



# HIER

## Harvard Institute of Economic Research

Discussion Paper Number 2114

Superstition and Rational Learning  
by

Drew Fudenberg  
and  
David K. Levine

March 2006

HARVARD UNIVERSITY  
Cambridge, Massachusetts

This paper can be downloaded without charge from:  
<http://post.economics.harvard.edu/hier/2006papers/2006list.html>

The Social Science Research Network Electronic Paper Collection:  
<http://ssrn.com/abstract=888774>

# Superstition and Rational Learning<sup>1</sup>

This version: 9/8/05 First version: 6/5/03

**Abstract:** We argue that some but not all superstitions can persist when learning is rational and players are patient, and illustrate our argument with an example inspired by the code of Hammurabi. The code specified an “appeal by surviving in the river” as a way of deciding whether an accusation was true, so it seems to have relied on the superstition that the guilty are more likely to drown than the innocent. If people can be easily persuaded to hold this superstitious belief, why not the superstitious belief that the guilty will be struck dead by lightning? We argue that the former can persist but the latter cannot by giving a partial characterization of the outcomes that arise as the limit of steady states with rational learning as players become more patient. These “subgame-confirmed Nash equilibria” have self-confirming beliefs at information sets reachable by a single deviation. According to this theory a mechanism that uses superstitions two or more steps off the equilibrium path, such as “appeal by surviving in the river,” is more likely to persist than a superstition where the false beliefs are only one step off of the equilibrium path.

---

<sup>1</sup> We thank Douglas Bernheim, an anonymous referee, and seminar participants at MIT, Yale, University of Texas Austin, Hong Kong University, University of Tokyo, the Federal Reserve Bank of Minneapolis and Stanford for helpful comments.

## **1. Introduction**

By a superstition we mean a belief which is objectively false. When can a superstition persist in the face of rational learning? Our basic insight is that superstitions concerning events that are off of the equilibrium path are more likely to persist than those that are not. Intuitively, if play converges to a steady state, we expect the players to learn (at least) the path of play. However, this does not rule out false beliefs about play off of the equilibrium path, and it does not even imply that the steady states must be Nash equilibria, because Nash equilibrium corresponds to a situation where players know not only the equilibrium path but also the consequences of unilateral deviations. Rational but very impatient learners will only play “greedy” strategies that maximize current payoff, so steady states with impatient rational learners can exhibit a wide range of false beliefs. Rational and patient learners “experiment” with other strategies. This experimentation reduces the set of durable superstitions. The question then is just how much experimentation rational learners will do, and how much this will restrict the possible off-path beliefs.

To carry out the analysis, we adopt the overlapping-generations model of our [1993b] paper, and consider the limit of the steady states of this learning model, as first the length of life becomes infinite, and then the discount factor approaches one; we call these the “patiently stable states.” Past work has shown that non-Nash equilibria can be steady states if learners are impatient, but that any patiently stable state must be equivalent to a Nash equilibrium. However, past results did not give a sufficient condition for patient stability, which leaves open the issue of the extent to which superstitions that arise in a Nash equilibrium are durable. We show that even patient players need not experiment enough to rule out all superstitions, and in particular false beliefs can survive about play that is more than one step off of the equilibrium path. This leads to our solution concept of “subgame-confirmed” equilibrium. Our central result is that this is a sufficient condition for patient stability.

As an example of a superstition that survived for quite some time, consider the Code of Hammurabi. The second of Hammurabi’s laws is “If any one bring an accusation against a man, and the accused go to the river and leap into the river, if he sink in the river his accuser shall take possession of his house. But if the river prove that the accused

is not guilty, and he escape unhurt, then he who had brought the accusation shall be put to death, while he who leaped into the river shall take possession of the house that had belonged to his accuser.” This law seems to be based on the superstition that the guilty are more likely to drown than the innocent. If people are this superstitious, why use such an elaborate mechanism? Why not simply assert that those who are guilty will be struck dead by lightning, while the innocent will not be? If this is believed, it will be as effective at preventing crime as the Hammurabi mechanism, and it does not require witnesses or judges or any of the other complicated and costly elements of the Hammurabi code.

To understand the logic behind our analysis, suppose that players are indoctrinated into a social norm as children – for example “if you commit a crime you will be struck by lightning” – and enter the world as young adults with a prior belief that it is very likely that the social norm is true. The players are patient, rational Bayesians, so when they are young they optimally decide to commit a few crimes to see what will happen. In the case of the lightning-strike norm, most young players will discover that the chances of being struck by lightning are independent of whether they commit crimes, and so go on to a life of crime, thereby undermining the norm. The Hammurabi case is more complex: the social norm is to not commit crimes and to only accuse the guilty. If older people adhere to this norm, what happens? Young potential criminals commit crimes, are accused of crimes, and are punished, so they learn that crime does not pay, and as they grow older stop committing crimes. But what about the young accusers? The critical fact is that the accusers only get to play the game after a crime takes place. As we have described the situation, there are few crimes, hence accusers only get to play infrequently.<sup>2</sup> Infrequent play reduces the value of experimentation, because there will likely be a long delay before the knowledge gained can be put to use. Our results suggest that even patient and rational young accusers will not experiment with false accusations, and so they will never learn that the river is as likely to punish the innocent as the guilty. In practice, there may be other sources of experimentation than the rational learning that is the focus of this paper, and one might expect that all actions will in fact have a positive, albeit very small, probability. We examine the robustness of our findings to such forces. Briefly, the robust implication of our theory is that “two-steps-off-the path

---

<sup>2</sup> It is also possible, for example, that the accuser is the criminal, in which case the accuser may get to play frequently. We discuss this and related issues in after explaining our main results.

superstitions” will be more durable than false beliefs either on or one step off of the path of play.

## 2. The Hammurabi Games

We begin by giving several stylized games inspired by the example of the Hammurabi superstition. These games are not intended to be detailed or accurate representations of the situation contemplated in the Code of Hammurabi. They are intended rather to capture the basic idea of superstitions that might or might not be located on the equilibrium path. Later we discuss how these examples and our results may help us to understand the actual Code of Hammurabi and other similar types of superstitions.

### Example 2.1: The Hammurabi Game

Our version of the “Hammurabi game” has two players, a suspect and an accuser. The suspect, player 1, moves first and may either **exit** or commit a **crime**. If the suspect **exits** the game ends. If the suspect chooses **crime**, the accuser, player 2, gets to move, and may either tell the **truth** or **lie**.

Both players get 0 if there is **exit**. If a **crime** is committed, and the accuser tells the **truth**, the suspect is thrown in the river, resulting in the suspect being punished with probability  $p$  and the accuser with probability  $1 - p$ . If the accuser **lies** a falsely accused third party not explicitly represented in the game is thrown in the river and the accuser is punished with probability  $1 - p$ .

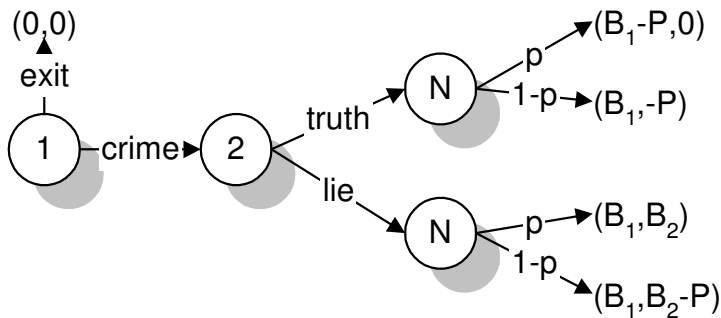
If the crime is committed the payoffs depend on whether the accuser tells the **truth** and whether he is punished.

	Accuser not punished	Accuser punished
<b>truth</b>	$B_1 - P, 0$	$B_1, -P$
<b>lie</b>	$B_1, B_2$	$B_1, B_2 - P$

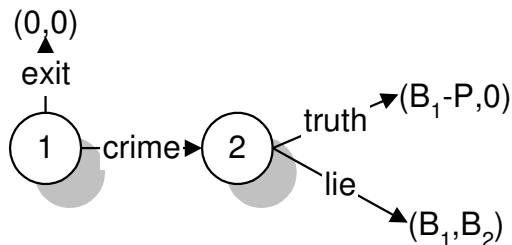
Here  $B_1$  is the benefit of the **crime** to the suspect,  $B_2$  is the benefit of a false accusation to the accuser and the cost of punishment  $P$  is the same for both. We assume that  $B_1 < pP$  so that the true probability of drowning is sufficient to deter **crime**, and that

$B_2 < pP$ , so that  $-(1-p)P > B_2 - P$ . This implies that an accuser who knows how often guilty people drown ( $p$ ) and believes that innocents never drown will prefer to accuse the guilty. Note that as long as the true probability that the accused drowns is independent of guilt, it is in fact optimal for the accuser to lie. Note also that our restrictions on the parameters are consistent with the idea that  $P$  is large, i.e. the players really dislike drowning.

The game is illustrated in the extensive form below.

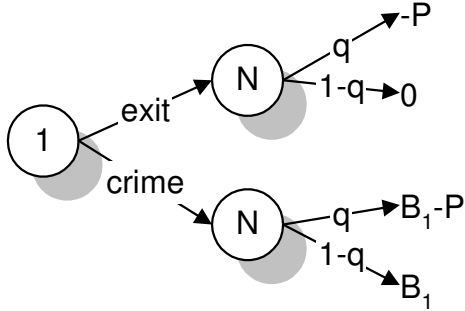


Example 2.2: The Hammurabi Game Without a River



In the Hammurabi game without a river is similar to the Hammurabi game, but there is no river. The suspect is always punished if the accuser tells the **truth**, and the accuser is never punished.

Example 2.3: The Lightning Game



In the lightning game there is no accuser, and the suspect is punished with probability  $q$ , regardless of whether a **crime** is committed or what the accuser does. Here we assume that  $B_1 < (1 - q)P$ .

Each of these three games has a strategy profile where crimes are always committed, and a profile in which there is no crime. The no-crime profile in the Hammurabi game is for the accuser to tell the **truth**, because he believes that if he **lies** he will be punished with probability 1. In the Hammurabi game without a river, no crime occurs when the accuser tells the **truth**; this is weakly optimal for the accuser because he is indifferent. In the lightning game, crime is deterred if everyone believes that if they commit a **crime** they will be punished with probability 1, and that if they **exit** they will be punished with probability  $q$ . Our results will imply that only the Hammurabi game with a river has a patiently stable state with no crime.

### 3. Simple Games

This paper focuses on a special class of games where there is a straightforward sufficient condition for patient stability. A *simple game* is a game of perfect information (each information set is a singleton node) in which each player has at most one information set on each path through the tree. He may have more than one information set, but once he has moved, he never gets to move again. The Hammurabi game with and without a river and the lightning game are simple games.

To begin we specify some notation. There are  $I + 1$  players in the game, where player  $i = I + 1$  is nature. The game tree  $X$  with nodes  $x \in X$  is finite. The terminal nodes are  $z \in Z \subset X$ . Nodes are partially ordered by precedence, so if  $x$  follows  $x'$  we write  $x' \leq x$ . Since information sets are singleton nodes, we also use  $X$  to denote the information sets. Information sets where player  $i$  has the move are denoted by  $X_i \subset X$ , while  $X_{-i} \equiv X \setminus X_i$  are the information sets for other players (or nature). The feasible

actions at information sets  $x \in X_i$  are denoted  $A(x)$ . The initial information set is denoted by  $x = 0$ . A pure strategy for player  $i$ ,  $s_i$ , is an action at each information set in  $X_i$ ,  $s_i(x) \in A(x)$ ;  $S_i$  is the set of all such strategies. We let  $s \in S = \times_{i=1}^{I+1} S_i$  denote a pure strategy profile for all players including nature, and  $s_{-i} \in S_{-i} = \times_{j \neq i} S_j$ . Each strategy profile determines a terminal node  $\zeta(s) \in Z$ . We suppose that all players know the structure of the extensive form – that is, the game tree  $X$  and action sets  $A(x)$ . Hence, each player knows the space  $S$  of strategy profiles and can compute the function  $\zeta$ . Each player  $i$  receives a payoff in the stage game that depends on the terminal node. Player  $i$ 's payoff function is denoted  $u_i : Z \rightarrow \mathfrak{R}$ . We let  $U \equiv \max_{i,z,z'} |u_i(z) - u_i(z')|$  denote the largest difference in utility levels.

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  $i$ ,  $\pi_i$ , assigns information sets in  $X_i$  a probability distribution over feasible actions,  $\pi_i(x) \in \Delta(A(x))$ ;  $\Pi_i$  is the set of all such strategies. For a fixed  $s_i$ , the marginal probability of a node  $x \in X(s_i)$  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  $s_i$ . We may extend this definition to mixed strategies  $X(\sigma_i)$  by requiring that the nodes or information sets be reachable with positive probability; we will make use of both mixed and behavior strategies for reasons that will become clear shortly. We will also need to refer to the information sets that are reached with positive probability under  $\sigma$ , 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  $\Pi_{-i}$ , the set of other players' behavior strategies. Let  $\mu_i$  denote the belief of player  $i$ . For a fixed  $s_i$ , the marginal probability of a node  $x \in X(s_i)$  is determined by  $\mu_i$ :

$$p_i(x | \mu_i) = \int p_i(x | \pi_{-i}) \mu_i(d\pi_{-i}).$$

The support of this distribution defined to be the set  $\bar{X}(s_i, \mu_i)$ . The distribution  $p_i(\cdot | \mu_i)$  generates a utility function on strategies:

$$u_i(s_i, \mu_i) \equiv u_i(s_i, p_i(\cdot | \mu_i)) \equiv \sum_{z \in Z(s_i)} p_i(z | \mu_i) u_i(z).$$

Frequently  $\mu_i$  has a continuous density  $g_i$  over  $\pi_{-i}$ . In this case we write  $p_i(x | g_i)$ ,  $u_i(s_i, g_i)$ , and  $\bar{X}(s_i, g_i)$ .

Since each player moves at most once along any path of play, Kuhn's Theorem implies that for any mixed strategy profile  $\sigma$  there exists a unique behavior strategy profile  $\underline{\pi}$  that is observationally equivalent to  $\sigma$ .<sup>3</sup> We say that player  $i$ 's belief  $\mu_i$  is *correct* at an opponent  $j$ 's information set  $x$  if  $\mu_i(\{\pi_{-i} | \pi_j(x) = \underline{\pi}(x)\}) = 1$ . In our learning model, there are many agents in the role of each player, and each agent will play a pure strategy, so that a state of the system will be a vector of probability distributions  $\bar{\theta} = (\bar{\theta}_1, \dots, \bar{\theta}_I, \bar{\theta}_{I+1})$ , where each  $\bar{\theta}_i$  is a distribution over the pure strategies of player  $i$ , and  $\bar{\theta}_{I+1} = \sigma_{I+1}^0$  is the exogenous distribution over Nature's move. Henceforth we will use  $\bar{\theta}$  to stand for mixed strategy profiles, and let  $\bar{\Theta}$  be the set of all mixed strategy profiles. We will need to make use of mixed strategy profiles in order to allow for the fact different agents in the role of a given player may have different beliefs.

#### 4. Final-Move Admissibility and Subgame Confirmed Nash Equilibrium

We turn next to concepts of equilibrium. Let the penultimate nodes be those all of whose immediate successors are terminal nodes; these nodes represent the "final moves" in the tree.<sup>4</sup> We refer to a profile as *final-move admissible* if no player plays a sub-optimal action at any final move (that is, penultimate node). We will see later that all play in the learning model must be final-move admissible; intuitively, at these nodes beliefs about  $\pi_{-i}$  are irrelevant and the player has a simple choice between alternatives with known payoffs.

In addition to this restriction, the steady states of the learning model will have the property that players have correct beliefs about play at nodes where they have infinitely many observations. Just which nodes these are is endogenous, and will depend on the discount factor, as increasing patience leads to an increased amount of

---

<sup>3</sup> Note that because we restrict attention to simple games, the issue of defining player  $i$ 's conditional play at an information set that player  $i$ 's own strategy makes unreachable does not arise.

<sup>4</sup> A node that is not a penultimate node can be the last move *by a player* if all of its successors correspond to moves by Nature. These are not "final moves" according to our definition; at such nodes a player's expected payoff depends on his beliefs about Nature.

“experimentation.” Our first notion of equilibrium, self-confirming equilibrium, corresponds to the case of myopic players who do no experimentation at all. Thus it imposes only the restriction that players learn what happens on the equilibrium path.

**Definition 4.1:**  $\bar{\theta}$  is a self-confirming equilibrium if for each player  $i$  and for each  $s_i$  with  $\bar{\theta}_i(s_i) > 0$  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  $x \in \bar{X}(s_i, \bar{\theta}_{-i})$ ,

It is important to note that this definition allows player  $i$  to rationalize each  $s_i$  in the support of  $\bar{\theta}_i$  with a different beliefs. This is because we want the equilibrium concepts we develop to correspond to steady states of learning models with anonymous random matching: In those models, there will be many agents in the role of each player, and different agents may hold different beliefs. Note also that *Nash equilibrium* differs by strengthening (b) to hold at all information sets. Finally, note that self-confirming equilibrium allows players to have any beliefs about opponents’ play that are not contradicted by their observations. The “rationalizable self-confirming equilibrium” of Dekel, Fudenberg and Levine [1999] strengthens this concept by restricting attention to beliefs that are consistent with almost common knowledge of the payoff functions.<sup>5</sup>

Our goal is to understand the consequences of optimal off-path experimentation by patient players. First we define what it means to be “one-step off the equilibrium path.”

**Definition 4.2:** In a simple game, node  $x$  is *one step off the path of  $\pi$*  if it is an immediate successor of a node that is reached with positive probability under  $\pi$ .

Our learning theory will imply that players have some knowledge of off-path play, but less knowledge about off-path play than about on-path play. The relevant equilibrium concept is

---

<sup>5</sup> See also Rubinstein and Wolinsky [1994].

**Definition 4.3:** Profile  $\pi$  is a *subgame-confirmed equilibrium* if it is a Nash equilibrium and if, in each subgame beginning one step off the path, the restriction of  $\pi$  to the subgame is self-confirming in that subgame.

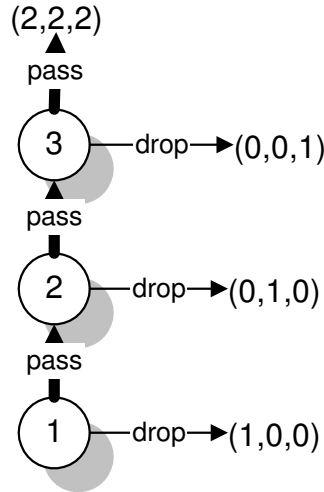
Subgame-confirmed equilibrium strengthens Nash equilibrium by imposing the requirement that (one-step) off-path play be a self-confirming equilibrium. Every subgame-perfect equilibrium is subgame confirmed; the subgame-confirmed condition is weaker in two respects. First, subgame-confirmed equilibrium imposes no constraints on play in subgames that are two or more steps off of the equilibrium path. Second, in the subgames that are one step off of the equilibrium path, subgame-confirmed equilibrium asks only that play is a self-confirming equilibrium, instead of requiring Nash equilibrium. In our learning model, it turns out that players need not acquire enough information about play two-steps off the equilibrium path to force their beliefs to be accurate there. So the relevant equilibrium concept only restricts beliefs to be accurate at subgames that are one step off of the equilibrium path, which is why the play in these subgames need only be a self-confirming equilibrium: The conclusion that play one step off the equilibrium path must be a Nash equilibrium would require that beliefs two steps off the path must be correct.

Before turning to the model of steady state learning, we consider some simple examples to illustrate the contrast between subgame-confirming equilibrium and subgame perfect equilibrium.

First, consider a simple game with no more than two consecutive moves. Here self-confirming equilibrium for any player moving second implies optimal play by that player. Consequently, subgame-confirmed equilibrium implies subgame perfection. Our next example shows how this fails when there are three consecutive moves.

#### Example 4.1: The Three Player Centipede Game

Three players move in order. If a player **drops** the game ends, if he **passes** the next player gets to move. Payoff are given in the diagram below: basically everyone prefers to **pass** if he thinks the next player is going do so, and **drop** if he thinks the next player is going to drop.



The unique subgame-perfect equilibrium is clearly for all players to **pass**. However we claim that **(drop, drop, pass)** is final-move admissible and subgame-confirmed.<sup>6</sup> It is obviously a Nash equilibrium, since player 1 is playing a best response to player 2's strategy of **dropping**. We must also have that **drop, pass** is self-confirming in the subgame beginning with player 2's move. It is, since if player 2 **drops**, he does not see player 3's move, and so may believe that player 3 is **dropping**, even though this is incorrect. The point is that subgame perfection requires beliefs to be correct in all subgames; subgame-confirmed equilibrium requires them only to be correct on the path of the subgame that starts one step from the equilibrium path.<sup>7</sup>

This example leaves open the issue of whether a subgame-confirmed equilibrium is path-equivalent to the requirement that the profile yield a *Nash* equilibrium at every node that is one step off of the path. A more elaborate 4-player centipede example shows how the two differ.<sup>8</sup> Intuitively, we create a conflict between player 1's and player 2's incentive constraints, so that for them both to play as specified, player 3 must randomize. However, we can structure the subgame starting with player 2 passing so that randomization by 3 is possible in self-confirming but not Nash equilibrium. Because the details are somewhat complicated, they can be found in Appendix B.

<sup>6</sup> The profile **(drop, drop, drop)** is subgame-confirmed but not final-move admissible.

<sup>7</sup> Past work had suggested that subgame perfection is not necessary for patient stability; see the discussions in Fudenberg and Kreps [1996] and Fudenberg and Levine [1999].

<sup>8</sup> The "*k*-step perfection" of Kalai and Neme [1992] imposes Nash equilibrium at all nodes *k* or fewer steps off of the path, so the example shows that subgame-confirmed equilibrium is not equivalent to "1-step perfection."

We should point out that replacing a terminal node by a move by Nature with the same expected payoff under the objective distribution on Nature's move can enlarge the set of subgame-confirmed payoffs unless the original terminal node was on the equilibrium path. This change has no impact in the usual model where players are assumed to have correct beliefs over all moves by Nature, but it is natural in a setting where players need to observe Nature's moves to learn them. Similarly, there are many other transformations of games that have no impact on the set of Nash equilibria, but that matter here: In a Bayesian learning model, the keys are what players are assumed to know at the start and what they observe when the game is played, that is, their priors and likelihood functions.<sup>9</sup> In a similar vein, but the opposite direction, there are changes to the game tree, such as moving Nature's move to the beginning of the game, but making it unknown to the players until later, that do not change what players learn at the end of the game: these types of changes cannot change the steady states of the learning model.<sup>10</sup>

### 5. Rational Steady-State Learning

*The Agent's Decision Problem:* We now consider an "agent" in the role of player  $i$ . This agent expects to play the game  $T$  times and wishes to maximize

$$\frac{1 - \delta}{1 - \delta^T} E \sum_{t=1}^T \delta^{t-1} u_t$$

where  $u_t$  is the realized stage game payoff at  $t$  and  $0 \leq \delta < 1$ .

The agent believes that he faces a fixed time invariant probability distribution of opponents' strategies, but is unsure what the true distribution is. This belief will be correct in the steady states we analyze, and approximately correct in the neighborhood of a stable steady state.<sup>11</sup>

---

<sup>9</sup> It is easy to extend our model to allow players to have objectively known distributions over certain moves by Nature, presumably ones that the players observe frequently either in this or some other game.

<sup>10</sup> However they can change the set of subgame confirmed equilibrium by changing the set of subgames; analogous issues arise in comparing subgame perfection and sequential equilibrium.

<sup>11</sup> A model of out-of-equilibrium learning must allow the players' beliefs to be systematically wrong, as the only way to avoid this is to assume that play in the overall system corresponds to an equilibrium. (Aumann [1987].) Thus the issue is not whether the beliefs are always correct, but whether we should expect the agents to detect the errors, which depends on the cost of the error and the difficulty of detecting it. Thus the assumption that players think the world is stationary is more plausible in cases when the world is at least approximately stationary, as in neighborhood of stable steady state. Aoyagi [1994], Foster and Vohra [1997], Fudenberg and Levine [1999], and Lambson and Probst [2004] study learning in games when agents try to detect various sorts of time-varying patterns such as cycles.

**Definition 5.1:** A belief  $\mu_i$  is non-doctrinaire if it is given by a continuous density function  $g_i$  that is strictly positive at interior points.

Note that this definition allows priors to go to zero on the boundary.<sup>12</sup>

Player  $i$  is assumed to have a prior  $g_i^0$  that is non-doctrinaire and independent. This belief is updated using Bayes Law: We let  $g_i(\cdot | z)$  denote the posteriors starting with prior  $g_i$  after  $z$  is observed. It is straightforward to show that non-doctrinaire priors imply non-doctrinaire posteriors.

**Optimal Play in the Agent's Decision Problem:** Each agent observes only his own play and the terminal nodes in games that he has played; the *private history* of agent  $i$  through time  $t$  is a sequence  $(s_i(1), z_i(1), \dots, s_i(t), z(t))$ . Let  $Y_i$  be the set of all such histories with length no more than  $T$ , and  $t(y_i)$  denote the length of history  $y_i \in Y_i$ . There is also a null history 0.

Let  $g_i(\cdot | y_i)$  be the posterior density over opponent's strategies given sample  $y_i$ , and let  $p_i(\cdot | y_i)$  be the corresponding distribution over terminal nodes. Let  $V_i^k(g_i)$  denote the maximized average discounted value (in current units) starting at  $g_i$  with  $k$  periods remaining. Bellman's equation is

$$V_i^k(g_i) = \max_{s_i \in S_i} \left[ (1 - \phi_k) u_i(s_i, g_i) + \phi_k \sum_{z \in Z(s_i)} p_i(z | g_i) V_i^{k-1}(g_i(\cdot | z)) \right]$$

where  $V_i^0(g_i) = 0$  and  $\phi_k = \delta(1 - \delta^{k-1}) / (1 - \delta^k)$ . Let  $s_i^k(g_i)$  denote a solution of this problem. It is convenient to abbreviate  $V_i^k(g_i(\cdot | y_i))$  as  $V_i^k(y_i)$ ,  $s_i^k(g_i(\cdot | y_i))$  as  $s_i^k(y_i)$ , and  $u_i(s_i, g_i(\cdot | y_i))$  as  $u_i(s_i | y_i)$ .

An *optimal policy* is a map  $r_i : Y_i \rightarrow S_i$  defined by  $r_i(y_i) = s_i^{T-t(y_i)}(g_i(\cdot | y_i))$ . Notice that there can be more than one optimal policy; for example several strategies may

---

<sup>12</sup> We use this definition, as opposed to the stronger version with densities that are uniformly bounded away from zero, because posterior beliefs will typically assign probability 0 to distributions that are inconsistent with the sample – that is, after seeing one “Heads,” the posterior density is 0 at the point “always Tails.” The assumption rules out players being *certain* that a particular opponent's action is dominated, but it allows them to believe that this is true with high probability; we view it as mild and very reasonable. At the technical level, Bayesian updating can have bizarre consequences when the true state is in a neighborhood that has prior probability 0, while the non-doctrinaire assumption lets us appeal to the Diaconis-Freedman [1990] result that Bayesian posteriors converge to the empirical distribution function.

be strategically equivalent. Note also that there will always be an optimal policy that is deterministic.

Recall that a strategy  $s_i$  is final-move admissible if it prescribes final-move-admissible actions at every final move of player  $i$ , and say that a policy is final-move admissible if it prescribes a final-move admissible strategy for every history.

**Lemma 5.1:** *Every optimal policy is final-move admissible.*

*Proof:* The optimal policy will assign probability 0 to an action unless it either (a) maximizes the current period's expected payoff or (b) increases expected payoff in future periods by providing information about actions that have a positive probability of being myopically optimal. However, final moves cannot generate any information, as they lead to terminal nodes, and we have assumed that players know the map from terminal nodes to payoffs.  $\square$

**Steady States in an Overlapping Generations Model:** We suppose that there is a continuum population, with a unit mass of agents in the role of each player. There is a doubly infinite sequence of periods; generations overlap, so there are  $1/T$  agents in each generation, with  $1/T$  new agents entering each population each period to replace the  $1/T$  oldest players who leave. Every period, each agent is randomly and independently matched with one agent from each of the other populations. In particular, the probability of meeting an agent of a particular age is equal to its population fraction  $1/T$ ; agents do not observe the ages or past experiences of their opponents.

We assume (by subdividing populations and adding player roles to the game if necessary) that each population  $i$  has a common prior, and uses a common deterministic optimal policy  $r_i$ .<sup>13</sup> Suppose we are given the fractions of each population  $\bar{\theta}_i(s_i)$  of each population that play the corresponding  $s_i$ . Using the rule  $r_i$  we may then work out the fractions  $f_i^T[\bar{\theta}](y_i)$  of the population with each experience  $y_i$ . The new entrants have no experience, so  $f_i^T[\bar{\theta}](0) = 1/T$ . We then calculate iteratively for each  $(y_i, r_i(y_i), z)$

$$f_i^T[\bar{\theta}](y_i, r_i(y_i), z) = f_i^T[\bar{\theta}](y_i) \sum_{\{s_{-i} | z \in Z(r_i(y_i), s_{-i})\}} \prod_{k \neq i} \bar{\theta}_k(s_k). \quad (*)$$

---

<sup>13</sup> For example, if one third the player 1's have prior  $p$  and two-thirds have prior  $q$ , we can view this as two distinct populations called "1p" and "1q." Each period, each player 2 then has probability 1/3 of matching with a player 1p.

Denote the resulting distribution over histories as  $f^T[\bar{\theta}] = (f_1^T[\bar{\theta}], \dots, f_T^T[\bar{\theta}])$ . We can then compute the population fractions playing each strategy:

$$\bar{f}_i^T[\bar{\theta}](s_i) = \sum_{\{y_i | r_i(y_i) = s_i\}} f_i^T[\bar{\theta}](y_i)$$

This is a polynomial map from the space  $\Sigma$  of mixed strategy profiles to itself, and so has a fixed point. These fixed points are the *steady states* of the system.<sup>14</sup>

**Patient Stability:** For each non-doctrinaire prior  $g^0$ , discount factor  $\delta < 1$  and length of life  $T$  there are optimal rules, and steady states with respect to those rules  $\bar{\Theta}(g^0, \delta, T)$ . If for some fixed  $g^0$  and  $\delta$  there is a sequence  $\bar{\theta}^T \in \bar{\Theta}(g^0, \delta, T)$  with  $\lim_{T \rightarrow \infty} \bar{\theta}^T \rightarrow \bar{\theta}$ , we say that  $\bar{\theta}$  is a  $\delta$ -stable state. If  $\bar{\theta}(\delta)$  are  $\delta$ -stable states and  $\lim_{\delta \rightarrow 1} \bar{\theta}(\delta) \rightarrow \bar{\theta}$ , we say that  $\bar{\theta}$  is a *patiently stable state*.

We will say that two profiles  $\bar{\theta}, \bar{\theta}'$  are *path equivalent* if they induce the same distribution over terminal nodes.

**Theorem 5.1:** (Fudenberg and Levine [1993b])  *$\delta$ -steady states are self-confirming equilibria; patiently stable states are Nash equilibria.*<sup>15</sup>

Note that a strategy profile is stable or patiently stable if there exists a non-doctrinaire prior such that the relevant conditions are satisfied. In general, we expect the set of steady states to depend on the prior.<sup>16</sup> Note also that since steady states exist for all lifetimes, and the space of population fractions is compact, patiently stable states exist.

---

<sup>14</sup> If we consider steady states of the deterministic dynamical system whose state is the fraction of agents with each history, the strategy frequencies in those steady states correspond to steady states as defined here. In our earlier work [1993b] we defined steady states in the larger space of fraction of agents with each history. However, it is technically easier to deal with steady states in the smaller space of strategy frequencies, since this space does not change as we vary the length of life. The two definitions are equivalent: given population fractions with each history and the optimal rule, we can easily compute the unique strategy frequencies; given the strategy frequencies and the optimal rules, we can work the optimal strategies forward to uniquely find the steady state population fractions with each history as shown in (\*).

<sup>15</sup> Our [1993b] paper states this result for the case where agents know the distribution of Nature's move, but the result extends to the present setting. The key fact is that our argument showed that in patiently stable state, each  $\bar{\theta}_i$  must maximize  $u_i(s_i, \bar{\theta}_{-i})$ , regardless of how  $\bar{\theta}_{-i}$  is generated.

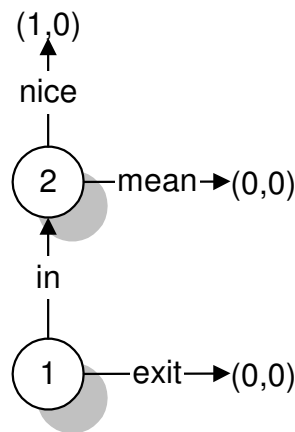
<sup>16</sup> Think for example of the steady states in a symmetric coordination game.

## 6. Patient Stability in Simple Games

This section presents our main results, and uses them to analyze the Hammurabi games that were presented in Section 2. The main result of this paper is loosely speaking that in simple games, a subgame-confirmed equilibrium is path-equivalent to a patiently stable steady state. To prove this, we must first rule out some types of weakly dominated strategy. The problem is illustrated by a simple two-player game “niceness” game.

### Example 6.1 The Niceness Game

Player 1 moves first, either **exit** or **in**. If he exits both players get zero. If he plays



**in**, player 2 can be **nice** or **mean**. Player 2 gets zero either way, but if he is **mean** player 1 gets zero, while if he is **nice**, player 1 gets one.

It is a subgame-confirmed equilibrium, indeed subgame perfect, for player 1 to **exit**, and player 2 to play **mean**. But player 1 knows his payoff to exit is zero, and with non-doctrinaire priors, his posterior is non-doctrinaire, so he has a positive expected payoff relative to his posterior by playing **in**. So in any steady state he must play **in**, which shows that being subgame-confirmed is not always sufficient for patient stability.

This problem can be avoided assuming that there are no ties in payoffs, but this would rule out the Hammurabi game with a river, since the suspect only cares whether he is punished or not, and there are a number of ways he may fail to be punished. Instead, we will make the weaker assumption that no player has two different actions at an information set that can possibly result in a tie in his own payoff. In addition, we require that this “no-ties” property holds when the game is transformed by moving all of Nature’s

moves to the end of the game and then replacing each of Nature's moves with a terminal node assigning the vector of expected utilities generated by that move. Notice that the first condition is satisfied for generic assignments of payoff vectors to terminal nodes, and that in a game in which the first condition is satisfied, the second is satisfied for generic assignments of probabilities to Nature. We refer to such games that satisfy both assumptions as having *no own ties*. This is satisfied in particular by the Hammurabi game: there are ties in the suspect's payoff function, but these ties all occur when he chooses to commit a crime, so two distinct own actions are not involved. Notice also that this assumption implies that a player playing in the final stage of the game has a unique best choice, and by backwards induction, every perfect information game with no own ties has a unique subgame perfect equilibrium.

We define a profile as *nearly pure* if there are no randomizations on the equilibrium path, and no player except Nature randomizes off the equilibrium path. Notice that our proposed Hammurabi game no-crime profile is nearly pure, as only Nature randomizes, and only off the equilibrium path.

**Theorem 6.1:** *In simple games with no own ties, a subgame-confirmed equilibrium that is nearly pure is path equivalent to a patiently stable state.*

The proof is in Section 7 below, here is a simplified sketch: To provide a sufficient condition, we can assume that beliefs are *independent*, so that players believe there is no correlation between how an opponent plays at different information sets, or how different opponents play; this implies that the only information a player has about play at a given node comes from observations of play at that node. When agents are patient and long-lived, most agents who are reached along the steady state path are playing a best response to the true distribution. These agents only play off path actions as experiments, or if their samples misleadingly suggest that off-path play is optimal. A strong-law-type argument says their samples are unlikely to be misleading, and for any fixed discount factor, the fraction of time that an agent experiments goes to 0. These results imply that agents off the path are reached a fraction of time that goes to 0, which in turn implies that off-path players are unlikely to get samples that suggest their nodes are likely to be reached. Thus, for a non-negligible set of priors, even very patient players do not do any experiments at off-path nodes.

Note that in simple games with no own ties, players will not randomize on the path of any Nash equilibrium. We do not know whether the restriction to nearly pure equilibria is necessary. In order for a subgame-imperfect equilibrium to be patiently stable, players must maintain incorrect beliefs at some parts of the game tree, which requires bounds on the amount of experimentation at off-path nodes. We have been unable to establish this bound when there is mixing on the equilibrium path. The difficulty is that the amount of experimentation at off-path nodes depends on how often those nodes are reached. We show in the proof of Theorem 6.1 that for a fixed discount factor  $\delta$  lack of randomization on the equilibrium path implies the frequency with which off-path nodes are reached goes to zero as  $T \rightarrow \infty$ . If there is randomization on the path, then some players will get sufficiently misleading samples of the path that they will “lock on” to playing a non-equilibrium action; this corresponds to players getting stuck on the wrong arm in a bandit problem. It is known that the fraction who get stuck falls to zero as  $\delta \rightarrow 1$ ; we can show that if the fraction falls faster than  $(1 - \delta)^2$  then Theorem 6.1 holds even with randomization on the equilibrium path. Unfortunately the rate of convergence of choosing the wrong arm in bandit problems does not appear to be known.<sup>17</sup>

The following partial converse to theorem 6.1 will show that patient stability has very different implications in the games with and without a river.<sup>18</sup> Recall that a profile is *final-move admissible* if no player plays a sub-optimal action at any penultimate node. Note that at penultimate nodes, beliefs are irrelevant to optimality.

**Theorem 6.2:** *A patiently stable state  $\bar{\theta}$  is a final-move admissible Nash equilibrium.*

---

<sup>17</sup> The argument underlying the need for  $(1 - \delta)^2$  is explained in footnote 23.

<sup>18</sup> We believe that if beliefs are “weakly independent” there is a full converse to theorem 6.1; that is, a patiently stable state must be path equivalent to a subgame confirmed Nash equilibrium. Because the key point of this paper is that superstitions that are subgame-confirmed can survive, we do not pursue this converse in detail. To sketch the argument, weak independence means that moves at a given node do not reveal information about play at other nodes. Without this assumption, off-path play may not be a response to incentives, but rather an attempt to gain information about play at some other part of the game tree. (For example, the accuser may experiment with false accusations in hopes of learning about the probability of crimes being committed.) With this assumption, however, we can conclude that players play optimally one-step off the path most of the time. Using an option value argument from our earlier paper, we can then show that most players beliefs about certain “decisive” off-path nodes must be nearly correct most of the time. Optimal play together with correct beliefs one-step off the path gives subgame confirmed equilibrium.

Note that this result applies to all games, not just to simple ones.

*Proof:* Our past work showed that a patiently stable steady state must be a Nash equilibrium. Lemma 5.1 shows that every optimal policy is final-move admissible, so the same is true of any steady state and any limit of steady states.

☑

Note that in games with length at most two, a final move admissible Nash equilibrium is subgame perfect, while a subgame-confirmed equilibrium is as well.

### Example 2.3 Continued: The Lightning Game

In the lightning game, the no-crime profile is a self-confirming equilibrium, since the information set for nature at which a **crime** is committed is not observed. It is not a Nash equilibrium, since the suspect is not playing a best response to Nature's strategy. Hence the no-crime profile is not patiently stable.

### Example 2.2 Continued: The Hammurabi Game Without A River

In the game without the river, profile **(exit, truth)** is a Nash equilibrium, because the accuser is off the path of play and so is willing to tell the truth. However, in his final move it is optimal for the accuser to **lie**, so **(exit, truth)** is not final-move admissible, hence is not patiently stable. The only Nash equilibrium where the accuser lies is **(crime, lie)**, so by Theorem 6.2 this is the only patiently stable state,

### Example 2.1 Continued: The Hammurabi Game

In the Hammurabi game, if the suspect **exits**, the only subgame that is one step off the equilibrium path is the game in which the accuser decides whether or not to **lie**. In this subgame, it is self-confirming for him to tell the **truth**, and believe he will not be punished for telling the **truth**, which gives payoff  $-(1-p)P$  each time a crime is committed; this equilibrium is supported by the belief that if he were to **lie** he would be punished with probability one, receiving  $B_2 - P$ . (This conclusion uses our assumption that  $pP > B_2$ .) So **(exit, truth)** is a subgame-confirmed equilibrium, and hence by Theorem 6.1, it is patiently stable. Moreover, **(exit, truth)** and **(crime, lie)** are the only

Nash equilibrium outcomes, so the set of patiently stable states is path-equivalent to the set of subgame-confirmed equilibria.

Before proceeding to the proof of Theorem 6.1, we provide a sufficient condition for patient stability that endogenizes the restriction to nearly pure strategies. We will say that a game has “length at most three” if no path through the tree hits more than three information sets.

**Lemma 6.3:** *In simple games with no own ties, no Nature’s move and length at most three, a subgame-confirmed equilibrium is path equivalent to a subgame-confirmed equilibrium in which players play pure strategies.*

The example in Appendix B shows the role of the assumption of length at most three. That game has length four, and as we saw there is a subgame-confirmed Nash equilibrium that is not path equivalent to a pure subgame-confirmed equilibrium. Our proof of Lemma 6.3 uses the following result on self-confirming equilibria in games of length at most two:

**Lemma 6.4:** *In simple games with no own ties, no Nature’s move and length at most two, every self-confirming equilibrium is path equivalent to a public randomization over Nash equilibria.*

*Proof:* Fix a self-confirming equilibrium  $\pi$ , and let the first player be player 1. Each strategy that has positive probability under  $\pi$  is a best response to some belief about other player actions in all other subgames. In particular it is a best response to the belief that following every other action  $s_1$  the player  $j$  that follows chooses the action that is worst for player 1 in that subgame; call these actions  $\underline{s}_j(s_1)$ . Moreover, because there are no own ties, in each subgame that is reached by  $\pi$ , player  $j \neq 1$  plays a pure strategy; call these  $s_j^*(s_1)$ . Thus for each  $s_1$  in the support of  $\pi$ , the profile

$$\begin{aligned} s_1 &= s_1, \\ s_j(s_1) &= s_j^*(s_1), \\ s_j(s_1) &= \underline{s}_j(s_1), s_1 \neq s_1 \end{aligned}$$

is a Nash equilibrium, so the self-confirming equilibrium  $\pi$  is path-equivalent to a public randomization over pure-strategy Nash equilibria.



*Proof of Lemma 6.3:* Fix a subgame-confirmed equilibrium of a game of length at most three. For each first-player action that has zero probability, specify that play in the resulting subgame will be one of the Nash equilibria that is worst for the first player moving. These continuation equilibria will be in pure strategies, and because the self-confirming equilibrium specified for these subgames were randomizations over Nash equilibria, picking the worst Nash equilibrium will preserve the first player's incentives not to deviate. Finally, the assumption of no own ties implies that the first player cannot randomize, so the strategies we have constructed are pure.



Lemma 6.3 and Theorem 6.1 yield the following corollary:

**Theorem 6.5:** *In simple games with no own ties, no Nature's move and length at most three, a subgame-confirmed equilibrium is path equivalent to a patiently stable state.*

Although the class of simple games with no Nature's move and length at most three is quite special, it includes many important games that have been extensively studied by experimentalists, including the ultimatum, best shot, chain store, peasant-dictator, and trust games. In the even more special, but also important case of games of length at most two, Theorem 6.1 and the equivalence of both final-move admissibility and subgame-confirmed equilibrium to subgame perfection gives rise to the following very sharp result:

**Theorem 6.6:** *In simple games with no own ties, no Nature's move and length at most two, the set of subgame-perfect equilibria is path equivalent to the set of patiently stable states.*

## 7. Proof of Theorem 6.1

We will now give the proof of Theorem 6.1.

**Theorem 6.1:** *In simple games with no own ties, a subgame-confirmed Nash equilibrium that is nearly pure is path equivalent to a patiently stable state.*

Let  $\hat{\pi}$  be a nearly-pure subgame confirmed equilibrium. Define a function on states  $\bar{\theta}$  (that is, distributions over strategies) as follows:

$$\lambda(\bar{\theta} | \hat{\pi}) = (\lambda_0(\bar{\theta} | \hat{\pi}), \lambda_1(\bar{\theta} | \hat{\pi})),$$

where  $\lambda_0$  is the maximum of the difference between the probabilities assigned by  $\bar{\theta}$  and  $\hat{\pi}$  to any pure action at any information set on the path of  $\hat{\pi}$ , and  $\lambda_1$  is the same maximum over information sets one step off the path of  $\hat{\pi}$ .

Now consider a  $\bar{\theta}$  such that  $\lambda(\bar{\theta} | \hat{\pi}) = (\varepsilon_0, \varepsilon_1)$ . Recall that  $\bar{f}^T[\bar{\theta}]$  is the play generated by the optimal dynamic learning rules in the environment defined by  $\bar{\theta}$  when players live  $T$  periods, and that  $f^T[\bar{\theta}]$  is the associated distribution over histories. In outline, our proof of the theorem relies on showing that there are (non-doctrinaire) priors such that the maps  $\bar{f}^T : \bar{\Theta} \rightarrow \bar{\Theta}$  map certain neighborhoods of  $\hat{\pi}$  to themselves, where the neighborhoods are defined by the  $\lambda$ -metric and  $\bar{\Theta}$  is the set of all mixed strategy profiles. We will conclude that the maps have a sequence of fixed points that converge to a suitable limit as  $T \rightarrow \infty$ . This limit need not be  $\hat{\pi}$ ; we only establish that the limit is path equivalent to it.

The proof uses a combination of new results specific to simple games and more general lemmas about rational learning and the law of large numbers, some of which are new and others we take from our previous work. This section states and proves the lemmas about simple games; Appendix A collects all of the more general statistical lemmas, and gives proofs for the lemmas that are new.

Turning to the details of the proof, we will measure the distance between two beliefs of player  $i$  by the distance (in the sup norm) between their expected values, that is by the maximum difference in the probabilities assigned to any pure action at any node, and we will measure the distance between beliefs and the state  $\bar{\theta}$  in the same way.

Since each  $\hat{\pi}_i$  is a best response to  $\hat{\pi}_{-i}$ , and there are no own ties, each player's action at each information set on the path of  $\hat{\pi}$  is a strict best response to the actual play of the other players. Therefore there is an  $\bar{\varepsilon} > 0$  such that each player's on-path actions are a strict best response to any  $\pi_{-i}$  that is within  $\bar{\varepsilon}$  of  $\hat{\pi}_{-i}$  at every information set. In addition, every player  $i$ 's actions at nodes one step off the path are also a strict best response to some strictly positive belief  $\hat{\mu}_i$  that supports  $\hat{\pi}$  as subgame confirmed. Moreover, there is such a  $\hat{\mu}_i$ , and an  $\tilde{\varepsilon} > 0$  such that for any belief within  $\tilde{\varepsilon}$  of  $\hat{\mu}_i$  any action that is not an (*ex ante*) best response to  $\hat{\pi}$  has expected payoff (relative to that belief) of at least  $\tilde{\varepsilon}$  lower than that of the best response.

We say that priors are  $n, \varepsilon$ -strongly accurate for a node  $x$  if fewer than  $n$  observations can not make the expected probability of actions at that node differ from  $\hat{\mu}$  by more than  $\varepsilon$ . Define  $\underline{n} \equiv 2^{11} / \bar{\varepsilon}^4$ . We say that priors are *strongly accurate* if they are  $\underline{n}, \bar{\varepsilon}$ -strongly accurate at all nodes.

Since we are free to choose any non-doctrinaire priors in order to prove the Theorem, we can specify that the priors are independent across opponents and information sets, and come from the Dirichlet family. Specifically, we set

$$g^0(\pi_{-i}) = \prod_{j \neq i, x \in X_j} g_x^0(\pi_j(x)),$$

where  $g_x^0(\pi_j(x))$  is a Dirichlet distribution on  $\Delta(A(x))$  with prior mean  $\hat{\mu}_j(x)$  and “initial intensity”  $\gamma(x)$ . Thus, when  $n$  observations have been acquired at  $x$  and observed play there corresponds to  $\hat{p}_x$ , the posterior mean (i.e. expected play) at  $x$  is the mixed strategy  $(\gamma \hat{\mu}_j(x) + n \hat{p}_x) / (\gamma + n)$ .

The first lemma shows that if priors are strongly accurate then beliefs about on-path play “are close to”  $\hat{\pi}$ . This is useful both in showing that most players in  $f^T(\bar{\theta})$  conform to the path of  $\hat{\pi}$  (Lemma 7.3) and in showing that there is little experimentation off of the path of play (Lemma 7.5.)

**Lemma 7.1:** *If priors are independent Dirichlet and strongly accurate, then for all  $\bar{\theta}$  such that  $\lambda(\bar{\theta} \mid \hat{\pi}) = (\varepsilon_0, \varepsilon_1)$  with  $\varepsilon_0 < \bar{\varepsilon} / 2$ , and all  $\delta, T$ , the fraction of agents in  $f^T[\bar{\theta}]$  whose beliefs about on-path play are more than  $\bar{\varepsilon}$  from  $\hat{\pi}$  is no more than  $\varepsilon_0 / 2$ .*

*Proof:* Since beliefs are independent, player  $k$  learns nothing about the on-path play of other players at information sets that come after hers in periods in which she deviates from  $\hat{\pi}$ . Consequently,  $k$ 's belief about on-path play at any information set at any date  $n$  is obtained by using the  $m \leq n$  observations of that information set that are available from periods where she did not deviate. Since the posterior mean of the agent's belief will be a convex combination of the prior and the sample, and strongly accurate priors are within  $\bar{\varepsilon}$  of  $\hat{\pi}$ , whenever the sample is within  $\bar{\varepsilon}$  of  $\hat{\pi}$ , the posterior will be within  $\bar{\varepsilon}$  of  $\hat{\pi}$  as well. From the assumption of strongly accurate priors, we know that there is no sample path of length less than  $\underline{n}$  that can make any player  $k$ 's posterior belief about  $j$ 's play be at least  $\bar{\varepsilon}$  from  $\hat{\pi}$ . It is thus sufficient to show that, of the agents with samples of length  $\underline{n}$  or more at node  $x$ , the fraction whose sample is more than  $\bar{\varepsilon}$  from

$\hat{\pi}$  is no more than  $\varepsilon_0/2$ . Since  $\bar{\theta}$  is within  $\bar{\varepsilon}/2$  of  $\hat{\pi}$ , we will show that of the agents with samples of length  $\underline{n}$  or more at node  $x$ , the fraction whose sample is more than  $\bar{\varepsilon}/2$  from  $\bar{\theta}$  is no more than  $\varepsilon_0/2$ . This will follow from a version of the law of large numbers.

Since on-path play of  $\hat{\pi}$  is pure, there is a single terminal node  $z^*$  to which  $\hat{\pi}$  assigns probability 1.<sup>19</sup> For each player  $j$  who plays on the equilibrium path of  $\hat{\pi}$ , let  $I_j(z)$  be the indicator function which takes on the value 1 if  $j$  deviated from  $\hat{\pi}$  and 0 if  $j$  conformed. Let  $\mu_j = EI_j(z)$  be the expected value of  $I_j$  under  $\bar{\theta}$ , and let

$$S_{j,n} = \frac{\left| \sum_{k=1}^n (I_j(z_k) - \mu_j) \right|}{n}$$

be the deviation of the sample average of  $I_j$  from its mean. Lemma A.1 from Appendix A implies that<sup>20</sup>

$$\Pr(\max_{\underline{n} \leq n \leq \bar{n}} S_{j,n} > \varepsilon) \leq \frac{2^7}{3} \frac{1}{\underline{n}} \frac{\mu_j}{\varepsilon^4} = \frac{\bar{\varepsilon}^4}{3} \frac{\mu_j}{2^4 \varepsilon^4},$$

where the equality comes from the definition of  $\underline{n}$ .

If the play prescribed by  $\bar{\theta}$  is within  $\varepsilon_0$  of  $\hat{\pi}$  at every information set on the path of play, then  $\mu_j \leq \varepsilon_0$ , and substituting this and taking  $\varepsilon = \bar{\varepsilon}/2$  we have

$$\Pr(\max_{\underline{n} \leq n \leq \bar{n}} S_{j,n} > \bar{\varepsilon}/2) \leq \frac{\varepsilon_0}{3}.$$

So, regardless of  $T$ , at most  $\varepsilon_0/3$  of the agents can have samples of length  $\underline{n}$  or more that differ from  $\bar{\theta}$  at information sets on the equilibrium path by at least  $\bar{\varepsilon}/2$ .

☑

Next we want to argue that players on the path of play are unlikely to have beliefs about off-path play that make them want to deviate. If player  $i$  plays on the path of  $\hat{\pi}$  and  $a$  is a deviation for player  $i$  from the path of  $\hat{\pi}$ , we say that his belief is  $a, \varepsilon$  *off-path deviation inducing* if there exists a strategy profile  $\tilde{\pi}_{-i}$  for the opponents that is within  $\bar{\varepsilon}$  of  $\hat{\pi}$  at on-path information sets such the strategy corresponding to  $\tilde{\pi}_{-i}$  at on-path information sets, and the strategy

<sup>19</sup> The nearly pure assumption simplifies the presentation, but is not essential for this Lemma; the extension to on-path mixing by Nature requires that we specify a different and generally larger  $\underline{n}$ .

<sup>20</sup> Note that  $\bar{n}$  on the left does not matter, since it does not appear on the right.

$$\pi_{-i}(\mu_i) = \int \pi_{-i}\mu_i(d\pi_{-i})$$

generated by the player's actual belief about play at off-path nodes, imply a loss of no more than  $\varepsilon$  from playing  $a$  rather than the path of  $\hat{\pi}$ . Note that in simple games, a player's belief about play following some other deviation  $a'$  are irrelevant for whether the beliefs are  $a, \varepsilon$  off-path deviation inducing, as is the player's belief about play at successors of  $a$  to which the player assigns sufficiently low probability.

**Lemma 7.2:** *Suppose that all agents have priors that are independent, Dirichlet, and strongly accurate. For any  $\varepsilon < \tilde{\varepsilon}$  and any state  $\bar{\theta}$  with  $\lambda(\bar{\theta} | \hat{\pi}) = \varepsilon_1 < \bar{\varepsilon}$ , and any  $\delta$ , as  $T \rightarrow \infty$  the fraction of agents in  $f^T(\bar{\theta})$  who play  $a$  and have beliefs that are  $a, \varepsilon$  off-path deviation-inducing goes to 0.*

*Proof:* In outline, we will show that for any  $\varepsilon' > 0$  the fraction of agents in  $f^T[\bar{\theta}]$  who play  $a$  and have beliefs are  $a, \varepsilon$  off-path deviation-inducing is no larger than  $\varepsilon'$ . This will follow from the fact that the true state  $\bar{\theta}$  is not off-path deviation-inducing and the strong law of large numbers.

To make this precise, let  $X(a, \bar{\theta}_{-i})$  be the set of nodes that have positive probability when player  $i$  plays  $a$  and the distribution of other player's play is given by  $\bar{\theta}_{-i}$ . Let  $x$  be the node where  $a$  is feasible. Define  $\hat{p}_x(a | y_i)$  to be the frequency in the history  $y_i$  with which  $a$  has been played when  $x$  has been reached. Let  $\bar{\pi}(a | \bar{\theta})$  be the behavior strategy corresponding to  $\bar{\theta}$  according to Kuhn's Theorem. Let  $n(x | y_i)$  be the number of times  $x$  has been hit given the sample  $y_i$ .

Now consider the information that player  $i$  has about play at successors of action  $a$ . Lemma A.2 shows that for all  $\varepsilon' > 0$  there is an  $N$  such that for all  $T, i, \bar{\theta}, x', a' \in A(x')$ ,

$$f_i^T[\bar{\theta}] \{ |\hat{p}_{x'}(a' | y_i) - \bar{\pi}_{-i}(a' | \bar{\theta})| > \varepsilon', \text{ and } n(x' | y_i) > N \} \leq \varepsilon'/3.$$

That is, at any node  $x'$ , only a few players (a) have seen that node be reached many times and (b) have observations that are substantially different from  $\bar{\theta}$ . Moreover, the share of such players can be made small by taking  $N$  sufficiently large. In particular, this is true at every node that is one step off of the equilibrium path, and every feasible action  $a'$  at such information sets. From that same lemma, for each node  $x'$ , and any  $N$  and  $\varepsilon'$ , there is an  $N'$  such that the fraction of players who have played  $a'$  more than  $N'$  times and

seen  $x'$  fewer than  $N$  times is less than  $\varepsilon'$ . Since  $X$  is finite, for any  $N$  and  $\varepsilon'$ , there is an  $N'$  such the fraction of players who have played  $a'$  more than  $N'$  times and seen any  $x' \in X(a', \bar{\theta}_{-i})$  fewer than  $N$  times is less than  $\varepsilon'/3$ . Since  $a$  has finitely many successors, the same statement is also true simultaneously for all successors of  $a$ . Hence, fewer than  $\varepsilon'/3$  player i's have samples that differ from  $\bar{\theta}_{-i}$  by more than  $\varepsilon$ .

Now fix an  $\varepsilon'$  such that  $\varepsilon' + \varepsilon_1 < \bar{\varepsilon}$  and the corresponding  $N, N'$ . By taking " $N$ " in the previous paragraph equal to  $N'$ , and considering the action  $a$  on the equilibrium path rather than  $a'$  one-step off-path, we may find an  $N''$  such that of those who have played  $a$  more than  $N''$  times, no more than  $\varepsilon'/3$  have fewer than  $N'$  observations on any  $x \in X(a, \bar{\theta}_{-i})$ , while of those who have more than  $N'$  observations on all  $x \in X(a, \bar{\theta}_{-i})$ , at most  $\varepsilon'/3$  have samples that differ from  $\bar{\theta}_{-i}$  by more than  $\varepsilon'$ .

Thus, discarding the two groups of size  $\varepsilon'/3$  we need to consider those players who play  $a$  but have played it fewer than  $N''$  times, and those who have more than  $N'$  observations on any  $x \in X(a, \bar{\theta}_{-i})$  and have samples that differ from  $\bar{\theta}_{-i}$  by less than  $\varepsilon'$ . But as  $T \rightarrow \infty$  the fraction of histories in which  $a$  is currently played, but has been played fewer than  $N''$  times necessarily goes to zero, so certainly drops to smaller than  $\varepsilon'/3$ .

This leaves those players who have more than  $N'$  observations on any  $x \in X(a, \bar{\theta}_{-i})$  and have samples that differ from  $\bar{\theta}_{-i}$  by less than  $\varepsilon'$ . Since priors are strongly accurate, they are accurate, so these players' beliefs at  $x \in X(a, \bar{\theta}_{-i})$  are within  $\varepsilon' + \varepsilon_1 < \bar{\varepsilon}$  of  $\hat{\mu}$ . Since  $X(a, \bar{\theta}_{-i})$  are the nodes reached with positive probability at  $\bar{\theta}_{-i}$  when  $a$  is played, beliefs at other reachable nodes given  $a$  are equal to the prior, that is  $\hat{\mu}$ .<sup>21</sup> By definition of  $\hat{\mu}$  and  $\tilde{\varepsilon}$  it follows that  $a$  has an expected loss of at least  $\tilde{\varepsilon}$ . Since  $\varepsilon < \tilde{\varepsilon}$  these players' beliefs are not  $a, \varepsilon$  off-path deviation inducing.  $\square$

Using Lemmas 7.1 and 7.2, we can conclude there are few deviations from the path of  $\hat{\pi}$ .

---

<sup>21</sup> We do not need the full strength of this assumption, as beliefs two steps off the equilibrium path can be shown not to matter, but proving this requires additional argument. As we are free to pick the prior, we chose it to make the proof as easy as possible.

**Lemma 7.3:** *Suppose that all agents have priors that are independent Dirichlet and strongly accurate. For any  $\varepsilon_0 > 0$  there is a  $T$  so that in  $\bar{f}^T(\bar{\theta})$  the fraction of players who deviate at a node on the path of  $\hat{\pi}$  is no greater than  $\varepsilon_0$ .*

*Proof:* Fix an  $\varepsilon \in (0, \bar{\varepsilon})$ . From the definition of an off-path deviation-inducing belief, a player who deviates at an on-path node either (i) does not play an  $\varepsilon$ -static best-response to his belief, (ii) has a belief that is  $a, \varepsilon$ -off-path deviation inducing for some  $a$ , or (iii) has a belief that is wrong by more than  $\bar{\varepsilon}$  about on-path play. The first class of agents goes to 0 with  $T$  by Lemma A.4, since  $\hat{\pi}$  is a strict equilibrium.<sup>22</sup> The second class goes to 0 with  $T$  from Lemma 7.2, and the third class is no more than  $\varepsilon_0/2$  from Lemma 7.1.

☑

Next we want to argue that play must be close to  $\hat{\pi}$  at nodes one step off of the equilibrium path. To do so, we first bound beliefs about play at those nodes.

**Lemma 7.4:** *For all  $\varepsilon_1$ , there exists an  $N$  such that if priors are independent Dirichlet and  $N, 2\varepsilon_1$ -strongly accurate at all nodes one step-off the path of  $\hat{\pi}$ , then for all  $\bar{\theta}, \varepsilon_0$  such that  $\lambda(\bar{\theta} \mid \hat{\pi}) = (\varepsilon_0, \varepsilon_1)$  and all  $\delta, T$ , the fraction of agents in  $f^T(\bar{\theta})$  whose beliefs about one-step-off-path play are more than  $2\varepsilon_1$  from  $\hat{\pi}$  is no more than  $\varepsilon_1/2$ .*

*Proof:* Denote by  $f_{2\varepsilon_1}$  the fraction of agents in  $f^T(\bar{\theta})$  whose beliefs about one-step-off-path play are more than  $2\varepsilon_1$  from  $\hat{\pi}$ . To bound  $f_{2\varepsilon_1}$ , recall that for any  $\varepsilon'$  Lemma A.2 yields an  $N$  such that fewer than  $\varepsilon_1/4$  players have seen a node more than  $N$  times and have a sample of play at that node that differs from the  $\bar{\theta}$  by more than  $\varepsilon'$ . Since the prior about this node is concentrated near  $\hat{\pi}$ , and  $\bar{\theta}$  is within  $\varepsilon_1$  of  $\hat{\pi}$  at this nodes, by choosing  $\varepsilon'$  sufficiently small, these players have beliefs that are within  $2\varepsilon_1$  of  $\hat{\pi}$  at those nodes. On the other hand, because we have assumed that priors are  $N, 2\varepsilon_1$ -strongly accurate one-step-off the path, players who have seen the node fewer than  $N$  times have beliefs that are within  $2\varepsilon_1$  of  $\hat{\pi}$  at those nodes.

☑

Finally we use Lemmas 7.1 and 7.4 to conclude that one step off the path of play, most players' actions are a best response to their priors.

---

<sup>22</sup> In addition to the strong law, Lemma A.4 relies on the fact that the posterior distribution converges to the empirical c.d.f. at a uniform rate, as shown by Diaconis and Freedman [1990].

**Lemma 7.5:** Let  $\varepsilon_1 \leq \bar{\varepsilon}/2$  and let  $\hat{\mu}$  be independent Dirichlet priors that support  $\hat{\pi}$  as subgame-confirmed. For any  $\varepsilon_1$  there exists  $N$  such that if  $\hat{\mu}$  is strongly accurate and is also  $N, 2\varepsilon_1$ -strong one step-off the path of  $\hat{\pi}$ , then for all  $\delta$  there is an  $\varepsilon_0 > 0$  such that if  $\bar{\theta}$  satisfies  $\lambda(\bar{\theta} | \hat{\pi}) = (\varepsilon_0, \varepsilon_1)$ , then in  $\bar{f}^T[\bar{\theta}]$  the fraction of players who fail to play a best response to their priors is less than  $\varepsilon_1/2$ .

*Proof:* The actual probability of being off the path of  $\hat{\pi}$  goes to zero as  $\varepsilon_0 \rightarrow 0$ , and lemma 7.1 shows that as  $\varepsilon_0 \rightarrow 0$  the fraction of the population who ever believes that the probability of being off the path is large must be small. By Lemma A.5, a player who believes that the chance of being at a node is small relative to  $(1-\delta)^2$  will not experiment at that node, so as  $\varepsilon_0 \rightarrow 0$  most players play a best response to their beliefs whenever they are at nodes that are off the path of play.<sup>23</sup> Lemma 7.4 shows that most players have beliefs about one-step-off-path play less than  $2\varepsilon_1 < \bar{\varepsilon}$  from  $\hat{\pi}$ ; since they have never experimented, their best response to their beliefs are a best response to their priors. ☑

*Proof of Theorem 6.1:* We show that  $\hat{\pi}$  is a path equivalent to a patiently stable state. (Recall that a patiently stable state is a limit first as  $T \rightarrow \infty$  then as  $\delta \rightarrow 1$  of the steady state path of play.) Recall that we have fixed  $\bar{\varepsilon}, \underline{n}$ . Fix  $\varepsilon_1 \leq \bar{\varepsilon}/2$ . We may then choose  $N$  independent of  $T$  so that for any  $\delta$  there is an  $\varepsilon_0$  such that Lemma 7.5 holds with the fraction failing to play a best-response to their priors no greater than  $\varepsilon_1$ . Fix a prior  $\hat{\mu}$  that supports  $\hat{\pi}$  as subgame confirmed, that is strongly accurate (relative to  $\underline{n}, \bar{\varepsilon}$ ) and is also  $N, 2\varepsilon_1$ -strongly accurate one-step off the path. We will keep this prior fixed as we vary  $\delta, T$ . Fix  $\delta$ . Since by Lemma 7.5 the fraction failing to play a best-response to their priors one-step off of the path is no greater than  $\varepsilon_1$ , and  $\hat{\mu}$  supports  $\hat{\pi}$  as subgame-confirmed, this implies that all but  $\varepsilon_1$  play according to  $\hat{\pi}$  one-step off the path, that is  $\lambda_1(\bar{f}(\bar{\theta}) | \hat{\pi}) \leq \varepsilon_1$ . By choosing  $T$  large enough we can conclude from Lemma 7.3 that  $\lambda_0(\bar{f}(\bar{\theta}) | \hat{\pi}) \leq \varepsilon_0$ . Hence there is a fixed point, that is, steady state, with a path within  $\varepsilon_0$  of  $\hat{\pi}$ . Since  $\varepsilon_0$  can be arbitrarily small, this implies that the limit for each  $\delta$  as

---

<sup>23</sup> Because the probability of off-path nodes goes to 0 as  $T$  goes to infinity (due to the assumption that  $\hat{\pi}$  is nearly pure) the exact fraction of players failing to play a best response does not matter. If we drop the nearly-pure assumption, then as  $T$  goes to infinity we need to ensure that the fraction of players deviating from the path (and thus making mistakes given actual play) goes to zero with  $\delta$  at a rate faster than  $(1-\delta)^2$  so that we can apply Lemma A.5.

$T \rightarrow \infty$  is path equivalent to  $\hat{\pi}$ . As this remains true for the limit as  $\delta \rightarrow 1$ , this completes the proof.

☑

## **8. The Economics of Superstition**

The “Hammurabi Game” is only loosely motivated by the Code of Hammurabi. The situation contemplated in the actual code is different in a variety of respects. To what extent is the basic insight that two-step off-path superstitions are patiently stable relevant to the actual code of Hammurabi? Here we examine the simplifications made in our “Hammurabi Game” and how robust are our conclusions.

### ◆ *Death Penalty*

The most obvious feature of the Code of Hammurabi that differs from our model is that players who drown in the river or who are executed do not get to play again. We assume that information gained by an accuser or criminal who is punished is not lost to society. In practice, “individuals” in our model could be thought of as criminal gangs or families, so that the information about guilt, innocence and punishment is not lost to the gang. Moreover, the death penalty does not explain the durability of the Code. That is, the relevant learning is the learning by false accusers that they are not likely to get punished – this particular information is not lost to society as it is known to individuals who survive. Our theory provides an explanation of why there are not enough of these surviving false accusers to upset the social norm.

### ◆ *Frequency of Play*

It is important in our analysis that the accuser does not have control over whether the crime takes place: If he did, he would control also the frequency with which he would be able to learn about the consequences of a false accusation. So we rule out situations such as the doctor who murders his wife and says “my wife is dead and my the one-armed man did it.” Rather, we have in mind the following general type of situation: a murder occurs in the town square. Only one person – with no connection to the victim – is in a position to have observed the murderer. This person is then called upon to testify as a witness at the trial.

We do not have a great deal of information about whether the types of crimes committed in the time of Hammurabi were more of the private or public variety. However, while capital punishment was much more common in the Code of Hammurabi than it is today, there are many lesser punishments as well. Since the appeals procedure involves a substantial likelihood of both death and your heirs losing your house, it seems unlikely that the appeals procedure would have been frequently used when small punishments for small crimes were involved. So the question is largely about the frequency of major crimes.

In assessing the opportunities to experiment with false accusations, other aspects of the overall social norm are likely to be important. That is, after claiming you are a witness to a major crime, making the same claim a second time is likely to be met with some skepticism. This is certainly true in modern times, where, for example, a record of past accusations is generally viewed as reasons to treat current accusations with skepticism. Someone who constantly finds a severed human finger in her bowl of chili is not likely to be believed. This highlights the essential feature of our model: the endogeneity of frequency of play. As long as once a witness has appeared they are unlikely to choose to do so again, their incentive to acquire information about the consequences of false accusations is diminished regardless of why they are unlikely to play again.

#### ◆ *Other Sources of Experiments*

In practice, there may be other sources of experimentation in addition to the rational learning that is the focus of this paper. For example, we can introduce an exogenous probability that crime does pay. Most work on learning in extensive-form games, including the work discussed in the conclusion, has treated the frequency and timing of off-path “experiments” as exogenous; introducing random payoff shocks that serve as a source of “experimentation” has the same consequence. The robust conclusion of this work is that if every node is reached a positive fraction of the time, then the limit equilibrium must necessarily be subgame perfect, so that superstition cannot persist; such a theory obviously cannot be used as an explanation of the Code of Hammurabi.

At the same time, one potentially troubling aspect of our model is that in the limit of arbitrarily long lifetimes our model predicts that there will be no crimes at all, yet

almost certainly in the time of Hammurabi there were both crimes and false accusations. In this context, it is important to realize that it is the steady states for long but finite lifetimes that are intended to reflect reality and not the limit steady state itself. In the steady states with finite lifetimes, crimes are committed, false accusations take place, and indeed some individuals learn that making false accusations is a good idea. What is true is that in the limit all these things disappear, so that what matters is the relative probabilities of the exogenous and endogenous experiments.

The details of the proof of our main theorem provide additional information about the robustness of our results to the presence of other sources of noise. It is useful to distinguish between primary experimentation, meaning experimentation on the equilibrium path, and secondary experimentation, which takes place at nodes that are not. The former we measured by  $\lambda_0$ , the latter by  $\lambda_1$ . In the Hammurabi game, the former corresponds to experimentation with crimes; the latter corresponds to experimentation with false accusations.

Our results show that with patient players superstitions one step off the path cannot persist – and of course introducing additional noise only reinforces this conclusion – so the question is whether there is enough secondary experimentation to eliminate superstition. However, primary experimentation plays a role, because it determines the frequency with which secondary experimentation is possible, and increased primary experimentation increases the incentive to conduct secondary experimentation, as in Lemma A.5.

The proof of Theorem 6.1 shows that if  $\lambda_0$  and  $\lambda_1$  are bounded by small positive  $\varepsilon_0, \varepsilon_1$  respectively, then rational learning induces sufficiently little experimentation that these bounds are preserved for large  $T$  by the mapping  $\bar{f}^T$  defining the steady state. An implication of this line of proof is that other sources of noise that do not cause  $\lambda_0$  and  $\lambda_1$  to become too large will not upset the equilibrium.

With respect to primary experimentation  $\lambda_0$ , the bound  $\varepsilon_0$  that the proof requires decreases to zero as the discount factor goes to one. This is again an implication of the proof of Lemma A.5. If there is a fixed exogenous probability of crimes being committed and the discount factor goes to one, the superstition will unravel. But for any particular discount factor, the proof shows there is a threshold  $\varepsilon_0$  so that if the probability of crime

$\lambda_0$  remains below it, not enough secondary experimentation is triggered to upset the equilibrium.

Turning more specifically to the Hammurabi game, if the exogenous probability of crime is sufficiently low, then the probability of being called as a witness is also small, and the incentive to experiment with false accusations is small. Certainly in the modern day, the probability of being called as a witness at a trial is exceptionally small; most people are not called even once in a lifetime.

In summary, if opportunities to make accusations are limited, either endogenously or exogenously, superstitions about the consequences of play are liable to last a long time. If a lot of experimentation takes place with false accusations, or if many opportunities arise to make false accusations, then it would be surprising if a superstition about the consequences of false accusations would in fact survive.

#### ◆ *Model Details*

There are a number of dimensions in which our “Hammurabi game” simplifies the Code; many realistic details can be added to the game without changing the mathematical results. First is the endogeneity of appeals: we could have smaller and greater crimes, with the incentive to appeal dependent on the size of the crime, and explicitly introduce the fact that by appealing the stakes are raised by the loss of house along with life. Naturally if an appeal is unlikely to take place then the incentive to make a false accusation is greatly increased. But in this case no information about the probability of drowning in the river following a false accusation is generated, and so the analysis does not change.

Similarly, it is likely in many cases the witness might prefer to testify truthfully rather than making a false accusation against an enemy. This simply adds a branch to the tree that is not relevant to our analysis. Or there might be a third option for the witness, “do not testify.” Provided that the witness suffers a loss from having the true criminal escape sufficient to compensate for the cost of mistakenly being punished when telling the truth, the addition of such an option does not change our analysis.

## 9. Conclusion

We have shown that a patiently stable state must be path-equivalent to a Nash equilibrium in weakly undominated strategies, and that in games with no own ties, a subgame-confirmed equilibrium is path equivalent to a patiently stable state if the equilibrium is near pure or if the game has length at most three. These results lead to sharp predictions in some games of interest, such as the Hammurabi, ultimatum, best-shot, peasant-dictator, and trust games.

We are working on an extension of our analysis to the more general class of “games with identified deviators.” We conjecture that in these games only subgame-confirmed equilibria can be patiently stable. However, the result that every subgame-confirmed equilibrium is equivalent to a patiently stable state seems unlikely to generalize, which leaves open the question of determining a more restrictive necessary condition.

Nash equilibrium is “as if” players know the equilibrium path and the consequences of unilateral deviations from the equilibrium path. This is why learning in an extensive form need not in general lead to Nash equilibrium: to rule out non-Nash profiles, players must have “enough” observations of off-path play to learn the consequences of deviating. Equilibrium refinements such as subgame-perfect equilibrium are “as if” players know play throughout the entire game tree. This requires “enough” observations of play at most information sets, not just those that can be reached by a single deviation. Thus the two key issues for learning in extensive form games are (1) How much off-path play is needed for various refinements, and (2) How much off-path play should we expect to see? Much work, such as Fudenberg and Kreps [1988], [1995], [1996], Jehiel and Samet [2004], Noldeke and Samuelson [1993] and Hart [2002] has treated the amount of off-path play as being determined by exogenous experimentation rates. Fudenberg and Kreps worked with a model of boundedly rational learning in the style of fictitious play, and developed various assumptions that ensured that every node one step off the path of play is reached infinitely often, such as the condition that at each date  $t$ , each player is constrained to play each action (at the information sets that are reached) with probability at least  $1/t$ . They point out that this condition implies that nodes on the path are reached infinitely often, while nodes that are two or more steps off of the path may only be reached finitely many times; they are agnostic about whether one

should expect more or less experimentation than this at off-path information sets.<sup>24</sup> In their work on the convergence of boundedly rational learning in games of perfect information, Jehiel and Samet [2004] assume that there is a fixed, time-invariant probability of experimentation; this implies that every node is reached a positive fraction of the time.<sup>25</sup> In Noldeke and Samuelson [1993] and Hart [2002], off-path play occurs as the result of an exogenous “mutation” that leads an agent to use another strategy; this serves as an “experiment” from the viewpoint of the population because all agents get to observe the result of the mutants play.

The present paper, like our [1993b] work, differs in deriving the experimentation rule from the solution to the agent’s optimal decision. It is clear that impatient agents need not experiment at all, so we have focused on the play of very patient agents. The main force driving our results is that even patient agents need not experiment at nodes that are off of the path of play; this is why all subgame-confirmed equilibria are patiently stable.

---

<sup>24</sup> If on-path players experiment at rate  $1/t$ , a player at a node that is one step off of the path needs to experiment at a higher rate to make sure that he uses all actions infinitely often; our results show that a rational player would not chose to experiment that much.

<sup>25</sup> They also impose independent beliefs, and look at maximization in the agent normal form. For this reason, the steady states of their model correspond to Selten’s [1975]  $\varepsilon$  trembling hand perfection in the agent normal form; they show that the learning process they consider converges to this outcome in games of perfect information.

## Appendix A: Proofs

Let  $\{x_n\}$  be a sequence of i.i.d. binomial random variables with mean  $\mu$ , and define

$$S_n = \frac{\left| \sum_{k=1}^n (x_k - \mu) \right|}{n}.$$

**Lemma A.1**<sup>26</sup>:  $\Pr(\max_{\underline{n} \leq n \leq \bar{n}} S_n > \varepsilon) \leq \frac{2^7}{3} \frac{1}{\underline{n}} \frac{\mu}{\varepsilon^4}.$

*Proof:* We derive specific bounds based on the method of proof of the strong law of large numbers given by Billingsley [1995], p. 85. By Markov's inequality,

$$\Pr(S_n^4 > \varepsilon^4) \leq \frac{ES_n^4}{\varepsilon^4},$$

so

$$\begin{aligned} \Pr(\max_{\underline{n} \leq n \leq \bar{n}} S_n > \varepsilon) &= \Pr(S_{\underline{n}} > \varepsilon \text{ or } S_{\underline{n}+1} > \varepsilon \dots \text{ or } S_{\bar{n}} > \varepsilon) \\ &\leq \sum_{n=\underline{n}}^{\bar{n}} \Pr(S_n > \varepsilon) \leq \sum_{n=\underline{n}}^{\bar{n}} \frac{ES_n^4}{\varepsilon^4}. \end{aligned}$$

By collecting terms and using known inequalities, Billingsley shows

$$E(S_n)^4 \leq \frac{4E(x_1 - \mu)^4}{n^2},$$

and in the binomial case  $E(x_1 - \mu)^4 = \mu(1 - \mu)^4 + (1 - \mu)\mu^4 \leq 2\mu$ . So we conclude that

$$\Pr(\max_{\underline{n} \leq n \leq \bar{n}} S_n > \varepsilon) \leq \sum_{n=\underline{n}}^{\bar{n}} \frac{ES_n^4}{\varepsilon^4} \leq \frac{8\mu}{\varepsilon^2} \sum_{n=\underline{n}}^{\bar{n}} \frac{1}{n^2} \leq \frac{8\mu}{\varepsilon^4} \sum_{n=\underline{n}}^{\infty} \frac{1}{n^2}.$$

Finally, to estimate the sum, when  $\underline{n} = 1$  it is equal to  $\zeta(2) \leq 8/3$  where  $\zeta$  is the

Riemann zeta function. For  $\underline{n} > 1$  we have the bound

---

<sup>26</sup> The Lemma is stated for the case of binomial random variables, where its strength is proportional to the mean  $\mu$ , but it is true more generally. The key requirement for this “strong law of small numbers” is that the variance of the  $\{x_n\}$  be near 0.

$$\sum_{n=\underline{n}}^{\infty} \frac{1}{n^2} \leq \int_{\underline{n}-1}^{\infty} \frac{1}{n^2} dn = \frac{1}{\underline{n}-1} \leq \frac{2}{\underline{n}} \leq \frac{16}{3} \frac{1}{\underline{n}},$$

which gives the desired result. ☑

Let  $a \in A(x)$ . Define  $\hat{\pi}(a | y_i)$  to be the frequency with which  $a$  has been played when  $x$  has been reached. Let  $\bar{\pi}_i(a | \bar{\theta}_i)$  be the behavior strategy profile corresponding to  $\bar{\theta}_i$  according to Kuhn's Theorem, and let  $\bar{p}_i(x | \bar{\theta}_{-i})$  be the marginal probability of reaching  $x$  derived from  $\bar{\pi}(a | \bar{\theta})$  given an  $s_i$  such that  $x \in X(s_i)$ . Let  $n(x | y_i)$  be the number of times  $x$  has been hit given the sample  $y_i$ , and  $n(s_i | y_i)$  be the number of times  $s_i$  has been played.

**Lemma A.2** *For all  $\varepsilon, \varepsilon' > 0$  there is an  $N > 0$  for all  $T, r, \bar{\theta}, i, a \in A(x), s_i, x \in X(s_i), x \in X_j, j \neq i$*

$$(A.2.1) \quad f_i^T[\theta] \{ |\hat{\pi}(a | y_i) - \bar{\pi}_j(a | \bar{\theta}_{-i})| > \varepsilon, \text{ and } n(x | y_i) > N \} \leq \varepsilon'$$

$$(A.2.2) \quad f_i^T[\theta] \{ n(x | y_i) \leq [\bar{p}_i(x | \bar{\theta}_{-i}) - \varepsilon] n(s_i | y_i), \text{ and } n(s_i | y_i) > N \} \leq \varepsilon'.$$

*References:* Fudenberg and Levine [1993b] Lemma B.2. The first statement says that fewer than  $\varepsilon'$  of the population have both a large number  $N$  of observations at some node  $x$  and also a biased sample of play at that node. The second statement says that fewer than  $\varepsilon'$  of the population have played a strategy  $s_i$  that makes  $x$  reachable more than  $N$  times and yet have reached node  $x$  appreciably less often than the theoretical probability. Both of these statements are consequences of the strong law of large numbers, and more specifically the Glivenko-Cantelli theorem, which shows that the empirical distribution at each information set converges to the theoretical distribution as the sample size increases at a rate that holds uniformly over all theoretical distributions, i.e. over all possible values of  $\bar{\theta}$ . (The reason this lemma needs a proof is to show that the bound also hold over all strategies  $s_i$ , which corresponds to it holding over all sampling rules.)

Let  $r_i^T$  be optimal rules when life is  $T$  periods, and let  $r_i^k$  be optimal rules when  $k$  periods of life remain. The next lemma says that if the population fraction playing a strategy is strictly positive in the limit as  $T \rightarrow \infty$ , then population fraction that has played it  $N$  times can't be much less than the population fraction. This is simply a matter of bookkeeping, and not related to probability theory.

**Lemma A.3:** If  $\bar{\theta}_i(s_i) > 0$ , then

$$f_i^T[\bar{\theta}] \{ n(s_i | y_i) > N \text{ and } r_i^T(y_i) = s_i \} > \bar{\theta}_i(s_i) - (N/T).$$

*Reference:* Fudenberg and Levine [1993b] Lemma 5.7.

We define the event  $Y_i(\varepsilon)$  to be those  $y_i$  such that  $\max_{s_i} u_i(s_i | y_i) \leq u_i(r_i^k(y_i) | y_i) + \varepsilon$ . That is,  $Y_i(\varepsilon)$  is the set of histories for player  $i$  such that  $r_i^k(y_i)$  is an  $\varepsilon$ -best-response to the marginal belief at  $y_i$ .

**Lemma A.4:** For all  $\varepsilon, \varepsilon' > 0$  and  $\delta < 1$  there is an  $N$  such that for all  $\bar{\theta}, T$  such that

$$f_i^T[\bar{\theta}] \{ y_i \notin Y_i(\varepsilon) \text{ and } n(r_i^k(y_i)) > N \} \leq \varepsilon'.$$

*Reference:* Fudenberg and Levine [1993b] proof of Theorem 6.1. The intuition for this result is that if an agent has many observations of play at node  $x$ , then one more observation is not likely to change the posterior beliefs by very much, so that the “option” or “information” value” of experimenting at this node is likely to be low. Consequently, an optimal rule must prescribe an  $\varepsilon$ -best response with high probability.<sup>27</sup> Thus, only a few players can be playing an action  $a_i = r_i^k(y_i)$  that they have already played more than  $N$  times and which is not an  $\varepsilon$ -best-response to their beliefs.

**Lemma A.5 :** *If priors are independent in the sense that*

$$\mu_i(\pi_{-i}) = \prod_{j \neq i, x \in X_j} \mu_j(\pi_j(x))$$

---

<sup>27</sup> In the classical one-armed bandit problem, the agent stops experimenting in finite time so one might expect that we could take  $\varepsilon'$  to be 0 by taking  $N$  sufficiently large. However, as we explained in our earlier work, the fact that players know the structure of the game tree means that in some games there can be large but “unrepresentative” samples for which the value of further experimenting is still high. We conjecture that these samples cannot occur in simple games, so that we could indeed set  $\varepsilon' = 0$  for the purposes of this paper, but it is easier to appeal to the more general result.

then

$$\max_{s_i} u_i(s_i | y_i, x) - u_i(r_i^k(y_i) | y_i, x) \leq [\delta U / (1 - \delta)^2] \max_z p(x | y_i, z)$$

*Proof:* Set  $\Delta = \max_{s_i} u_i(s_i | y_i, x) - u_i(r_i^k(y_i) | y_i, x)$ . By assumption  $r_i^k(y_i)$  yields information that will only be of value only if  $x$  is reached again. The greatest value the information could have at that time is  $U$ . Let  $p_t = p(x | \tilde{y}_t)$  where  $\tilde{y}_t$  means that  $x$  was not reached during the previous  $t$  periods. Then

$$\begin{aligned} 0 &\leq -(1 - \delta)\Delta + \delta p_0 U + (1 - p_0)p_1 \delta^2 U + (1 - p_0)(1 - p_1)p_2 \delta^3 U \dots \\ &\leq -(1 - \delta)\Delta + \delta p_0 U + (1 - p_0)p_0 \delta^2 U + (1 - p_0)(1 - p_0)p_0 \delta^3 U \dots = \\ &-(1 - \delta)\Delta + \frac{p_0 \delta}{1 - \delta(1 - p_0)} U < -(1 - \delta)\Delta + \frac{p_0 \delta}{1 - \delta} U, \end{aligned}$$

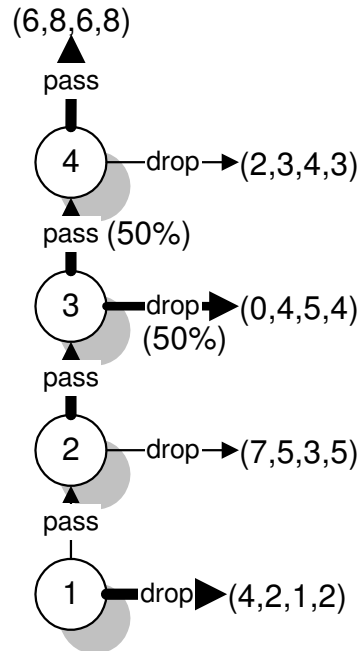
where the second inequality follows from the facts that  $p_t$  is non-increasing in  $t$ , and that increasing the probability  $p_t$  of stopping in period  $t$  decreases the expected waiting time and so increases the expected present value.

Note that only  $p_0$  is relevant; the strength of the belief is not. Also  $p_0 \leq \max_z p(x | y_i, z)$ , which gives the result.

☑

## Appendix B: The Four Player Centipede Game

Each of four players may either **drop** out, or **pass** the move to the next player, with payoffs shown in the diagram below. The bold lines indicate the subgame-confirmed equilibrium we propose to study. Here players 1, 2, and 4 are playing the best response to the actual distribution of opponent's play. Some player 3's **drop** and some **pass**, even though the payoff to **pass** is strictly higher; this is consistent with the definition because self-confirming equilibrium allows player  $i$  to rationalize each  $s_i$  in the support of  $\bar{\theta}_i$  with a different belief.



We claim first that in a subgame-confirmed equilibrium in which player 1 **drops** out, player 2 must **pass** with positive probability, and that player 3 must randomize. To see this, note first that player 2 must **pass** with positive probability in any Nash equilibrium where 1 **drops**, so the same is true for the more restrictive concept of subgame-confirmed equilibrium. Now consider the subgame starting with player 2's move. If 2 plays **pass** with positive probability in a self-confirming equilibrium, then 2's payoff to **pass** must be at least 5, so 3 must **pass** with probability at least .25. However, it is not consistent with subgame-confirmed equilibrium for player 1 to drop, player 2 to pass with positive probability, and player 3 to pass with probability 1, as then player 1 would get more than 4 from **pass**, regardless of 2's randomization probabilities, and moreover player 1 would know this, because all nodes would be at most one step off of the equilibrium path.

We claim next that the equilibrium above, in which player 1 drops, is not path equivalent to an equilibrium with Nash play at all nodes at most one step off of the path of play. To show this, consider the Nash equilibria of the subgame starting with player 2's move. If 2 **drops** with probability 1, then player 1 would **pass**, so player 2 must pass with positive probability for player 1 to **drop**. If 2 passes with positive probability, then Nash equilibrium requires optimal play by player 3. If player 3 **drops** with probability 1, then 2 would not be willing to pass, so 3 must **pass** with positive probability; this

requires that 4 play optimally and **pass**, so that 3 strictly prefers to **pass**, and then 2 strictly prefers to pass as well. But when players 2, 3, and 4 all **pass** with probability 1, player 1 prefers **pass** to **drop**.

The heart of this example is that there is a conflict between player 1's and player 2's incentive constraints, so that for them both to play as specified, player 3 must randomize. Yet in a Nash equilibrium of the subgame starting with 2's move, if player 2 passes and player 3 randomizes, player 4 must **pass**, so 3 must **pass** with probability 1.<sup>28</sup>

---

<sup>28</sup> Thus the self-confirming equilibrium in the subgame beginning with player 2's move in which player 3 randomizes is a counterexample to a claim made in Fudenberg and Levine [1997] that in games of perfect information self-confirming equilibria are public randomizations over Nash equilibrium. It is true for games where no path through the tree hits more than two information sets, as we prove in the process of proving Lemma 6.4.

## References

- Aoyagi, M. [1996] "Evolution of Beliefs and the Nash Equilibrium of Normal Form Games," *Journal of Economic Theory*, **70**: 444-469.
- Aumann, R. [1987] "Correlated Equilibrium as an Expression of Bayesian Rationality," *Econometrica* **55**,1-18.
- Billingsley, P. [1995]: *Probability and Measure, 3<sup>rd</sup> edition*, Wiley, New York.
- Dekel, E., D. Fudenberg and D. K. Levine [1999]: "Payoff Information and Self-Confirming Equilibrium," *Journal of Economic Theory* **89**: 165-185.
- Diaconis, P. and D. Freedman [1990]: "On the Uniform Consistency of Bayes Estimates for Multinomial Probabilities," *The Annals of Statistics* **18**: 1317-1327.
- Foster, D. and R. Vohra [1997]: "Calibrated Learning and Correlated Equilibrium" , *Games and Economic Behavior*, **21**:40-55.
- Fudenberg, D. and D.M. Kreps [1988] "A Theory of Learning, Experimentation, and Equilibrium in Games," mimeo.
- Fudenberg, D. and D. M. Kreps [1995]: "Learning in extensive games, I: self-confirming equilibrium, *Games and Economic Behavior* **8**, 20-55.
- Fudenberg, D. and D. M. Kreps [1996]: "Learning in Extensive Form Games, II: Experimentation and Nash Equilibrium," mimeo.
- Fudenberg, D., D. M. Kreps, and D. K. Levine [1988]: "On the Robustness of Equilibrium Refinements," *Journal of Economic Theory* **44**, 354-380.
- Fudenberg, D. and D. K. Levine [1993a]: "Self-Confirming Equilibrium," *Econometrica* **61**, 523-546.
- Fudenberg, D. and D. K. Levine [1993b]: "Steady State Learning and Nash Equilibrium," *Econometrica* **61**, 547-573.
- Fudenberg, D. and D. K. Levine [1997]: "Measuring Subject's Losses in Experimental Games," *Quarterly Journal of Economics*, 112: 508-536.
- Fudenberg, D. and D.K. Levine [1999] "Conditional Universal Consistency" *Games and Economic Behavior*, **29**: 104-130.
- Jehiel, P. and D. Samet [2004] "Learning to Play Games in Extensive Form by Valuation," mimeo.

- Kalai, E. and A. Neme, [1992] "The Strength of a Little Perfection," *International Journal of Game Theory*, 20, 335-355.
- Kreps, D. and R. Wilson [1982]: "Sequential Equilibria," *Econometrica* **50**, 863-894.
- Kuhn, H. [1953]: "Extensive games and the problem of information," *Annals of Mathematics Studies* no. 28, Princeton University Press, Princeton, NJ.
- Lambson, V. and D. Probst [1994] "Learning by Matching Patterns," *Games and Economic Behavior*, forthcoming
- Noldeke, G. and L. Samuelson [1993] "An Evolutionary Analysis of Forward and Backward Induction," *Games and Economic Behavior* **5**, 425-454
- Rubinstein, A. and A. Wolinsky [1994] "Rationalizable Conjectural Equilibrium: Between Nash and Rationalizability," *Games and Economic Behavior*, **6**, 299-311.
- Selten, R. [1975] "Reexamination of the perfectness concept for equilibrium points in extensive games," *International Journal of Game Theory* **4**, 25-55.