



ELSEVIER

Available online at www.sciencedirect.com

 ScienceDirect

Journal of Economic Theory 144 (2009) 2354–2371

JOURNAL OF
**Economic
Theory**

www.elsevier.com/locate/jet

Self-confirming equilibrium and the Lucas critique

Drew Fudenberg, David K. Levine*

*Departments of Economics, Harvard University and Washington University in St. Louis, Campus Box 1208,
St. Louis, MO, United States*

Received 10 August 2007; accepted 6 July 2008

Available online 24 November 2008

Prepared for the Conference in Honor of Robert E. Lucas Jr.

Abstract

We examine the role of off-path “superstitions” in macro-economics, and show how a false belief about off-path play is the key element underlying both the Lucas Critique and the game-theoretic concept of self-confirming equilibrium. However, the impact of false beliefs in these two cases is different: In the Lucas case, a policy maker’s incorrect beliefs about off-path play can lead to the adoption of mistaken policy innovation. However, the consequences of such an innovation provide evidence that the belief that motivated them was wrong. In contrast, play may never escape an undesirable self-confirming equilibrium, as the action implied by the mistaken belief does not generate data that contradicts it; escape from the self-confirming equilibrium requires that players do a sufficient amount of experimentation with off-path actions.

© 2008 Elsevier Inc. All rights reserved.

JEL classification: A1; A2

Keywords: Lucas Critique; Self-confirming equilibrium; Superstitions

* Corresponding author.

E-mail addresses: dfudenberg@harvard.edu (D. Fudenberg), david@dklevine.com (D.K. Levine).

“The fact that nominal price and wages tend to rise more rapidly at the peak of the business cycle than they do in the trough has been recognized from the time when the cycle was first perceived as a distinct phenomenon. The inference that permanent inflation will therefore induce a permanent economic high is no doubt equally ancient, yet it is only recently that this notion has undergone the mysterious transformation from obvious fallacy to the cornerstone of economic policy.”

Robert E. Lucas Jr. [23]

1. Introduction

Thinking of equilibrium as the result of non-equilibrium learning suggests that players are likely to be better informed about the consequences of actions on the equilibrium path than off the equilibrium path. Indeed, if play converges to equilibrium and players have many observations, then we should expect players to have correct beliefs about the equilibrium outcome.¹ But by definition, off-path actions are never observed in equilibrium, which raises the possibility that incorrect beliefs about off-path play might persist for quite some time. Following Fudenberg and Levine [15], we call a belief that is objectively false a “superstition.”²

There are two manifestations of the role of off-path superstitions. One is the sort of “econometric policy evaluation” that was the focus of the famous Lucas critique: namely, a behavioral parameter that is estimated under one government policy may not be invariant to changes in that policy. Moreover, the false belief that the parameter is policy-invariant may lead to the adoption of mistaken government policies. For example, econometric estimation on the equilibrium path may make certain types of poor economic policies (“permanent inflation”) appear desirable.³

Although the macroeconomic literature has focused on the Lucas critique, the opposite sort of off-path superstitions can be longer-lasting, and thus potentially worse: Some desirable policies may fail to be adopted because of superstitions that make them appear undesirable. This idea is the basis of self-confirming equilibrium. In the Lucas critique case, we may expect that if a policy is implemented based on superstitions about its consequences, we will learn that these superstitions are false, and the defective policy will be dropped. Indeed – the argument that permanent inflation might be a desirable policy seems to have dropped off the radar since the mid-70s. On the other hand, if a policy is not implemented at all because of a superstition about its consequences, no new data is generated, and the situation is likely to persist – which is why a “self-confirming equilibrium” is an “equilibrium.”

Our main goal here is to illustrate the basic concepts of superstitions and self-confirming equilibrium through a series of examples inspired by the macroeconomic literature. We also use a simple two-armed bandit model to examine how quickly we might expect mistakes to be

¹ At least if players observe the path of play, which we will assume throughout this essay. See Dekel et al. [7] for a discussion of some cases where players observe less information than this.

² Fudenberg and Levine [15] analyze off-path experimentation in greater detail. They show that superstitions about play two or more steps off the equilibrium path can be persistent as the discount factor goes to 1, even though patient players will experiment enough with off-path actions to reject false beliefs about play one step off of the equilibrium path.

³ Michael Woodford argues that what we call the “self-confirming case” is also covered by the original Lucas critique. Robert E. Lucas Jr. in private communication indicates that his primary concern was with the inaccuracy of the prevalent econometric models, and that he was not concerned at the time with the game theoretic distinction we make here. Thus although the historical record is ambiguous, it is consistent with our formulation of the Lucas critique.

detected, and the welfare cost of initially incorrect beliefs when players are rational Bayesians. Our focus is on how these welfare costs vary depending on whether the false beliefs lead to a non-equilibrium action (as in the Lucas critique) or whether the false beliefs prevent the adoption of a policy innovation that would increase welfare (the self-confirming case). When players are impatient, the self-confirming type of errors can persist indefinitely, and have a higher welfare cost, but with patient players the Lucas-type mistake can actually be more costly.

2. The “Lucas critique” model

The Lucas critique can be illustrated with a small variation on the simple peasant-dictator game used to illustrate the basic problem of time consistency. In the peasant-dictator game, there is a peasant who moves first and must decide whether or not to undertake an investment: whether to plant corn or eat the corn seed. If the peasant eats the corn, the dictator gets nothing and the peasant gets one. If the peasant plants the corn, then the dictator must decide how much of the corn growing in the field to appropriate. Suppose that planting the corn results in four units of corn. We may imagine that the dictator decides between the “high tax” action of taking all the corn, yielding a utility of four for herself and a utility of zero for the peasant, and the “low tax” action of taking one unit of corn, resulting in one unit of utility for herself and three units of utility for the peasant.

The traditional time-consistency problem, as analyzed, for example, by Kydland and Prescott [21], is easy to see. Backwards induction implies that after the fact, it is optimal for the dictator to take all the corn. Anticipating this, the peasant prefers to eat the corn, resulting in one unit of utility for herself and none for the dictator. By way of contrast, *ex ante* if the dictator could commit to a “low tax” action, the peasant would grow the corn, and the dictator would get one, the peasant three, both being better off than in the sub-game perfect equilibrium.

To put this into the context of Lucas’ critique, suppose that the dictator can make a commitment to a policy prior to the move by the peasant, and that three policies are available: Always set the low tax, always set the high tax, and set the low tax unless there is a “war” in which case set the high tax. Suppose that the probability of a “war” is 50–50. It is objectively ascertainable whether or not a war takes place, so that it is possible to determine if the dictator follows the commitment, and if a war does take place, it does so after the decision to grow is made. (In practice dictators have committed themselves to policies such as “high tax only in case of war” through institutional arrangements, for example, by requiring a vote of parliament to raise taxes.)

In this game, commitment to a low tax means the peasant will grow the corn, that the dictator will get one and the peasant three. Commitment to a high tax means that the dictator gets nothing and the peasant one. Commitment to a high tax only in case of “war” means that by growing the corn the peasant gets an expected utility of one and a half, while by not growing the corn he gets a utility of only one. In other words, it is still optimal for the peasant to grow the corn, and the dictator now gets two and a half. In other words the optimal commitment is to set the high tax only in case of “war.”

What does this have to do with Lucas? Suppose that this game is played for a number of generations and an enterprising econometrician comes along. The econometrician regresses government income on the tax rate. In the “high taxes only in case of war” policy, there is variation in the tax rate – sometimes taxes are low and sometimes they are high, so it is possible to estimate that low taxes result in a government income of one, and high taxes a government income of four. The econometrician recommends to the dictator that to maximize government revenue the best policy would be to charge a high tax all the time. Once the peasant learns of this change in policy,

then (pace Lucas) he will stop growing corn and start eating it: government revenue will fall to zero. This, in a nutshell, is the Lucas Critique: a structural relation (between taxes and revenues) estimated under one policy regime (high tax only in case of war) leads to a recommendation of a regime change (high tax always) that in turn results in a change in the structural relation.

There are two points we want to emphasize. First, the problem pointed out by Lucas is mitigated by the fact that policy makers will eventually discover that the new policy is a mistake. In our example, after the new policy is implemented government revenues will fall to zero, so we expect that it will ultimately be rescinded. That is not to say that the social cost of failed policy experiments must be small, or that the mistakes will always be discovered quickly. We will examine more carefully the rate of learning and possible welfare losses from Lucas critique types of policy errors in Section 5. On the other hand, the problem pointed out by Lucas is a manifestation of a deeper issue: At an equilibrium, we see only the equilibrium path, and our information about what happens off the equilibrium path is either conjecture (such as the econometrician who conjectures the structural relationship will not change) or based on limited evidence from previous deviations from the equilibrium path.

One conclusion is that there is no substitute for genuine understanding of causality. That is, deep theoretical models supported by strong evidence that the parameters are true constants – aspects of preferences, for example – tested across a broad range of places and times, may give us confidence in understanding the consequences of policies for which there is no direct experience. A second key point is that we may find ourselves in a situation opposite of the one envisaged by Lucas: A superstition about the consequences of off-path play, rather than leading us to test a bad policy, may instead convince us not to implement a good policy. We examine this next.

3. Self-confirming equilibrium

3.1. An example

We consider a simple game involving Foreign Direct Investment (FDI). We imagine that there are two countries, East and West, and that they must decide between two economic policies: one of awarding monopolies to foreign investors, and one of forcing foreign competitors to compete. As a more concrete example, you may wish to think of the monopolies in question as either patent or copyright protection; in fact patent monopolies originated in the late Middle Ages as a way to induce skilled artisans to relocate (see for example Landes [22]).

Following the choice of policy by the two countries, there is a single multinational investor who must determine how to allocate two units of FDI in each of the two countries. The options are: (a) Invest one unit each in East and West; (b) invest everything in East; (c) invest everything in West, and (d) do not invest at all. We normalize payoffs so that if there is no investment in a country, this generates zero for both the country and the multi-national. If there is a unit of investment in a country, under monopoly this results in a return of one for the country and two for the investor, while under competition this results in a return of three for the country and one for the investor. Note the key feature of monopoly in this example is that it is beneficial to the monopolist, but the social costs exceed that benefit.

To generate some variation in policy, let us also suppose that there is a probability of 10% that one and only one of the governments is “socialist” and refuses to grant monopolies. In other words the probability that neither government is socialist is 90%, the probability that both governments are socialist is 0%, an assumption discussed further below, and the probability that a particular government is socialist and its counterpart is not is 5%.

One particular sub-game perfect equilibrium of this game is for all non-socialist governments to grant a monopoly, and for the investor to split investment between countries granting a monopoly, or, if one country is socialist, to invest entirely in the country that is not socialist.

Now we suppose that economic policy can be set for both countries by an international agency, called the World Intellectual Property Organization (WIPO). We make the heroic assumption that the WIPO cares only about economic welfare; its utility is the sum of the utility to the government and to the investor. Suppose over some period of time we are in the equilibrium described above in which the WIPO policy is that both countries grant monopolies – except for the occasional socialist government that does not feel itself bound by WIPO rules. At the end of this period, WIPO econometricians estimate the relationship between welfare and monopoly using data on each individual country. They conclude that not offering a monopoly results in a welfare of zero. Offering a monopoly results in a welfare of three 90% of the time and six 10% of the time, so a welfare of 3.3. So the policy of offering a monopoly appears to be a good one. The problem is that if neither country offered a monopoly the multi-national would continue to invest and the welfare per country would in fact be four.

Notice that we have assumed that both countries are not simultaneously socialist. If they were, then occasionally data would be generated showing that when neither country offered a monopoly welfare is four. However: if the probability of a socialist government is only 10% and each country has an independent chance of being socialist, then both countries are socialist only 1% of the time, so the rate at which data is generated is quite low. This once again raises the issue of how quickly players learn. We examine this issue in Section 5.

3.2. Overview of learning and self-confirming equilibrium

In the WIPO game, we have described a “self-confirming” equilibrium, in which each agent’s strategy is a myopic best response to his beliefs about the play of the other player(s) and are consistent with what is observed when the agent plays the game. That is, if the WIPO believes that failing to grant monopoly will result in zero welfare, this is consistent with what is observed on the equilibrium path in which monopoly is granted except when there is a socialist government, and given those beliefs, it is optimal to grant the monopoly. In other words, beliefs are “self-confirming” because they do not induce actions that generate observations that disconfirm the beliefs. However, the outcome in which monopoly is always granted is also a Nash equilibrium: firms can “threaten” to never invest unless the WIPO insists all non-socialist countries enforce patents.

In this section of the paper, we informally present the assumptions on the learning process that underlie the concept of self-confirming equilibrium. In particular, we want to highlight the assumptions under which the self-confirming equilibria of a game correspond to the long-run outcome (steady states or asymptotic steady states) of a learning model. We will then discuss how modifying the assumptions can lead to a smaller set of possible long-run outcomes and thus provide foundations for some “equilibrium refinements.”

A background assumption of all of the models we consider is that players learn the strategies of their opponents from repeated observations. To fix ideas, we will focus on models where agents keep explicit track of their beliefs about opponents’ play, and use these beliefs to guide their behavior, so that the issues become how the player update their beliefs and how the be-

liefs guide their actions.⁴ This makes it easier to discuss what we see as the main substantive assumptions and issues.

First, if players learn from the data, it should be the case that beliefs are “asymptotically empirical”: Once players have a great many observations of play at an information set, their beliefs should resemble the empirical distribution. This condition is implicit in the idea that players eventually learn the path of play, and is satisfied by Bayesian agents who think they are seeing draws from an exchangeable distribution.

Second, self-confirming equilibrium reflects an element of strategic myopia. That is, players do not attempt to influence the future play of their opponents through punishments, rewards, or manipulation of their learning procedure. For example, patent-granting Nash equilibrium of the WIPO game might be sensible in a repeated-game model. It might be that forward looking multi-national firms might feel that the reputational benefit of refusing to invest unless granted a monopoly is worth the short term cost. The literature on learning has typically ignored these forward looking incentives, often making assumptions about random matching that eliminate them.⁵ In our discussion we will maintain the assumption of myopic incentives. For small individual players, or short-lived governments, this may be a reasonable assumption.

Third, for learning theory to be relevant, players should not have priors that are too strongly held. In the extreme case, players might have point beliefs about their opponents’ strategies, in which case no information would change their mind – and they would stubbornly believe anything regardless of the evidence. More interesting, from the perspective of learning theory, are prior beliefs that might be better motivated and strongly held, but not completely so, such as the belief that opponents behave rationally, or the belief that small changes in price will result in small changes in demand. The set of self-confirming equilibria includes outcomes that cannot arise with priors of this type; restrictions on prior beliefs can lead to refinements of self-confirming equilibrium, most notably in “rationalizable self-confirming equilibrium (RSCE).” This requires players’ beliefs about the opponent’s strategies to be consistent with the idea that the payoff functions are “almost common knowledge” (see Monderer and Samet [24]). For example, the “patent equilibrium” in the WIPO game can be ruled out if we assume that WIPO knows (or is highly confident that it knows) the payoff functions of the firms, because it can then deduce how the firms will respond to a change in policy. The idea of an RSCE is close in spirit to Lucas’s argument that we should look for deep structural parameters, in this case the preferences and information of the players in the game.⁶

Fourth, in addition to the strategic myopia assumed above, there should be asymptotic myopia, in the sense that eventually, once they have a great deal of evidence, players play a myopic best response to their beliefs, no longer experimenting to gather further information. If, for example, there are periodic regime shifts, then it would be desirable to experiment, even asymptotically. By way of contrast, asymptotic myopia is satisfied by rational Bayesian agents who believe the

⁴ Models of reinforcement learning, where players track only the rewards of each action, can have very similar properties even though the agents’ beliefs are not explicit.

⁵ For formal modeling of learning in myopic play, see Fudenberg and Levine [13], Fudenberg and Kreps [10], and Noldeke and Samuelson [25]. There are also limited results on rational forward looking play, see for example Kalai and Lehrer [19]. Jehiel [18] consider Bayesian but boundedly rational players who care only about payoffs for the next k periods, and who believe that opponent’s play only depends on outcomes in the past m periods.

⁶ In the context of inflation, Michael Woodford suggests that a refinement in which the government “expects the worst” along the lines of ambiguity aversion theory might be appropriate.

world is stationary (and have non-doctrinaire priors), as then the option value of experimentation decreases to zero on almost all sample paths. (See, for example, Fudenberg and Levine [13].)

Finally, for all of the self-confirming equilibria to be possible long-run outcomes, it is necessary that there not be too much experimentation at any point in the process, as otherwise players might learn the true distribution of off-path play. Extreme myopia leads to self-confirming equilibrium, as in this case players do not experiment at all. If players are more patient, then there is an option value to experimenting with off-path play, and as the players become more and more patient we expect them to experiment more and more. If they experiment infinitely often at on-path information sets, they will learn the distribution of play at all information sets that are “relevant” to the determination of Nash equilibrium. Fudenberg and Kreps [9] use a model of exogenously given experimentation to show that this is enough to prevent convergence to non-Nash outcomes; Fudenberg and Levine [13] provide an analogous result about rational Bayesian learning with patient players, where the decision whether to experiment is endogenous.

However, these results do not provide a foundation for the use of the Nash equilibrium concept when players are patient, as they do not pin down the extent of experimentation at off-path information sets. It turns out that the possible long-run outcomes when players are patient is a refinement of Nash equilibrium called “subgame-confirmed Nash equilibrium (SCNE).” To understand the intuition for this equilibrium concept, note that in a world in which players do a lot of experimenting at every information set that is on the equilibrium path, they will come to have correct beliefs about play one step off of the equilibrium path. This is why non-Nash outcome cannot persist. However, experimentation by on-path players forces their opponents to actually respond to some “off path” play, so that there is a cost to choosing suboptimal responses, while Nash equilibrium allows agents one step off the path of play to choose their actions arbitrarily. Specifically, the “patent equilibrium” of the WIPO game is a Nash equilibrium – but involves the firm making investments that are not in fact optimal if the subgame has to actually be played. For this reason, if players are patient enough to rule out non-Nash self-confirming equilibria, then some of the Nash equilibria can be ruled out as well, which raises the question of just which outcomes can occur when players are patient. Fudenberg and Levine [15] provide a sufficient condition: The outcome of any SCNE can be generated by patient rational learning.

To sum up, then: If play converges, then learning alone should lead at least to a self-confirming equilibrium.⁷ Learning plus reasoning about opponents’ incentives should lead to the further refinement of RSCE. Learning plus substantial experimentation with off-path play should lead instead to the further refinement of SCNE. Conveniently, in two-stage games of perfect information, both RSCE and SCNE are equivalent to backwards induction. In general, however, neither RSCE, nor SCNE allows arbitrarily long chains of backwards induction.

In the case of RSCE, long chains of backwards induction break down because RSCE recognizes that a small prior probability that an opponent’s payoff function differs from its expected value can become a large posterior probability after observing an unexpected outcome.⁸ In the

⁷ Throughout this discussion we have avoided the issue of whether the learning system will in fact converge, which is an important open research area. If the system does not converge to a single profile but for example cycles, then the assumption that the players view the system as stationary is open to question. See Fudenberg and Kreps [10]. In the macroeconomic literature the issue of stability of learning procedures has been given more attention – see in particular Cho, Williams and Sargent [5] who note the equivalence of Nash and self-confirming equilibrium paths in their model, but show how the instability of the self-confirming equilibrium leads the government to repeatedly “discover” too strong a version of the natural rate hypothesis.

⁸ See Fudenberg, Kreps, and Levine [11], Dekel and Fudenberg [6], and Börgers [4].

case of SCNE long chains of backwards induction break down because there is little value to experimenting at nodes that are off the equilibrium path and consequently reached only infrequently. In short, learning theory points us either to general self-confirming equilibria (if players are not terribly patient, and so do not experiment a great deal), or towards refinements of self-confirming equilibria if players use knowledge of opponents' payoffs, or if they experiment a great deal.

3.3. Self-confirming equilibrium

To provide a formal definition of self-confirming equilibrium we must first specify some notation. For notational simplicity and conceptual clarity, we will restrict attention to games of perfect information. There are $I + 1$ players in the game, where player $i = I + 1$ is nature. We consider an extensive-form game, with a finite game tree consisting of nodes X . The terminal nodes are $z \in Z \subset X$. The nodes at which player i has the move are denoted by X_i . The feasible actions at a node x are denoted $A(x)$. A pure strategy for player i , s_i , is an action at each node $x \in X_i$, $s_i(x) \in A(x)$; S_i is the set of all such strategies. We denote by s, s_{-i} profiles of all strategy profiles, and those of all players except player i respectively. Each strategy profile determines a terminal node $\zeta(s) \in Z$. In interpreting the model, we suppose that all players know the structure of the extensive form, so that each player knows the space S of strategy profiles and can compute the function ζ . Each player i receives a payoff in the stage game that depends on the terminal node. Player i 's payoff function is denoted $u_i : Z \rightarrow \mathfrak{R}$.

Let $\Delta(\cdot)$ denote the space of probability distributions over a set. Then a mixed strategy profile is $\sigma \in \times_{i=1}^{I+1} \Delta(S_i)$. In addition to mixed strategies, we define behavior strategies. A behavior strategy for player i , π_i , assigns nodes $x \in X_i$ a probability distribution over feasible actions, $\pi_i(x) \in \Delta(A(x))$; Π_i is the set of all such strategies. For a fixed s_i , the marginal probability of reaching a node $x \in X_i$ depends on the behavior strategies of the other players and is denoted $p_i(x|\pi_{-i})$. Let $Z(s_i)$ be the subset of terminal nodes that are reachable when s_i is played, that is $z \in Z(s_i)$ if and only if for some $s_{-i} \in S_{-i}$, $z = \zeta(s)$. Similarly, define $X(s_i)$ to be all nodes that are reachable under s_i , and extend this definition to mixed strategies $X(\sigma_i)$ and behavior strategies $X(\pi_i)$ by requiring that the nodes or information sets be reachable with positive probability. We will also need to refer to the nodes that are reached with positive probability under σ , denoted $\bar{X}(\sigma)$.

We now model the idea that each player has a belief about his opponents' play (including the play of Nature). Because many different mixed strategies can be observationally equivalent, it is easiest to model beliefs as a probability measure over Π_{-i} , the set of other players' behavior strategies. Let μ_i denote the belief of player i . We may then define utility with respect to those beliefs by

$$u_i(s_i, \mu_i) \equiv \sum_{z \in Z(s_i)} u_i(z) \int p_i(z|\pi_{-i}) \mu_i(d\pi_{-i}).$$

For a given mixed strategy profile σ let π^σ be the observationally equivalent behavior strategy. We say that player i 's belief μ_i is *correct* at an opponent j 's node x if $\mu_i(\{\pi_{-i} | \pi_j(x) = \pi_j^\sigma(x)\}) = 1$. This will be used to capture the idea that asymptotically players will have correct beliefs about play at nodes that are reached sufficiently often. At the least, this condition will apply to nodes that are "on the equilibrium path," and it may apply to some other nodes as well, depending on how patient the players are and thus how much experimentation there is.

Our first notion of equilibrium, *self-confirming equilibrium*, corresponds to the case of myopic players who do little or no experimentation. Consequently it imposes only the restriction that players learn what happens on the equilibrium path.

Definition 3.1. $\hat{\sigma}$ is a *heterogeneous self-confirming equilibrium* if for each player i and for each s_i with $\hat{\sigma}_i(s_i) > 0$ there are beliefs $\mu_i(s_i)$ such that

- (a) s_i is a best response to $\mu_i(s_i)$, and
- (b) $\mu_i(s_i)$ is correct at every $x \in \bar{X}(s_i, \hat{\sigma}_{-i})$.

It is important to note that this definition allows player i to rationalize each s_i in the support of $\hat{\sigma}_i$ with a different belief. This is relevant to models of anonymous random matching in large populations. In those models, and in other settings with representative agents, there will be many different individuals in the role of each player, and they may hold different beliefs. One such example is the representative voter in the Alesina and Angeletos [1] model that we discuss in Section 4.2. There is no reason to suppose that all voters have the same homogeneous beliefs just because they have the same preferences.

In other settings such as that of the WIPO/FDI games there is a single unitary agent for each player role, and we would use the following stronger definition.

Definition 3.2. $\hat{\sigma}$ is a *unitary self-confirming equilibrium* if for each player i there are beliefs μ_i and for each s_i with $\hat{\sigma}_i(s_i) > 0$ such that

- (a) s_i is a best response to μ_i , and
- (b) μ_i is correct at every $x \in \bar{X}(\hat{\sigma})$.⁹

By way of contrast, *Nash equilibrium* strengthens (b) to hold for all $x \in X$. In terms of observed outcomes, there are four ways that self-confirming equilibrium can differ from Nash equilibrium. First, two players might have different beliefs about the play of a third player, as in the example of Fudenberg and Kreps [8]. This example relied on player 3's off-path information set being reachable both by a deviation of player 1 and by a deviation by player 2; it cannot arise in the games of perfect information we consider here, nor in games with only two players. The second sort of non-Nash outcome arises when a player's beliefs about the off-path play of the opponents corresponds to a correlated strategy; this too cannot occur in games with only two players. Next, different agents in the role of the same player can have different beliefs, as we indicated above and will illustrate below. Finally, if the distribution of Nature's move is not known *a priori*, but learned in the same way as the distribution of opponents' play, then even in a two-player game SCE can differ from Nash as the players may have different beliefs about the off-path play of Nature.

The role of heterogeneous beliefs and the contrast to unitary SCE is easy to see in a two-player "Stackelberg game" of complete information: player 1 picks an action $a_1 \in \{U, D\}$, player 2 observes a_1 and plays $a_2 \in \{N, L, R\}$ and then the game ends. Suppose that when player 1 plays D , player 2 plays N , and all agents in the role of player 1 know this. However, some player

⁹ Battigalli [2] defined an equivalent concept of "conjectural equilibrium" for two-player games with the distribution of Nature's moves known, and showed that in these games, unitary SCE is outcome-equivalent to Nash equilibrium. He conjectured that this result extends to games with more players.

1's incorrectly believe that player 2 will respond to U with L , and that payoffs are such that this belief makes D the best response for player 1, while other player 1's correctly believe that player 2 responds to U with R , and that this belief makes U optimal. Then we can have a heterogeneous SCE in which some player 1's play U and some play D , even though this is not the outcome of any Nash equilibrium.

Many experiments from the laboratory show the significance of self-confirming equilibrium. Two famous experiments, thought largely to contradict standard equilibrium theory, turn out largely to confirm the importance of self-confirming equilibrium. In the celebrated ultimatum bargaining experiment in which one party makes a binding take-it-or-leave-it offer for dividing a sum of money, backwards induction predicts that the first mover should get essentially all of the pie. This fails rather badly in the laboratory, where offers are closer to 50–50.

A useful way to examine the data is by examining the *losses* that players suffer relative to the most that they could have earned given the objective play of the other players. The theory of self-confirming equilibrium informs us that we should also distinguish between knowing losses and unknowing losses – the first correspond to losses that a player would know given his own play and the objective play of the other players, and the latter to losses that he could know about only by experimentation with alternative actions. So, for example, the player moving first in ultimatum bargaining has no knowing losses – he may well believe that any higher demand will be rejected, and so that his current demand is the best one. In the ultimatum bargaining experiments, Fudenberg and Levine [14] show that the failure of backwards induction is driven by knowing losses – the fact that second movers are willing to reject ungenerous offers. However, the unknowing losses of the first mover are in fact 3–5 times greater than the knowing losses of the second mover, indicating that self-confirming effects – the lack of knowledge of off-the-equilibrium-path play of opposing players – is quantitatively more significant than, for example, the preferences for altruism and spite that are implicit in second mover play.

In a similar vein, consider the surprising results of the centipede game, in which at each stage players have a chance to opt out or continue. If they continue the social value doubles, but if at the next stage their opponent opts out, the current mover's payoff decreases. Here the backwards induction solution is to opt out in the first stage, but this happens in the laboratory less than 3% of the time. The root cause is again interpersonal preferences, as some players generously and intentionally give away money at the end of the game, which reverses the incentives for staying in earlier in the game. Yet here again, the unknowing losses of players who foolishly drop out too soon are seven times as large as that of the altruists playing at the end of the game.

To put it a different way: laboratory experimentation has established that there are systematic deviations from the assumption that individuals act to maximize expected utility functions for their own monetary income only, and a great deal of research has been focused on studying these alternative preferences. The Lucas effect – the lack of knowledge of off the equilibrium path play – has received less attention, but it is quantitatively far more significant, on the order of five times as important.

3.4. Rationalizable self-confirming equilibrium

As we explained above, RSCE is relevant when players have strong prior information about their opponents' payoffs and use this information in forming beliefs about off-path play. This knowledge leads to backwards induction in two-stage games of perfect information, such as the examples in Section 4, but it does not imply backwards induction in longer games.

To define this concept formally, we begin with the idea of “versions” of player i ; this is $v_i = (\pi_i, \mu_i)$ where π_i is a behavior strategy and μ_i is a probability measure over the behavior strategies of other players. A *belief model* is a collection of sets of versions (V_1, \dots, V_I) . We say that a belief model is *belief closed* if for every $(\pi_i, \mu_i) \in V_i$, the beliefs μ_i are consistent with being a probability measure over the types of other players, that is if

$$\mu_i \left(\left\{ \sum \alpha_j \pi_j \mid (\pi_j, \mu_j) \in V_j \right\} \right) = 1.$$

Definition 3.3. $\hat{\sigma}$ is a *rationalizable self-confirming equilibrium (RSCE)* if there is a belief-closed model V such that

- (a) $(\pi_i, \mu_i) \in V_i$ maximizes $u_i(\cdot, \mu_i, x)$ at all $x \in X(\pi_i)$,
- (b) if $(\pi_i, \mu_i) \in V_i$ then they induce the same distribution over terminal nodes as $\hat{\sigma}$.

3.5. Subgame confirmed Nash equilibrium

Definition 3.4. $\hat{\sigma}$ is a *subgame confirmed Nash equilibrium (SCNE)* if it is a Nash equilibrium, and for each node x that is one step off the equilibrium path, the strategies $(\hat{\sigma}|_x)$ are a self-confirming equilibrium of the subgame beginning at x .

To repeat the earlier discussion, the idea of this definition is that patient on-path players will experiment enough to learn the play one step off the path. This is all that is needed for play to yield a Nash outcome. Because of this experimentation, players one step off the path of play will learn the “continuation path” if there are no further deviations. The “ k -step perfection” of Kalai and Neme [20] is a stronger condition that requires a Nash equilibrium at every node k or fewer steps off the path. However, there is currently not a learning-theoretic foundation for k -step perfection.

In the simple setting of two-stage games of perfect information, as we mentioned, both RSCE and SCNE are the same as backwards induction. In the simplified macroeconomic models discussed below, these concepts form the learning theoretic background for the shorthand that backwards induction corresponds to greater information about opponents payoffs (RSCE) or enough patience to be willing to invest in actively learning about off-path play (SCNE).

4. Applications in macroeconomics

To illustrate the significance and limitations of self-confirming equilibrium in macroeconomics, we examine three applications that have appeared in the literature.

4.1. Hahn’s conjectural equilibrium

In many respects, the notion of self-confirming equilibrium is anticipated by Hahn’s [17] notion of a “conjectural equilibrium.” In that paper he argues that firms accurately perceive their profit level, but incorrectly anticipate a fall in profits if they change the price they are currently charging. This leads to a model in which the price level can be arbitrary, as in the fixed-price models of Benassy [3] and others. This application is most interesting because it illustrates the limitations of self-confirming equilibrium. In a setting where firms engage in a variety of small variations in price, as do their rivals, where the real price is constantly varying due to small

changes in inflation, and payoffs are continuous in actions we would anticipate a gradual process of local adjustment, leading at least to a local maximum. This applies also to games played in the laboratory. In ultimatum bargaining, for example, if the grid of allowed offers is fine (as it generally is) and if subjects play many times, it would be natural for them to experiment with asking for a nickel more to see if they can get it. However, subjects typically only get to play ten or so times; and because it takes several trials to estimate the fraction of their opponents that reject any particular offer there is not incentive for them to do these experiments. By way of contrast, business firms get a great deal of feedback about the consequences of their pricing, so it is doubtful that prices are sticky on account of a belief by firms that whatever price they happen to be setting is optimal.

In our WIPO example, there is also constant generation of small amounts of information. However, the self-confirming equilibrium is locally stable in the sense that the small amounts of information – the elimination of monopoly in only one country – is misleading about the consequences of eliminating monopoly in both countries. Put differently – when we consider dynamic stability, we must recognize that while a self-confirming equilibrium that is not Nash is not robust to the discovery of the entire truth, it may be robust to smaller discoveries.

4.2. Alesina–Angeletos

We give a simplified version of Alesina and Angeletos [1] model of redistributive tax policy. The key idea is that voters care both about fairness and about efficiency. The government, reflecting the wishes of the voters, chooses tax policy that can be more or less redistributive. Economic outcomes are influenced both by investment and luck. Even if the social optimum is a less redistributive policy that encourages investment, there may be a self-confirming equilibrium in which there is high redistribution. The idea is that the high level of redistribution discourages investment. What voters see then is that pre-tax income is mostly due to luck, and that the redistributive policy is more fair. They also believe that if the policy were changed to be less redistributive, then investment would not increase. This becomes self-confirming, because they choose the more redistributive policy.

Specifically, we consider the following simplified Alesina–Angeletos model. First the government moves, choosing either high redistribution (H) or low redistribution (L). Then a representative individual chooses either to invest (I) or not to invest (0). Finally, nature moves, assigning either good luck (G) or bad luck (B) with equal probability.¹⁰ We normalize the base payoff of the investor to 0, with a premium of 2 for good luck, and a benefit of investment of 2 and a cost of 1. These base payoffs are also the actual payoffs under the low redistribution policy. Under the high redistribution policy the individual gets one minus the cost of investment (if any). The government gets the same utility as the investor plus a “fairness bonus” of $\frac{1}{2}$ for the high redistribution policy.

First we analyze backwards induction, which corresponds to the RSCE in the case where players know the government knows the payoffs of the investor. In the subgame in which the government chooses low redistribution it is optimal to invest, and the utility of the investor and the government is 2. In the subgame in which the government chooses high redistribution it is optimal not to invest and the utility of the investor is one and that of the government is $1\frac{1}{2}$. So the optimal policy is low redistribution resulting in the utility of 2.

¹⁰ The actual Alesina–Angeletos model is considerably more sophisticated, with heterogeneity in ability and continuous choices of investment and taxes in an overlapping generations setting.

There is, however, a unitary self-confirming equilibrium in which the government sets high redistribution, believing that if it chooses low redistribution there will still be no investment. (Since this is off the equilibrium path, the belief is untested.) On the equilibrium path there is no investment and the investor gets 1, the government $1\frac{1}{2}$. According to the government's beliefs, if it chooses low redistribution there will still be no investment, and the investor will still get 1, but the "fairness bonus" is lost so the government gets only 1. Consequently this is a self-confirming equilibrium. Notice that this outcome is also the outcome of a Nash equilibrium: since low redistribution is off the equilibrium path, it is in fact a best response for the investor not to invest in response to low redistribution.¹¹ However, while the Nash equilibrium has the same outcome as the self-confirming equilibrium here, this highlights a deficiency of Nash equilibrium: why would the investor respond to low redistribution by "punishing" the government and themselves with low investment? By way of contrast, the self-confirming equilibrium makes sense, provided the government does not discover that its beliefs are wrong.

Here too it is the case that experimentation with a slightly less redistributive policy should yield information that the increase in output more than compensates for the decrease in fairness. However, it is likely that redistributive policy impacts output only with a substantial lag, and there are many confounding factors, meaning that the signal is quite noisy. So it seems plausible that this information about the relationship would emerge only slowly. There may also be institutional constraints that make experiments rarer and more difficult than experiments by firms. This is consistent with the view that European policy has become less redistributive but only gradually so.

4.3. The Sargent–Williams–Zha inflation model

The following simplified version of the Sargent, Williams and Zha [26] model of inflation has a non-Nash self-confirming equilibrium based on incorrect belief about a Philips curve. In a sense this is the self-confirming opposite of the simple "Lucas critique" model we discussed at the start of the paper.

We assume that there is a policy maker who chooses a monetary policy, which we take to be either high or low inflation, and a representative consumer who moves after observing the monetary policy and chooses either high or low unemployment. The policy maker prefers low inflation but is willing to choose 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 choose low unemployment. It follows that backwards induction leads the policy maker to choose low inflation and receive a payoff of 3. There is, however, a unitary 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 1.¹²

¹¹ In fact in any two-player game where players have correct beliefs about Nature, any outcome of a unitary self-confirming equilibrium is the outcome of a Nash equilibrium, as shown in Fudenberg and Levine [12] and Fudenberg and Kreps [10].

¹² As in the Angeletos–Alesina model, this outcome is also the outcome of a Nash equilibrium where the consumer really does choose high unemployment after low inflation, but the same critique applies: why should the consumer do this?

Obviously this model is a highly simplified version of a misperceived Phillip's curve. Sargent, Williams and Zha argue that a misperceived Phillips curve resulting in a 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 about the Phillips curve that allows misperception but also allows the misperception to be corrected as data accumulated. They argue that this learning model can explain many of the details of U.S. monetary policy and inflation during 1970s and 80s. In Sargent, Williams and Zha [27] they use a related model to explain hyperinflationary episodes and monetary reform in South America.

5. The effectiveness of learning

In both the Lucas case and the self-confirming case, the short-run state of the economy may be quite different from the Nash equilibrium that will obtain if players eventually learn that their beliefs are incorrect. Whether this short-run outcome is of practical interest or not depends on how long that short-run is, that is, on how quickly players learn. In this section we investigate more carefully the length of time it takes to reach the long run. To compare the outcomes in various cases, we introduce the concept of "learning effectiveness," which measure how a player's present value payoff compares to the case where she knows the true parameters.

We analyze the special case where there is an existing status quo policy that has been in use for some period of time, and, as a result, has a known expected payoff that we may normalize to 0. We suppose that a single alternative not previously contemplated is proposed for the first time. In other words, we consider a two-armed bandit problem in which one arm has a known value, and the other (the "new" arm) has an unknown value.

In both the self-confirming and Lucas cases learning effectiveness is higher when the player is more patient, but the role of patience in the two cases is very different. In the Lucas case, 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. As we will see, learning effectiveness is actually higher in the self-confirming case than in the Lucas 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.

Consider first the Lucas case. 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 ("low unemployment"). We suppose, moreover, that for the reasons described by Lucas the econometrician is in error. If the new "high inflation" regime is introduced, how long will it take before the policy maker discovers the error? To further focus thinking, let us suppose that the outcome of the new "high inflation" regime is binomial, and that the prior is the conjugate beta distribution

$$\frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha, \beta)}$$

where p is the probability of success, α, β are parameters describing the prior and

$$B(\alpha, \beta) = \int_0^1 q^{\alpha-1} (1-q)^{\beta-1} dq$$

is the beta function. On a success the posterior is

$$\frac{p^\alpha (1-p)^{\beta-1}}{B(\alpha+1, \beta)}$$

while on failure it is

$$\frac{p^{\alpha-1} (1-p)^\beta}{B(\alpha, \beta+1)}.$$

If the payoff to success is $+1$ and the payoff to failure is -1 then the mean and variance of the payoff with respect to the beta distribution are

$$\mu = \frac{\alpha - \beta}{\alpha + \beta},$$

$$\sigma^2 = \frac{4\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)},$$

so it is convenient to parameterize the prior by μ, σ rather than α, β .

In the Lucas case, the prior is favorable for the new arm, so $\mu > 0$, while the actual mean of the new arm is $\mu^* < 0$. We will contrast this with the “self confirming case,” where the alternative arm is *a priori* thought to be inferior, so the prior mean $\mu < 0$, while in fact it is superior, so that $\mu^* > 0$. Because there is a single risky arm, the optimal policy here takes the form of a stopping time: Once the agent uses the safe and uninformative arm, he uses it in all subsequent periods. Hence if we let T be the stopping time after which the safe arm is used the agent’s expected average discounted payoff with respect to the true distribution is simply $(1 - E\delta^T)\mu^*$.

In the Lucas case, $\mu^* < 0$ and the full-information optimum has payoff 0, so the agent’s loss compared to full information is $-(1 - E\delta^T)\mu^*$, which is minimal when $E\delta^T = 1$. In the self-confirming case, where $\mu^* > 0$, the full-information payoff is $\mu^* > 0$, so the agent’s loss compared to full information is $E\delta^T\mu^*$. To facilitate the comparison of these two cases, we define the *learning effectiveness* to be $\lambda = E\delta^T$ in the Lucas case, and $\lambda = 1 - E\delta^T$ in the self confirming case. This is equivalent to defining

$$\lambda = E \sum_{t=1}^{\infty} \delta^{t-1} I_t,$$

where I_t an indicator function which is 1 when the objectively correct arm is used and 0 otherwise. Thus the agent’s loss compared to full information is $(1 - \lambda)|\mu^*|$ in both cases.

Consider first the case of complete myopia, so $\delta = 0$. In the self-confirming case since $\mu < 0$, the player will never try the alternative arm, and will remain stuck on the wrong arm forever. Consequently $\lambda = 0$. In the Lucas case the player will pull the wrong arm once, and again $\lambda = 0$. This is obvious, and not very interesting, but we can use it to bootstrap to the case of small but non-zero discounting. For any given beliefs, if δ is sufficiently small in the self-confirming case where $\mu < 0$ it will still not be optimal to try the alternative arm since the possible benefits (a gain forever) are outweighed by the cost of an anticipated mistake in the first period, so it is still true that $\lambda = 0$.

Now consider the Lucas case, and suppose that the true value of p is near zero so that the true mean of the risky arm is approximately -1 . Moreover, the arm has little variance, so we may imagine as an approximation that the player draws -1 every period in which he uses the risky arm. If the prior is given by α, β and the Gittins index is $-g < 0$ the stopping time T when -1 is drawn every period is the solution to

$$\frac{\alpha - \beta - T}{\alpha + \beta + T} = -g$$

yielding

$$T = \frac{\alpha(1+g) - \beta(1-g)}{1-g}.$$

If δ is small then the Gittins index g is near zero, and we have approximately

$$T = \alpha - \beta = (\alpha + \beta)\mu.$$

So for small δ and small sampling variance on the uncertain arm, we have the approximation in the Lucas case

$$\lambda = \delta^{(\alpha+\beta)\mu}.$$

The worse the arm is thought to be – that is, the smaller is the prior mean μ – and the less strongly held that belief – that is, the smaller is $\alpha + \beta$ – the greater is the effectiveness of learning.

Before doing more precise calculations, we can further firm up our intuition about the length of time it takes to learn in the Lucas and self-confirming cases by considering the situation where the player is very patient, that is δ is near one. In the Lucas case this will increase the length of time it takes to switch back to the old correct arm, since it is no longer optimal to stop when the posterior mean reaches zero, but rather to continue until it is sufficiently negative. However, although the time it takes to learn increases, it is easy to show that λ approaches 1, so that the fraction of discounted time spent on the wrong arm goes to zero. An impatient player cares only about the first period, and in the Lucas case the first period action is a mistake. If the player is patient, the mistakes occur only for a small (relative to the discount factor) fraction of his life, and so are not very costly.

In the self-confirming case the effect of increased patience is different. While an impatient player never learns, a patient player who experiments will learn that the new arm is better, unless he gets bad draws the first few times he tries the arm and so switches back to the wrong arm forever. Once the discount factor is high enough that the player is willing to experiment once, he either learns very quickly or not at all. The more patient he is, the more likely it is that he will experiment enough times to avoid getting misleading samples, so the higher the probability that he learns correctly. Put differently, the effective time to learn is either zero or infinity.

For a relatively myopic player, utility is higher in the Lucas case than in the self-confirming case, since learning takes place even if slowly in the former, while not at all in the latter. A relatively patient player gets about the full-information payoff in either case. In the intermediate cases the ranking between the two cases in learning effectiveness is ambiguous, but the way in which learning takes place is quite different. In the Lucas case learning is gradual and sure, while in the self-confirming case it is all or nothing.

To examine this more closely we report simulation results based on the Gittins index tables in the Appendix of Gittins [16]. Gittins considers a binomial with a value of 1 on success and 0 on failure. In these units, our certain arm yields a sure payoff of 0.5, so the optimal rule is to stop as

soon as the index falls below 0.5. Let α_t, β_t be the posterior parameters after t observations; that is, α_t is the sum of the prior value of α and the number of successes (payoffs of +1), and β_t is sum the prior value β and the number of failures (payoffs of 0 for Gittins, payoff of -1 for us). We consider two different discount factors, $\delta = 0.5$ and $\delta = 0.9$, where the latter may be thought of roughly as corresponding to annual data. Conveniently, for $\delta = 0.5$ and $\alpha_t, \beta_t \leq 20$ in the Gittins table stopping when the index falls below 0.5 corresponds to stopping when $\alpha_t - \beta_t \leq -1$. For $\delta = 0.9$ and $\alpha_t, \beta_t \leq 40$ the stopping rule is $\alpha_t - \beta_t \leq -2$. We extrapolated these rules for larger values of α_t, β_t , and ended the simulation when $t = 10$ for the case $\delta = 0.5$, and when $t = 66$ when $\delta = 0.9$ leaving a maximum possible error of less than 0.001 in the calculation of λ .

In the table below, the shaded rows correspond to the Lucas case in which the prior mean on the risky arm is positive and the true mean is negative. The unshaded arms report the results for the symmetric self-confirming cases obtained by switching the signs of the true mean μ^* and the prior mean μ . We report a range of priors for each discount factor; in each case a Monte Carlo with 10,000 trials was run.

| Actual Prior | | | | | Effectiveness | |
|--------------|-------|----------|---------|----------|---------------|--|
| μ^* | μ | α | β | δ | λ | |
| -0.2 | 0.2 | 3 | 2 | 0.5 | 0.102 | |
| -0.2 | 0.2 | 3 | 2 | 0.9 | 0.392 | |
| -0.5 | 0.5 | 3 | 1 | 0.5 | 0.061 | |
| -0.5 | 0.5 | 3 | 1 | 0.9 | 0.475 | |
| -0.9 | 0.2 | 3 | 2 | 0.9 | 0.706 | |
| 0.2 | -0.2 | 2 | 3 | 0.5 | 0.000 | |
| 0.2 | -0.2 | 2 | 3 | 0.9 | 0.506 | |
| 0.5 | -0.5 | 1 | 3 | 0.5 | 0.000 | |
| 0.5 | -0.5 | 1 | 3 | 0.9 | 0.000 | |
| 0.9 | -0.2 | 2 | 3 | 0.9 | 0.965 | |

The results confirm that the learning efficiency in the self-confirming case tends to be more extreme than in the Lucas case. In the self-confirming case the agent either does not experiment at all, so $\lambda = 0$, or (when δ is sufficiently large) the agent experiments and is very likely to learn the true optimum, so λ is close to 1. (Recall that in the self-confirming case the risky experiment is the full-information optimum, so the only way that λ can be positive but less than 1 is if the agent starts out experimenting but eventually reverts to the safe arm.) In the Lucas case, using the risky arm is a mistake and the agent eventually learns this, regardless of parameters. Nevertheless in the cases we consider the initial losses have a non-trivial cost; this is why the learning effectiveness with patient players is lower here than in the self-confirming case.

Acknowledgments

We would like to thank Robert E. Lucas Jr., Lee Ohanian, Michael Woodford, an anonymous referee and associate editor for useful comments, Jie Zheng for careful proofreading, and NSF Grants SES-03-14713 and SES-06-646816 for financial support.

References

- [1] A. Alesina, G.-M. Angeletos, Fairness and redistribution, *Amer. Econ. Rev.* 95 (2005) 960–980.
- [2] P. Battigalli, *Comportamento razionale ed equilibrio nei giochi e nelle situazioni sociali*, undergraduate dissertation, Bocconi University, 1987.
- [3] J.-P. Benassy, Neo-Keynesian disequilibrium theory in a monetary economy, *Rev. Econ. Stud.* 42 (1975) 503–523.
- [4] T. Börgers, Weak dominance and approximate common knowledge, *J. Econ. Theory* 64 (1994) 265–276.
- [5] I.K. Cho, N. Williams, T.J. Sargent, Escaping Nash inflation, *Rev. Econ. Stud.* 69 (2002) 1–40.
- [6] E. Dekel, D. Fudenberg, Rational play with payoff uncertainty, *J. Econ. Theory* 52 (1990) 243–267.
- [7] E. Dekel, D. Fudenberg, D.K. Levine, Learning to play Bayesian games, *Games Econ. Behav.* 46 (2004) 282–303.
- [8] D. Fudenberg, D. Kreps, A theory of learning, experimentation, and equilibrium in games, Mimeo, 1988.
- [9] D. Fudenberg, D. Kreps, Learning in extensive games, II: Experimentation and Nash equilibrium, Mimeo, 1994.
- [10] D. Fudenberg, D. Kreps, Learning in extensive games, I: Self-confirming equilibria, *Games Econ. Behav.* 8 (1995) 20–55.
- [11] D. Fudenberg, D. Kreps, D.K. Levine, On the robustness of equilibrium refinements, *J. Econ. Theory* 44 (1988) 354–380.
- [12] D. Fudenberg, D.K. Levine, Self-confirming equilibrium, *Econometrica* 61 (1993) 523–546.
- [13] D. Fudenberg, D.K. Levine, Steady state learning and Nash equilibrium, *Econometrica* 61 (1993) 547–574.
- [14] D. Fudenberg, D.K. Levine, Measuring subject's losses in experimental games, *Quart. J. Econ.* 112 (1997) 508–536.
- [15] D. Fudenberg, D.K. Levine, Superstition and rational learning, *Amer. Econ. Rev.* 96 (2006) 630–651.
- [16] J.C. Gittins, *Multi-Armed Bandit Allocation Indices*, John Wiley & Sons, New York, 1988.
- [17] F. Hahn, Exercises in conjectural equilibria, *Scand. J. Econ.* 70 (1977) 210–226.
- [18] P. Jehiel, Learning to play limited forecast equilibria, *Games Econ. Behav.* 22 (1999) 274–298.
- [19] E. Kalai, E. Lehrer, Rational learning leads to Nash equilibrium, *Econometrica* 61 (1993) 1019–1045.
- [20] E. Kalai, A. Neme, The strength of a little perfection, *Int. J. Game Theory* 20 (4) (1992) 335–355.
- [21] F. Kydland, E. Prescott, Rules rather than discretion: The inconsistency of optimal plans, *J. Polit. Economy* 85 (1977) 473–490.
- [22] D.S. Landes, *The Unbound Prometheus: Technological Change and Industrial Development in Western Europe from 1750 to Present*, Cambridge University Press, 1969.
- [23] R. Lucas, Econometric policy evaluation: A critique, *Carnegie-Rochester Conf. Ser. Public Pol.* 1 (1976) 19.
- [24] D. Monderer, D. Samet, Approximating common knowledge with common beliefs, *Games Econ. Behav.* 1 (1989) 170–190.
- [25] G. Noldeke, L. Samuelson, An evolutionary analysis of backward and forward induction, *Games Econ. Behav.* 5 (1993) 425–454.
- [26] T. Sargent, N. Williams, T. Zhao, Shocks and government beliefs: The rise and fall of American inflation, *Amer. Econ. Rev.* 96 (2006) 1193–1224.
- [27] T. Sargent, N. Williams, T. Zhao, *The Conquest of South American Inflation*, NYU, 2006.