.[29] More recent work has begun to look into the "black box" of how group behavior depends on the behavior of the individuals making up the group.

Political Contests

To understand how game theory succeeds and fails in helping us understand political contests let us start by viewing the groups as single players and focusing on the simple prototypical example of a winner-take-all contest. We suppose that there are two players $i = 1, 2$ each of which determines the level of resources $b^i \geq 0$ to devote to the political contest. The player i who commits the greatest level of resources wins the contest and receives a prize worth V^i receiving a utility of $V^i - b^i$ while the loser $-i$ gets no prize and receives utility $-b^{-i}$. In case of a tie each player has an equal chance of winning.

This game is known as the all-pay auction and has been studied by many authors including Dasupta (1986), Hillman and Samet (1987), Hillman and Riley (1989), Barut and Kovenock (1998), and Hillman and Riley (2006). Traditionally the resources commited by player i are referred to as the bid. There is a unique Nash equilibrium and as is often the case when objectives are in conflict it necessary involves randomization. The player i with the highest value $V^i > V^{-i}$ is advantaged and player $-i$ disadvantaged. The advantaged player bids uniformly in $[0, V^{-i}]$. The disadvantaged player bids 0 with probability $1 - V^{-i}/V^i$ with the remaining probability being uniformly distributed on $[0, V^{-i}]$. The equilibrium utility of the disadvantaged player is 0 and that

[28] See USA (2004).

[29] Two prominent examples among many others are Becker (1983)'s work on lobbying and theAcemoglu and Robinson (2001) model of the extension of the franchise.

of the advantaged player is $V^i - V^{-i}$ which is exactly the same as it would be in the unique weakly dominant equilibrium of a second price auction. One simple conclusion we can draw from this model concerns rent dissipation: if the value of the prize is the same to both players then the entire value of the prize is dissipated through the resources wasted in the contest - only if one party is greatly advantaged is it able to benefit in a significant way from the contest.

This model is a special case of a broader and also widely studied class of political contests in which the higher bid does not necessarily win, but merely increases the chances of winning. For example, in voting various random factors such as the intervention of courts, the ways votes are counted, and the weather in particular voting districts, mean that the player who commits fewer resources may still have a chance of winning. In warfare the fact that superior forces do not guarantee victory should be reasonably self-evident. One widely used conflict resolution function is the Tullock specification, in which a player bidding b^i has a probability of winning equal to

$$\frac{(b^i)^\alpha}{(b^i)^\alpha + (b^{-i})^\alpha}$$

where $\alpha \geq 0$. If $\alpha \leq 1$ this function is concave and all Nash equilibria are in pure strategies; as $\alpha \to \infty$ the function approaches the all-pay auction in the sense that a slightly higher bid results in a probability of winning close to 1 and equilibrium must be in mixed strategies.[30] Two empirical studies using this type of approach are the Coate and Conlin (2004) study of referendum voting in Texas and the Esteban, Ray and Mayoral (2012) study of ethnic conflict. Roughly speaking these models seem to do a good job empirically in predicting in the one case the outcome of elections and in the other the level of ethnic strife.

So far so good on the game-theoretic front. An important problem is studied using standard methods of game theory and good results are obtained. The problem is that as we have described the model it is a contest between players while in fact the contests are between groups of players, and sometime the groups are fairly large. For example in the case of farm lobbying there are over two million farms in the United States, so that when we speak of the "farm lobby" we are in fact speaking of a great many people. The problematic aspect of this is that when we construct the game based on the individual members of the group we immediately see that there is an enormous public goods problem - in voting theory this is called the paradox of voting. The chances of an individual vote changing the outcome of an election are so small that the incentive to vote is negligible, - so why does anybody bother to vote? Similarly, why do farmers contribute to lobbying efforts when their individual effort makes little difference? Everybody of course would like their group to win the contest, but - from a purely selfish point of view - they would prefer that everyone else contribute to the effort while they do not.

[30] See, for example Levine and Mattozzi (2016).

A Public Good Game

We now want to peek into the black box of group decision making. To focus thinking, let us again study a simple game that takes place within a group. Suppose there are $i = 1, 2, \ldots, N$ group members each of whom chooses an effort level $\alpha^i \in [0, 1]$ and receive utility $u^i(\alpha) = g^i(\alpha^{-i}) - c^i(\alpha^i)$ where these functions are continuous and the cost $c^i(\alpha^i)$ is strictly increasing with $c^i(0) = 0$. Here no individual benefits from her own contribution - a reasonable approximation where the benefits are spread over a large group. Formally this model is similar to the coconut production model of Diamond (1982) or the gift-giving model of Levine and Pesendorer (2007): each group member provides effort which is useful only to other group members - it represents in effect a gift to those other members. This is a multi-party prisoner's dilemma game and in one-time play standard game theory tells us to expect the unique dominant strategy equilibrium: nobody contributes anything.

There are two main ways to explain the fact that we do observe lobbying, voting and wars despite a theory that seems to predict we should not - we can either change the preferences of the agents, leaving the game tree fixed, or we can model these group decisions with a different extensive form game. Political economists have focused on two quite simple forms of social preference: these can be described as altruism or ethical behavior and they have also considered models where groups cooperate due to peer pressure or coercion.[31]

Altruism

The oldest and most traditional model of social preference in economics is that of altruism. Here is a particular specification from Esteban, Ray and Mayoral (2012): Define the group utility to be $U(\alpha) = \sum_i u^i(\alpha)$ and suppose each player i's objective function is $(1 - \lambda)U(\alpha) + \lambda u^i(\alpha)$ where $0 \leq \lambda \leq 1$ is a measure of how selfish they are. Eliminating irrelevant constants this is equivalent to maximizing $v^i(\alpha) = (1 - \lambda)U(\alpha) - \lambda c^i(\alpha^i)$, so we see that players are willing to bear the cost of contributing if the group benefit is large enough.

Although many papers on voting also use variations on this model, for example, Schram and Sonnemans (1996), Fowler (2006), Fowler and Kam (2007), Jankowski (2007), Edlin, Gelman and Kaplan (2007), Faravelli and Walsh (2011), andEvren (2012), there are some important issues with this approach. First, in order to scale properly it is necessary that preferences depend on the total utility of others (and not, for example, on their average utility). This implies that if you double group size you double the incentive to contribute - and this is not entirely plausible. Moreover, while altruism towards people we know well may be strong, in the anonymous setting of the laboratory the forces of altruism are weak[32] and, for example, the altruism of farmers towards millions of other farmers who live in other states does not seem likely to be strong.

Second, this model of altruism allows multiple and Pareto- inefficient equilibria. For example

[31] See also the more detailed review of the literature in Gintis (2015). Economists have also introduced a variety of theories to explain more sophisticated social behavior such as retaliation and reciprocity - see Fehr and Schmidt (1998) and Levine (1998).

[32] See, for example, Levine (1998).

in the simple 2x2 coordination game with payoff matrix [33]

$$\begin{array}{ccc} & L & R \\ L & 2,2 & 0,0 \\ R & 0,0 & 1,1 \end{array}$$

despite the fact there is no conflict of interest, there are three equilibria: L, L, R, R and a symmetric mixed equilibrium where the probability of L is 1/3rd. In a political setting it seems likely that groups might be able to solve this coordination problem and agree on L, L. A concept called rule-consequentialism introduced by Harsanyi (1977) proposes a particular solution to this problem that is used in a mumber of papers on voting literature[34] where it is often called the ethical voter model. The idea is that each group member asks what would be in the best interest of the group, that is, what vector α would maximize $U(\alpha)$? Then having determined this hopefully unique solution. each member "does their part" by playing their own piece α^i of the solution. Conceptually this is supposed to capture the duty or social obligation of voting - the idea that it is unethical to free ride. However note that consequentialist models assume that $\lambda = 0$ so have difficulty grappling with the fact that in many circumstances people are altruistic, but only partially so.

Alternate within-Group Extensive Forms: Peer Pressure and Coercion

As a practical matter public goods problems are generally overcome in part through coercion. In some cases coercion may have a specific legal form, such as tax laws, mandatory voting laws, or a military draft. In other cases there is no explicit legal coercion - for example a social norm among police not to report wrong-doing by other police - yet sociologists such as Coleman (1988) argue that peer pressure plays a key role in providing discipline that can overcome public goods problems. For a game theorist reading the sociology literature it is natural to interpret peer pressure in the form of social approval, social disapproval, and social exclusion as a punishment or reward.[35] We are probably familiar with this in our everyday life and Coleman (1988) discusses specifically the case of honesty among diamond merchants who all belong to the same social circle and would suffer an enormous loss if excluded. Much of Elinor Ostrom's work, especially Ostrom (1990), documents how public goods problems are overcome by peer monitoring and appropriately scaled peer punishments for free-riding. Coleman and Ostrom as well as Olson (1965) argue that - within the limits of available monitoring and punishment - these peer pressure mechanisms do a good job of solving public goods problems.

Levine and Modica (2016) provide a simple model incorporating the idea of peer monitoring and punishment chosen to achieve group objectives. Here a group would like to collude to maxi-

[33] This can be interpreted as which side of the road to drive on in a place like Hong Kong where the physical infrastructure such as highway on and off ramps favors driving on the left.

[34] See, for example, Riker and Ordeshook (1968), Coate and Conlin (2004), Feddersen and Sandroni (2006), Li and Majumdar (2010), Ali and Lin (2013).

[35] This is in contrast to models of social conformity such as Akerloff and Kranton (2005) which do not explicitly consider punishments or rewards.

mize $U(\alpha) = \sum_i u^i(\alpha)$, and does so by agreeing on a punishment for group members who do not contribute. Specifically, they suppose that the group agrees on a social norm $a \in [0,1]$, and has a monitoring technology which generates a noisy signal of whether or not a member complies with the norm. Suppose the signal is $z^i \in \{0,1\}$ where 0 means "good, followed the social norm" and 1 means "bad, did not follow the social norm." Suppose further that if member i violates the social norm by choosing $\alpha^i \neq a^i$ then the signal is 1 for sure while if he adhered to the social norm ($\alpha^i = a^i$) then the signal is 1 with probability π. Play takes place in two stages: first individual group members choose actions α^i. Then signals about each group member are seen. When the bad signal is received the group member receives a punishment of size P^i.[36] For the social norm a to be incentive compatible we need $-\pi P^i \geq c^i(a^i) - P^i$ which is to say $P^i \geq c^i(a^i)/(1-\pi)$. If all agents adhere to the social norm there will still be punishments due to the erroneous bad signals; the social cost of the punishment is πP^i, and we suppose that the group can agree to minimize this cost so that it will choose $P^i = c^i(a)/(1-\pi)$. The resulting cost is then $c^i(a^i)\pi/(1-\pi)$. Let $C(a) = \sum_i c^i(a^i)$. The utility of the group taking account of enforcement costs is then $U(a) - C(a)\pi/(1-\pi)$ which is equivalent to $V(a) = (1-\pi)U(a) - \pi C(a)$. [37]

Observe that if a is the solution to maximizing $(1-\pi)U(a) - \pi C(a)$ then it must be that a^i maximizes $v^i(a) = (1-\pi)U(a) - \pi c^i(a^i)$. Consequently peer pressure leads to a model predicting the same level of contributions as altruism (with cost scaled by π) although the altruism model does not predict that punishment will be observed.

Also, the peer pressure model makes predictions about changes in behavior resulting from changes in monitoring technology a which the model of altruism says should not matter. Similarly the peer pressure model makes predictions about which groups are likely to be effective that are absent in the altruism models. Specifically the theory suggests that free-riding problems will be most difficult to overcome in settings where monitoring and punishment are costly. Another example involving traffic is instructive. On entrances to highways people typically follow the selfish optimum of entering as quickly as possible. When traffic is heavy this creates strong negative externalities by hindering other motorists, and traffic studies such as McDermott (1967) have shown that forcing motorists to wait briefly before entering by means of traffic metering significantly improves traffic flow and commute times. A model of altruism predicts that motorists would voluntarily delay entry to at least partially offset the negative impact on other motorists, which as far as we are aware does not occur. A model of peer monitoring and punishment predicts that because the cost of both monitoring and punishment are very high a social norm of delaying entry will not emerge.

We have so far discussed one-off group decision problems. Often in practice these sorts of problems are repeated. There are two main ways that repeated play could matter. First, even without peer pressure or coercion, contribution can be supported in a repeated public goods game by straregies of the form "Stop contributing if last period's contributions were too low." With these strategies, everyone in the group is punished, and no particular individual is singled out. In a small

[36] Here the coercion takes the form of punishment, but it could equally well be the witholding of a reward.

[37] The latter form enables us to account for the case $\pi = 1$ when the signal has no information so the group must choose to provide no effort.

population of say large banks, who engage in ongoing lobbying efforts, these sorts of equilibria might be plausible. [38]However, with a large population and noise in observing individual actions, a common punishment applied to the whole group leads to an anti-folk theorem (see Levine and Pesendorfer (1995)) which asserts that little incentive against free riding can be provided.[39] For example, it is not very plausible that a voter participates in a presidential election due to a fear his personal failure to vote could trigger a massive decline in future voter turnout. To understand why this must be so observe that to provide incentives free-riding by a single individual must be able to trigger the common punishment. When individual free-riding is only imperfectly observed and the population is large, the chances that an erroneous observation of free-rising will trigger the common punishment becomes close to one. However if the common punishment is almost certainly going to be triggered by someone else no matter what you do you hardly have any incentive not to free-ride.

The other way that repeated play could matter is to enforce the application of punishments targeted at individual deviators, and indeed this is the base model studied in Levine and Modica (2016). However, if the monitoring and punishment are costless to the monitors then punishment is an equilibrium without repetition. Here we have made that assumption for simplicity. Levine and Modica (2016) show that the big picture of enforcement costs that depend on monitoring parameters is not greatly different when monitoring and punishment are costly to the monitors and the monitoring game is repeated - although naturally the cost to the group of operating the scheme is higher. Hence, for example, farmers live in farm communities and socialize with other farmers, so the social network underlying the farm lobby is strong. By contrast we might think of urbanites as the "anti-farm lobby" and as this broad and diverse group is not closely connected through a social network it is less able to overcome the problem of free-riding.

The Rate of Change of Social Norms

Social behavior has been studied extensively in the laboratory (see, for example, Fehr and Gachter (2000)) and we observe and can measure social norms of trust, altruism, fairness, retaliation and spite. Typically in laboratory studies we try to remove the possibility of explicit post-experiment peer punishment by providing participants with a high degree of anonymity so as to isolate social preferences. But is this indeed what we are seeing? That is, social norms that are enforced through peer pressure and other means may be partially internalized, meaning that group members "punish themselves" through guilt, shame, and the inability to perfectly lie about their behavior after the fact. No doubt part of our social norms are intrinsically part of our biological make up, part are internalized and part are enforced. We do not have a good empirical grasp of the relative importance of these three, nor is laboratory work especially useful in helping us understand their relative importance or how quickly social norms change. We do observe enormous variation in

[38]Acemoglu and Jackson (2014)extend this idea to a setting where each agent lives for two periods and interacts with agents from the previous and next generations via a coordination game.

[39]Fudenberg, Levine, and Pesendorfer (1998) extend this result to game with non-anonymous small players, provided that the number of "punishment actions" is bounded as the number of small players grows,

social norms measured in the laboratory if we look across different societies.[40] This indicates a high degree of mutability in the sense that - given enough time - different social norms do emerge and suggests therefore that they are indeed endogenous. We find evidence of this also in the nature of the social norms that are observed. For example Henrich et al (2001) study play of the ultimatum bargaining game in different societies. Here there is a first mover who is given a fixed budget and must choose how much to offer to the second mover. The second mover may either accept the offer, in which case the budget is divided as agreed upon, or reject it, in which case neither player gets anything. In most societies bad offers are likely to be rejected and the first mover generally asks for and gets somewhat more than half the budget. One society is an extreme outlier: good offers are rejected and bad offers accepted. As it happens this makes perfectly good sense if we understand the social norms outside the laboratory in that society: It is a competitive gift-giving society, where social status is determined by the size of gift given to others, so a good offer is in effect an insult. Moreover this social norm is very functional: this primitive society lives by whale hunting, and as whales are only rarely caught, for the society to survive there must be an insurance mechanism by which the lucky hunters provide whale to the unlucky hunters. Competitive gift giving in one such insurance mechanism. This functionality of social norms is consistent with the ideas and observations of Coleman (1988), Ostrom (1990) and Olson (1965). However, we are left with the key question of how long it takes for functional social norms to emerge in a particular setting.

As Binmore (1998) has emphasized, social norms can often be understood as Nash equilibria. This is true in particular when they rely on explicit punishments, as in the peer punishment model of Levine and Modica (2016). As with any equilibrium, these equilibria must be learned, and there are certainly evolutionary forces that would push towards more functional social norms over less functional social norms.[41] This puts the question of social norms on the same footing as the question we addressed in the first half of this essay: how quickly does learning and adjustment take place? Do social norms change rapidly through some sort of agreement explicit or implicit in response to changed circumstances? Or are we essentially the victims of social norms that made sense long ago but no longer do?

As is the case with learning about equilibrium, there is evidence both for very rapid and for very slow learning about social norms.

Let us start with an example of a very rapid change in social norms. Throughout most of the history of East Germany the social norm was to support the government , and this was enforced in various ways, not least by ratting out those who spoke against the government to the secret police. Moreover, from the perspective of East Germans as a collective this may have been a functional social norm , as had they succeeded in overthrowing the government the result might well have been the same as in Hungary in 1956 or in Czechoslovakia in 1968, which is to say, invasion by the Soviet Union and an even more repressive government put in place. However on July 6, 1989, Mikail

[40] See especially Henrich et al (2001).

[41] See, for example, Ely (2003), Levine and Modica (2014) as well as the broader literature on group selection reviewed, for example in Bergstrom (2001).

Gorbachev gave a speech indicating that the Soviet Union would not invade in response to changes in government. Within four months the citizens of East Germany had successfully overthrown the existing government and the Berlin Wall fell. In a lab setting, Peysakhovich and Rand (2014) are able to create different social norms in "laboratory time" which is to say minutes and not months. They show that participants who play repeated prisoner dilemmas that do or do not support cooperation learn social norms that carry over to subsequent one-shot games. In particular participants from environments that support cooperation are subsequently more prosocial, more likely to punish selfishness, and more generally trusting.

In the opposite direction we have the striking work of Bigoni et al (2013). They study the North-South divide in Italy, conducting experiments about social norms in four cities. They find that social norms of trust and cooperation are stronger in the North. The more striking finding however is that the only historical difference that correlates with the details of which cities have the strongest social norms are the frequency of warfare more than 150 years ago. Roughly speaking it seems that those cities that faced a higher frequency of conflict developed stronger norms of trust and cooperation in order to prevail in those conflicts. While based on a very small sample - this does suggest a remarkable stickiness in social norms.

6 What constitutes a good theory?

Where are we and where do we want to go? In understanding where we are, the political contest papers of Coate and Conlin (2004) and Esteban, Ray and Mayoral (2012) are helpful. These are recent empirical papers addressing seemingly rather different issues, voter participation in Coate and Conlin (2004) and ethnic conflict in Esteban, Ray and Mayoral (2012) case. Both papers use similar game theoretic models of a sophisticated sort. We can conclude form this that game theory is scarcely dead and indeed it is a basic working tool for economists. Moreover, it is not a purely theoretical too, but is used successfully for empirical investigation. The fact that game theoretic models of different situations have a similar structure - voting and warfare in these two papers - highlights one of the lesser known strengths of game theory: It enables us to bring lessons learned from one domain to another domain.

These papers also make clear that there are important practical circumstances where we need to explicitly introduce behavioral forces and in particular social preferences into our models. We see that there are a great many such models and we do not have a clear idea which is "correct." It is both good and bad that in some cases, such as the formation of social norms within groups, results of different models are similar. On the one hand it means we may have some confidence in our conclusions. On the other it does not do much to help us understand which model is best for moving ahead. Similarly we have a number of different learning models, and no clear conclusion as to which is "right." Moreover some of the learning models seem to be right in some circumstances and others in other circumstances, yet we do not have a good understanding of which is which.

So where do we go from here? Perhaps a good place to start is to think about what we want in a good theory, so we can then think about what we need to do to get there. First and foremost we want precise and valid predictions. Our impression is that existing game theoretic methods often do pretty well on the validity front, at least when we use appropriately robust versions of the theory. However, precision means that we make precise statements about what will happen: saying "there are lots of equilibria and we do not know which one will occur" is an imprecise theory. 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.

Second we need theoretical simplicity. A long-standing analogy between globes and models may help here. A good globe lets you have a sense of the Earth as a whole. A globe that is the size of a room is already impractical, though of necessity it will omit some details, and a globe on a 1-1 scale is pointless: We might as well wander around the actual surface of the earth as use a map drawn on 1-1 scale. At our current level of technology it is not feasible to model human behavior by analyzing the laws of physics that govern the movement of atoms and molecules in our body and brain. Similarly neuro-scientific models of the brain are poorly suited to introduce into a model of many millions of interacting people. As time goes on we may be able to deal with more sophisticated models of human behavior, but is likely always to be the case that simple models will be helpful for understanding and intuition about how more elaborate models work.

Third is the e need for breadth. A single unified theory is subject to a great deal of possible falsification and testing. On the other hand a theory designed to explain only one fact cannot be tested at all and we think that economics should not just about cross-correlating environments with possible explanations. As Ellison (2006) says when discussing behavioral industrial organization :

> Think of the set of behavioral biases as the column headings, and put all of the standard models in IO as the row headings: how will a monopolist price, how will a monopolist selling durable goods price; how will a monopolist price discriminate; how will oligopolists selling differentiated goods set prices; how will some action be distorted to deter or accomodate entry, etc. It takes little knowledge or imagination to come up with literally thousands of paper topics...Will this lead to thousands of behavioral IO papers in the next few years? I am pretty sure (and hope) the answer is no.

In addition to these desiderata for theories, we would like to argue for a methodological value, namely that of conservatism, meaning that we should try to preserve well-established and useful tools. We are going to throw out things like expected utility and exponential discounting, for example, only if this is essential for empirical validity. And we must also be conservative in that preserve the conclusions of existing theory where it works well. There is absolutely no point in introducing a 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

To summarize, we argue that we need to improve learning models to better capture the idea that specific events trigger reevaluation of models and can lead to very rapid learning - possibly hidden Markov models will prove a suitable way of doing this. With respect to social preference we have a range of models ranging from models of signaling intent Levine (1998), to models of impulsive behavior (Dreber et al (2015)). The paper of Fehr and Schmidt (1998) is important because it points in the right direction: a simple model broadly applied not to a single observation but to many different experiments including ones we already understand well. While their specific model has not survived the test of time, it does provide a good road map for where we should go

References

Acemoglu, Daron, and Matthew O. Jackson (2014): "Social norms and the enforcement of laws," NBER w20369.

Acemoglu, Daron, and James A. Robinson (2001): "A theory of political transitions," *American Economic Review*: 938-963.

Akerloff, G. and R. E. Kranton (2005): "Identity and the Economics of Organizations," *Journal of Economic Perspectives.*

Ali, S. Nageeb and Charles Lin (2013): "Why People Vote: Ethical Motives and Social Incentives," *American Economic Journal: Macroeconomics* 5:73–98.

Aumann, R. and L. Shapley [1976]: "Long Term Competitin: A Game Theoretic Analysis," unpublished notes.

Axelrod, R. and W. D. Hamilton (1981): "The evolution of cooperation," *Science.*

Bajari, P., H. Hong, H. and D. Nekipelov (2013):" Game theory and econometrics: A survey of some recent research. " In *Advances in Economics and Econometrics: Tenth World Congr*ess.

Barut, Yasar, and Dan Kovenock (1998): "The symmetric multiple prize all-pay auction with complete information," *European Journal of Political Economy* 14: 627-644.

Battigalli, P. and D. Guatoli (1997): "Conjectural Equilibria and Rationalizability in a Game with Incomplete Information," in *Decisions, Games and Markets*, Kluwer Academic Publishers, Norwell, MA.

Becker, Gary S. (1983): "A theory of competition among pressure groups for political influence," *Quarterly Journal of Economics*: 371-400.

Bergstrom, Ted (2001): "Evolution of social behavior: individual and group selection models, " mimeo Department of Economics, UCSB.

Benaïm, M., Hofbauer, J. and Hopkins, E. (2009): "Learning in games with unstable equilibria," *Journal of Economic Theory* 144: 1694-1709.

Bigoni, Maria, Stefania Bortolotti, Marco Casari, Diego Gambetta, Francesca Pancotto (2013): "Cooperation Hidden Frontiers: The Behavioral Foundations of the Italian North-South Divide," University of Bologna.

Binmore, K. G. (1998): *Game theory and the social contract: just playing* (Vol. 2), MIT Press.

Binmore, Kenneth G., and Larry Samuelson (1992): "Evolutionary stability in repeated games played by finite automata," *Journal of Economic Theory* 57: 278-305.

Blackwell, David (1956): "An analog of the minimax theorem for vector payoffs," *Pacific Journal of Mathematics* 6: 1-8.

Block, Juan, Drew Fudenberg and David K. Levine (2016): “The Evolution of Conventions Revisited,” mimeo.

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: 164-192.

Brown, George W. (1951): “Iterative solution of games by fictitious play," *Activity Analysis of Production and Allocation* 13: 374-376.

Camerer, Colin, and Teck-Hua Ho (1999): Experienced-weighted attraction learning in normal form games." *Econometrica*: 827-874.

Cason, T., Friedman, D., and Wagener F. (2005): “The Dynamics of Price Dispersion or Edgeworth Variations,” *Journal of Economic Dynamics and Control* 29: 801-822;

Chen, W., Y. Chen and D. K. Levine (2015): ”A Unifying Learning Framework for Building Artificial Game-Playing Agents," *Annals of Mathematics and Artificial Intelligence* 73: 335-358.

Cheung, Yin-Wong, and Daniel Friedman (1997): “Individual learning in normal form games: Some laboratory results," *Games and Economic Behavior* 19: 46-76.

Cho, I.K. and K. Kasa (2015): “Learning and Model Validation,” *Review of Economic Studies* 82: 45-82.

Coate, S., M. Conlin (2004): “A group rule–utilitarian approach to voter turnout: theory and evidence,” *American Economic Review* 94: 1476–1504.

Coleman, James S. (1988): “Social capital in the creation of human capital," *American Journal of Sociology*: S95-S120.

Cournot, A (1883): *Recherches sur les Principes Mathematiques de la Theorie de la Richesses*.

Bó, Pedro Dal (2005): “Cooperation under the shadow of the future: experimental evidence from infinitely repeated games," *American Economic Review*: 1591-1604.

Bó, Pedro Dal, and Guillaume R. Fréchette (2011): “The evolution of cooperation in infinitely repeated games: Experimental evidence," *The American Economic Review* 101: 411-429.

Dal Bó and Pujals [2015] “The Evolutionary Robustness of Forgiveness and Cooperation,” mimeo.

Dasgupta, P. (1986): “The theory of technological competition,” In: Stiglitz, J.E., Mathewson, G.F. Eds., *New Developments in the Analysis of Market Structure*, MIT Press, Cambridge, pp. 519–547

Diamond, Peter A. (1982): “Aggregate demand management in search equilibrium," *The Journal of Political Economy*: 881-894.

Florian, Michael ed., *Traffic Equilibrium Methods*, Springer-Verlag, 1976.

Domitrovic, Brian (2012): “Chairman Bernanke Buries the Phillips Curve: Bravo!” *Forbes*, April 3.

Dreber, Anna, Drew Fudenberg, David K. Levine and David Rand (2015): “Altruism and Self Control.” mimeo

Dreber Fudenberg and Rand [2014]: “Who Cooperates in Repeated Games: The Role of Altruism, Inequity Aversion, and Demographics,” *Journal of Economic Behavior and Organization* 98: 41-55.

Dutta, R., David K. Levine and Salvatore Modica (2016): "Collusion Constrained Equilibrium," EUI.

Edlin, Aaron, Andrew Gelman, and Noah Kaplan (2007): "Voting as a Ra tional Choice: Why and How People Vote to Improve the Well-being of Others," *Rationality and Society* 19:293–314

Ellison, Glenn (2006): *Bounded rationality in industrial organization*, Econometric Society Monographs.

Ely, Jeffrey C. (2002): "Local conventions," *Advances in Theoretical Economics* 2.

Embrey, Matthew, Guillaume Frechette, and Sergei Yuksel (2015): "Cooperation in the Finitely Repeated Prisoner's Dilemma," http://cess.nyu.edu/frechette/print/Embrey_2014b.pdf

Erev, I. and E. Haruvy (2013): "Learning and the economics of small decisions," Forthcoming in *The Handbook of Experimental Economics*, Vol. 2 Roth, A.E. and Kagel, J. Eds.)

Esteban, J., L. Mayoral, D. Ray (2012): "Ethnicity and conflict: An empirical study," *American Economic Review*

Evreñ Ozgur (2012): "Altruism and Voting: A Large-turnout Result that Does not Rely on Civic Duty or Cooperative Behavior," *Journal of Economic Theory* 147:2124–2157.

Eyster, E. and M. Rabin 2005. Cursed equilibrium. *Econometrica*, 73: 1623-1672

Faravelli, Marco, and Randall Walsh (2011): "Smooth Politicians and Paternalistic voters: a theory of large elections," No. w17397, National Bureau of Economic Research.

Feddersen, T., A. Sandroni (2006): "A theory of participation in elections," *American Economic Review* 96: 1271–1282.

Fehr, E., and Gächter, S. (2000): "Fairness and Retaliation: The Economics of Reciprocity," *The Journal of Economic Perspectives* 14: 159–181.

Fehr, Ernst, and Klaus M. Schmidt (1999): "A theory of fairness, competition, and cooperation," Q*uarterly Journal of Economics*: 817-868.

Foster, Dean P., and Rakesh Vohra (1999): "Regret in the on-line decision problem," *Games and Economic Behavior* 29: 7-35.

Fowler, James H. (2006): "Altruism and Turnout," *Journal of Politics* 68: 674– 683.

Fowler, James H., and Cindy D. Ka (2007): "Beyond the self: Social identity, altruism, and political participation," *Journal of Politics* 69: 813-827.

Friedman, James W. (1971): "A non-cooperative equilibrium for supergames," *The Review of Economic Studies*: 1-12.

Fudenberg, D., R. Iijima, and T. Strzalecki (2015): "Stochastic Choice and Revealed Perturbed Utility," *Econometrica* 83: 371-2409.

Fudenberg, Drew, and David Kreps (1988): "Learning, experimentation, and equilibrium in games," http://scholar.harvard.edu/files/fudenberg/files/theory_of_learning_working_paper_1988.pdf?m=136007066

Fudenberg, D. and D. Kreps (1993): "Learning Mixed Equilibria" *Games and Economic Behavior*, 5: 320-367,

Fudenberg, D. and D. 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: 354-380.

Fudenberg, D. and D. K. Levine (2009): "Self Confirming Equilibrium and the Lucas Critique," Journal of Economic Theory, 144: 2354-2371

Fudenberg, Drew, and David K. Levine (1993a): "Self-confirming equilibrium," *Econometrica*: 523-545.

Fudenberg, Drew and David K. Levine (1993b): "Steady State Learning and Nash Equilibrium" *Econometrica* 61: 547-574.

Fudenberg, Drew, and David K. Levine (1995): "Consistency and cautious fictitious play," *Journal of Economic Dynamics and Control* 19: 1065-1089.

Fudenberg, D. and D. K. Levine (2014): "Learning with Recency Bias," *Proceedings of the National Academy of Sciences* 111: 10826-10829

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

Fudenberg, D., D.K. Levine, and W. Pesendorfer (1998): "When are nonanonymous players negligible?" *Journal of Economic Theory*, 79: 46-71.

Fudenberg, Drew, and Eric Maskin (1986): "The folk theorem in repeated games with discounting or with incomplete information," *Econometrica* 54: 533-554.

Fudenberg D. and E. Maskin (1990): "Evolution and Cooperation in Noisy Repeated Games" *American Economic Review* 80: 274-279.

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. ACM.*

Fudenberg, Drew, David Rand and Anna Dreber (2012):"Slow to Anger and Fast to Forgive: Cooperation in an Uncertain World," *American Economic Review* 102: 720-749.

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

Gintis, Herbert (2015): "Homo Ludens : Social Rationality and Political Behavior"

Gittins., J.C. [1988] *Multi-Armed Bandit Allocation Indices*, John Wiley & Sons, New York.

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

Harsanyi, John C. (1977): "Morality and the theory of rational behavior," *Social Research*: 623-656.

Hart, S., and A. Mas-Colell (2001): "A general class of adaptive strategies," *Journal of Economic Theory* 98: 26-54.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H. and McElreath, R. (2001): "In search of homo economicus: behavioral experiments in 15 small-scale societies," *American Economic Review*: 73-78.

Hillman, A., Riley, J., 1989. Politically contestable rents and transfers. Economics and Politics 1, 17–40

Hillman, A. L. and J. Riley (2006): "Politically Contestable Rents and Transfers," *Economics and Politics* 1: 17-39.

Hillman, A., Samet, D., 1987. Dissipation of contestable rents by small numbers of contenders. Public Choice 54, 63–82.

HHommes, C., Sonnemans, J., Tuinstra, J. and H. Van De Velden (2007): “Learning in cobweb experiments," *Macroeconomic Dynamics* 11: 8-33.

Ho, T. H. , C. F. Camerer, J-K. Chong [2007] “Self Tuning Experience Weighted Attraction Learning in Games,” Journal of Economic Theory 133: 177-198

Hooker, Brad (2011): “Rule Consequentialism,” *The Stanford Encyclopediaof Philosophy.*

Jankowski, Richard (2007): “Altruism and the Decision to Vote Explaining and Testing High Voter Turnout,” *Rationality and Society* 19:5–34.

Kalai, E. and E. Lehrer (1993): “Rational Learning Leads to Nash Equilibrium,” *Econometrica* 61: 1019-1045.

Kaldor, Nicholas (1934): “A Classificatory Note on the Determination of Equilibrium,” *Review of Economic Studies* 1: 122-36.

Kreps, David M., and Robert Wilson (1982): “Sequential Equilibria." *Econometrica*: 863-894.

Larcom, S., F. Rauch, and T. Willems (2015): “The Benets of Forced Experimentation: Striking Evidence from the London Underground Network,” http://users.ox.ac.uk/~econ0360/FerdinandRauch/Tube.pdf

Levine, David K. (1998): “Modeling altruism and spitefulness in experiments." *Review of economic dynamics* 1: 593-622.

Levine, David K. and Wolfgang Pesendorfer (1995): “ When are agents negligible?” *American Economic Review.*

Levine, D. K. and S. Modica (2014): "Dynamics in Stochastic Evolutionary Models," Theoretical Economics, forthcoming

Levine, D. K. and S. Modica([2015): “Peer Discipline and Incentives Within Groups," *Journal of Economic Behavior and Organization*, forthcoming.

Levine, D. K. and A. Mattozzi (2015): “Voter Participation with Collusive Parties," EUI

Levine, David K., and Wolfgang Pesendorfer (2007): “The evolution of cooperation through imitation," *Games and Economic Behavior* 58: 293-315.

Li, Ming and Dipjyoti Majumdar (2010): “A psychologically based mo del of voter turnout,” *Journal of Public Economic Theory* 12:979–1002.

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."

McKelvey, Richard D., and Thomas R. Palfrey (1995): “Quantal response equilibria for normal form games," *Games and Economic Behavior* 10: 6-38.

Milgrom, Paul, and John Roberts (1982): “Predation, reputation, and entry deterrence," *Journal of Economic Theory* 27: 280-312.

Monderer, D. D. Samet and A Sela (1997): “Belief affirming in learning processes," *Journal of Economic Theory* 73: 438-452

Morgan, J., H. Orzen, and M. Sefton (2006): "An Experimental Study of Price Dispersion," *Games and Economic Behavior*.

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: 279-300.

Myerson, Roger B. (1978): "Refinements of the Nash equilibrium concept." *International Journal of Game Theory* : 73-80.

Myerson, R.B. (1991): *Game Theory,* Harvard university press.

Nachbar, John H.(1987): "Prediction, optimization, and learning in repeated games," *Econometrica*: 275-309.

Olson Jr., Mancur (1965): *The Logic of collective action: public goods and the theory of groups*, Harvard Economic Studies.

Osborne, M.J. and Rubinstein, A. (1994): *A Course in Game Theory*, MIT press.

Ostrom, Elinor (1990): *Governing the commons: The evolution of institutions for collective action*, Cambridge university press.

Peysakhovich, A. and D. Rand (2014): "Habits of Virtue: Creating Norms of Cooperation and Defection in the Laboratory," mimeo.

Riker, William H. and Peter C. Ordeshook (1968): "A Theory of the Calculus of Voting," *American Political Science Review* 62: 25–42

Roemer, John (2010): "Kantian Equilibrium," *Scandinavian Journal of Economics* 112:1–24.

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: 164-212.

Roth, Alvin E., and J. Keith Murnighan (1978): "Equilibrium behavior and repeated play of the prisoner's dilemma," *Journal of Mathematical Psychology* 17: 189-198.

Salmon, Timothy C. (2002): "An evaluation of econometric models of adaptive learning," *Econometrica* 69: 1597-1628.

Sims, Christopher A. (2003): "Implications of rational inattention," *Journal of monetary Economics* 50: 665-690.

Sargent, T. (1999): *The Conquest of American Inflation*, Princeton University Press, Princeton NJ.

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

Schram, Arthur, and Joep Sonnemans (1996): "Why people vote: Experimental evidence." *Journal of Economic Psychology* 17: 417-442.

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: 47-70.

Shapley, Lloyd (1964): "Some Topics in Two Person Games," *Annals of Mathematical Studies* 5: 1-28.

Smith, Vernon L. (1962): "An experimental study of competitive market behavior," *The Journal of Political Economy* : 111-137.

Sutan, A. and M. Willinger [2004]: “Coordination In Cobweb Experiments With(out) Belief Elicitation” mimeo BETA-Theme

USA (2004): *9/11 Commission Report*.

Van Huyck, John B., Raymond C. Battalio, and Richard O. Beil (1990): “Tacit coordination games, strategic uncertainty, and coordination failure," *The American Economic Review*: 234-248.

von Neumann, J., and O. Morgenstern (1944): *The Theory of Games and Economic Behavior*, Princeton, Princeton University Press.

Wilcox, Nathaniel T. (2006): “Theories of learning in games and heterogeneity bias," *Econometrica* 74: 1271-1292.