



HIER

Harvard Institute of Economic Research

Discussion Paper Number 2034

Steady State Learning and the
Code of Hammurabi

by

Drew Fudenberg and David K. Levine

May 2004

Harvard University
Cambridge, Massachusetts

This paper can be downloaded without charge from:

<http://post.economics.harvard.edu/hier/2004papers/2004list.html>

Steady State Learning and the Code of Hammurabi¹

Drew Fudenberg and David K. Levine²

This version: 5/3/04 First version: 6/5/03

Abstract: The code of Hammurabi specified a “trial by surviving in the river” as a way of deciding whether an accusation was true. This system is puzzling for two reasons. First, it is based on a superstition: We do not believe that the guilty are any more likely to drown than the innocent. Second, if people can be easily persuaded to hold a superstitious belief, why such an elaborate mechanism? Why not simply assert that those who are guilty will be struck dead by lightning? We attack these puzzles from the perspective of the theory of learning in games. We give a partial characterization of patiently stable 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 certain information sets reachable by a single deviation. We analyze this refinement and use it as a tool to study the broader issue of the survival of superstition. According to this theory Hammurabi had it exactly right: his law uses the greatest amount of superstition consistent with patient rational learning.

¹ We are grateful to NSF grant SES-9986170 and SES-0112018 and to Adam Szeidl for careful proofreading.

² Fudenberg: Department of Economics Harvard, dfudenberg@harvard.edu, <http://fudenberg.fas.harvard.edu>, Levine: Department of Economics UCLA, david@dklevine.com, <http://www.dklevine.com>.

1. Introduction

The first known written record of a mechanism is 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 is puzzling for two reasons. First, it is based on the superstition that the guilty are more likely to drown than the innocent. Second, 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.

Our perspective on these puzzles is that of the theory of rational Bayesian learning in extensive-form games.³ We argue that Hammurabi had it exactly right: his law uses the greatest amount of superstition consistent with patient rational learning. Using a model we developed in [1993b], we imagine society to consist of overlapping generations of finitely lived players. These players are indoctrinated into the 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 prior beliefs that it is very likely that the social norm is true. However, the players are rational Bayesians, and are relatively patient, 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

³ We discuss some of the related literature in the conclusion.

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. Infrequent play reduces the value of experimentation, because there will likely be a long delay before the knowledge gained can be put to use. We show that even 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.

To formalize this intuition, we 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.” Our [1993b] paper showed that these limits are necessarily Nash equilibria, but being a Nash equilibrium is not sufficient for patient stability. The present paper’s technical contribution is to refine this conclusion, providing a more restrictive necessary condition for patient stability that is also sufficient in the stylized “accusation game” that we use to illustrate the Hammurabi mechanism. Specifically, we show that, for the appropriate choice of priors, the Hammurabi mechanism does describe a patiently stable outcome of this game, but that the “lightning-strike” mechanism does not.

To see the impact of patient stability, consider a game with a single potential criminal and a single accuser. Player 1 moves first and may either **exit** or commit a **crime**. If player 1 **exits** the game ends; if he chooses crime, 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 player 2 tells the **truth**, player 1 receives a very low payoff, so that regardless of player 2’s payoff function, it is a Nash equilibrium for player 1 to exit and player 2 to tell the truth.⁴ We show that in patient stability requires that players act rationally one step off of the equilibrium path; if accusers have grudges against individuals other than the criminal, the Nash equilibrium in which they tell the truth will fail this additional test. This test is useful also for dealing with a broader set of issues concerning off-path play. For example, there may be several players playing in the subgame following a crime. Patient stability requires that they learn each other’s behavior, at least to the extent of self-confirming equilibrium.

⁴ Of course this outcome is not subgame perfect, but we will show that subgame perfection is not necessary for patient stability. Past work had suggested that this is the case, see the discussions in Fudenberg and Kreps [1994] and Fudenberg and Levine [1999].

To address the question of whether the Hammurabi mechanism is patiently stable, we give for the first time a sufficient condition for patient stability: at each subgame reachable from the equilibrium path by a single deviation, play in that game must be a self-confirming equilibrium in the sense of our [1993a] paper. In particular, we show that the Hammurabi equilibrium satisfies this condition. In future work we expect to be able to show that for a broader class of games this condition is also necessary for patient stability. To complete our analysis of the Hammurabi games, we give a weaker necessary condition that fails in the “lightning strike” game.

2. The Hammurabi Games

Example 2.1: The Hammurabi Game

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 $1 - p$.

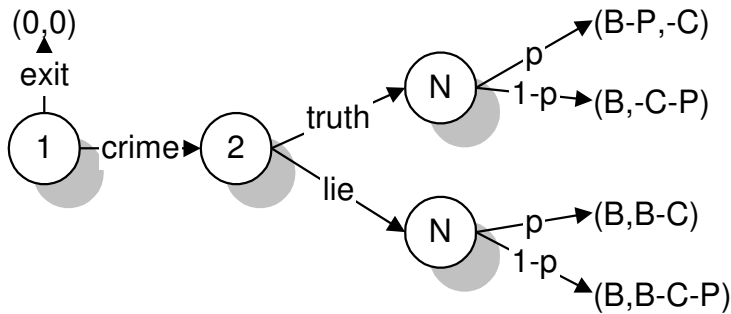
If the crime is committed the payoffs depend on whether the accuser tells the **truth** and whether he is punished.

	Accuser not punished	Accuser punished
truth	$B - P, -C$	$B, -C - P$
lie	$B, B - C$	$B, B - C - P$

Here we interpret C as the social cost of the **crime**, which to keep the game simple, we have borne by the accuser. To avoid excess notation, we take the benefit to the accuser of a false accusation, or **lie**, B to be the same as the benefit of the **crime** to the suspect, and the cost of punishment P to be the same for both. We assume that $B < pP$ so that the

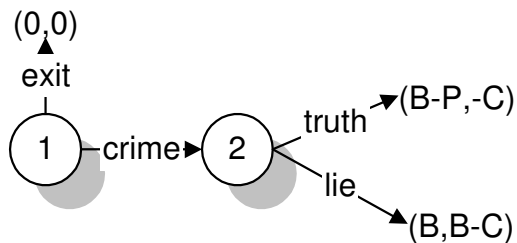
true probability of punishment is sufficient to deter **crime**. Note that as long as the probability that the accused drowns is independent of guilt, it is optimal for player 2 to lie.

The game is illustrated in the extensive form below.



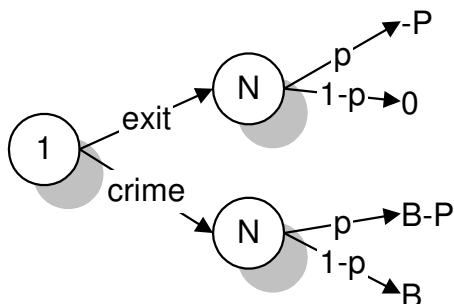
Example 2.2: The Hammurabi Game Without a River

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 p , regardless of whether a **crime** is committed or what the accuser does. Here



we assume that $B < (1 - p)P$.

Each of these three games has a configuration where crimes are always committed, and a configuration in which there is no crime. The no-crime configuration 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 p . 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 necessary and sufficient condition for social 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

ζ . Each player i receives a payoff in the stage game that depends on the terminal node. Player i 's payoff function is denoted $u_i : Z \rightarrow \Re$. 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 play i , π_i , assigns information sets in X_i a probability distribution over feasible actions, $\pi_i(x) \in \Delta(A(x))$; Π_i is the set of all such strategies. 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 also need to refer to the information sets that are reached with positive probability under σ , denoted $\bar{X}(\sigma)$.

We now model the idea that each player has beliefs about his opponents' play (including the play of Nature.) Let μ_i be a probability measure over Π_{-i} , the set of other players' behavior strategies. Throughout this paper we make the assumption that beliefs are *independent*, that is, that players do not believe that there is a correlation between how an opponent plays at different information sets, or how different opponents play.⁵ In other words,

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

For a fixed s_i , the marginal probability of a node $x \in X(s_i)$ is determined by the beliefs μ_i :

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

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

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

⁵ This means that what we call self-confirming equilibrium is independent self-confirming in the terminology of our [1993a] paper.

Frequently μ_i has a continuous density g_i over π_{-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, there is a unique behavior strategy profile $\underline{\pi}$ associated with any mixed strategy profile σ by Kuhn's Theorem.⁶ We say that player i 's belief μ_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, 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.

4. Subgame Confirmed Nash Equilibrium

We turn next to concepts of equilibrium. Our first notion of equilibrium is that of self-confirming equilibrium – this imposes the minimal restriction that players should 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 in the steady states of our learning model, 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 opponent's 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.⁷

⁶ Note that because we restrict attention to simple games, the usual issue of defining player i 's conditional play at an information set that player i 's own strategy makes unreachable does not arise.

⁷ See also Rubinstein and Wolinsky [1994].

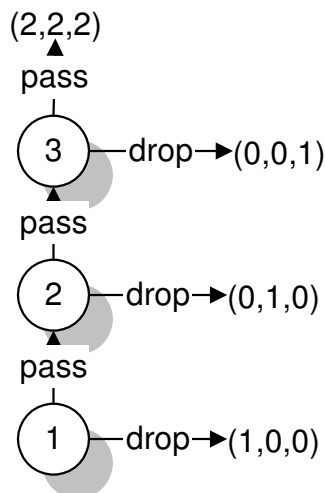
We strengthen Nash equilibrium through the refinement of subgame-confirmed Nash equilibrium. This requires self-confirming equilibrium in subgames one-step off the equilibrium path. As we will show it corresponds to the steady states of learning procedures in which rational Bayesian players experiment with off-path play.

Definition 4.2: *In a simple game, node x is one step off the path of π if it is an immediate successor of a node that is reached with positive probability under π . Profile π is a subgame-confirmed Nash equilibrium if it is a Nash equilibrium and if, in each subgame beginning one step off the path, the restriction of π to the subgame is self-confirming in that subgame.*

Before turning to the model of steady state learning, we first illustrate the notion of subgame-confirmed equilibrium through some simple examples. First, it is interesting to contrast subgame-confirming with subgame perfection. In a simple game with no more than two consecutive moves, self-confirming equilibrium for any player moving second implies optimal play by that player, so subgame-confirmed Nash equilibrium implies subgame perfection. The 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 to do so, and **drop** if he thinks the next player is going to drop.



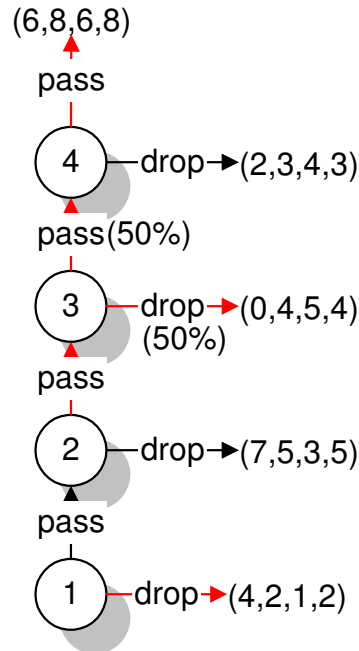
The unique subgame-perfect equilibrium is clearly for all players to **pass**. However we claim that **(drop, drop, pass)** is subgame-confirmed. It is obviously a Nash equilibrium, since player 1 is playing a best response to player 2's strategy of **dropping**. We must also have that **drop, pass** 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 Nash equilibrium requires them only to be correct on the path of the subgame that starts one step from the equilibrium path.

The next example shows that subgame-confirmed Nash equilibrium is not equivalent to the requirement that the profile yield a *Nash* equilibrium at every node that is one step off of the path.⁸

Example 4.2 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 red lines indicate the equilibrium we propose to study.

⁸ 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.”



Inspection of the game shows that in a subgame-confirmed Nash equilibrium in which player 1 **drops** out, player 3 must randomize, so in particular the equilibrium above, in which player 3 randomizes 50-50, is not path equivalent to a pure strategy subgame-confirmed Nash equilibrium, and that it is also not path equivalent to an equilibrium with Nash play at all nodes at most one step off of the path of play. In particular, the self-confirming equilibria of the subgame starting with player 2's move that are consistent with player 1 **dropping** require player 3 to randomize.

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 1.⁹

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

⁹ This 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 Theorem 5.2.

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

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.¹⁰

Definition 5.1: *Beliefs μ_i are non-doctrinaire if μ_i is given by a continuous density function g_i strictly positive at interior points.*

Note that this definition allows priors to go to zero on the boundary.¹¹

Player i is assumed to have a prior g_i^0 that is non-doctrinaire and independent. The assumption of independence makes updating beliefs very simple: We let $g_i(\cdot | z)$ denote the posteriors starting with prior g_i after z is observed:

$$g_i(\pi_i | z) = p_i(z | \pi_{-i}) g_i(\pi_{-i}) / p_i(z | g_i).$$

It is straightforward to show that non-doctrinaire priors imply non-doctrinaire posteriors.

Optimal Play in the Agent's Decision Problem: The agent observes only his own play and the terminal nodes in games that he has played; the *private history* of the agent 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

¹⁰A 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

¹¹We 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.”

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

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 will be 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 $r_i(y) = s_i^{T-t(y)}(g_i(\cdot | y_i))$. Notice that there can be more than one optimal policy; for example several strategies may be strategically equivalent. Note also that there will always be an optimal policy that it deterministic. The combined assumptions of independent beliefs and simple games makes each of an agent's actions correspond to an independent "bandit problem," in the sense that using a given action provides no information about what would have happened had an alternative action been chosen.¹²

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 $1/T$ players in each generation, with $1/T$ new players entering each population each period to replace the $1/T$ player 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 has a common prior, and uses a common deterministic optimal rule $r_i(y_i)$. 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 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 $(y_i, r_i(y_i), z)$

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

¹² Even with independent beliefs this property need not hold when an agent has several actions that lead to the same information set of an opponent.

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 Σ of mixed strategy profiles to itself, and so has a fixed point. These fixed points are the *steady states* of the system.¹³

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 there is a sequence $\bar{\theta}^T \in \bar{\Theta}(g^0, \delta, T)$, $\lim_{T \rightarrow \infty} \bar{\theta}^T \rightarrow \bar{\theta}$ we say that $\bar{\theta}$ is a g^0, δ -*stable state*. If $\bar{\theta}(\delta)$ are g^0, δ -*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]) g^0, δ -*steady states are self-confirming equilibria; patiently stable states are Nash equilibria*.¹⁴

Note that the definitions of stability and patient stability are satisfied 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, and Theorem 5.1 does not assert that there is a single prior for which all of the Nash equilibria are steady states.

¹³ 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 (*).

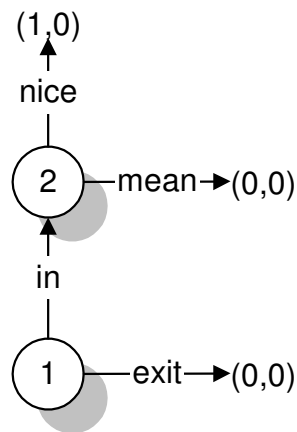
¹⁴ Our 1993 paper states this result for the case where agents know the distribution of Nature's move, but the result extends to the present setting. The key fact is that our argument showed that in patiently stable state, each $\bar{\theta}_i$ must maximize $u_i(s_i, \bar{\theta}_{-i})$, regardless of how $\bar{\theta}_{-i}$ is generated.

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 Nash equilibrium, indeed subgame perfect, for player 1 to play **in**, 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 the 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. A weaker assumption is to assume that no player has two different actions at an information set that can possibly result in a tie in his own payoff. We require also that this assumption hold

with respect to Nature's play. That is, we may convert a game with Nature's moves into a game without Nature's moves by moving all of Nature's moves to the end of the game and then replacing Nature's moves with a terminal node assigning the expected utility generated by Nature. In this game as well, no player should have two different actions at an information set that can possibly result in a tie in his own payoff. 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: the ties are for the suspect, but 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 profile is nearly pure – only Nature randomizes, and only off the equilibrium path.

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.*

The proof is in Section 7 below. 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. Note also that, although this paper only considers the case of independent beliefs, Theorem 6.1 applies immediately to the more general case where correlated beliefs are allowed; difficulties could only arise if priors were restricted to have specific types of correlation.

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.

Theorem 6.2: *In a simple game, a patiently stable state $\bar{\theta}$ is a Nash equilibrium in weakly undominated strategies.*

Proof: Our past work showed that a patiently stable steady state must be a Nash equilibrium, so it remains to show that the steady state must assign probability 0 to weakly dominated strategies. This follows from our maintained assumptions that priors are non-doctrinaire and independent. The optimal rule in the agent's dynamic programming problem 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. Non-doctrinaire priors imply that it will never be a myopic best response to play a weakly dominated strategy, and in a simple game with independent priors, a weakly dominated strategy has no information value. Notice that the theorem makes no assertions about iterated dominance.¹⁵

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 lightning profile is not patiently stable.

In the game without the river, the no-crime profile is Nash, but fails any test of off-path rationality by the accuser, who finds it weakly dominant to **lie**. In the Hammurabi game, the no-crime profile is again a Nash equilibrium, and it also passes the test of off-path rationality, but the beliefs of the accuser about his off-off-path play are incorrect. We will show that the no-crime profile is patiently stable in the Hammurabi game, but that the no-crime profile is not patiently stable in the game without the river.

¹⁵ We are unaware of a counterexample with correlated priors.

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, it is weakly dominant to **lie**, so **(exit, truth)** 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**, believe he will not be punished for telling the **truth**, and believe that if he were to **lie** he would be punished with probability one. 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 Nash equilibrium is path equivalent to a subgame-confirmed Nash equilibrium in which players play pure strategies.*

Example 3.2 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 equilibria that is not path equivalent to a pure subgame-confirmed Nash 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 π , and let the first player be player 1. Each strategy that has positive probability under π is a best response to some beliefs about other player actions in all other subgames. In particular it is a best response to the beliefs 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 π , player $j \neq i$ plays a pure strategy; call these $s_j^*(s_1)$. Thus for each s_1 in the support of π , 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 π is path-equivalent to a public randomization over pure-strategy Nash equilibria.

☑

Proof of Lemma 6.3: Fix a subgame-confirmed Nash 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 Nash 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.

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 (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 λ_0 is the maximum of the difference between $\bar{\theta}$ and $\hat{\pi}$ at any information set on the path of $\hat{\pi}$, and λ_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 λ -metric. 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; the appendix 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 π_{-i} that is within $\bar{\varepsilon}$ of $\hat{\pi}_{-i}$ at every information set. In addition, every player's actions at nodes one step off the path are also a strict best response to some strictly positive beliefs $\hat{\mu}$ that support $\hat{\pi}$ as subgame confirmed. Moreover, there is such a $\hat{\mu}$, and a $\tilde{\varepsilon}$ such that for any beliefs within $\bar{\varepsilon}$ of $\hat{\mu}$ any action that is not an (*ex ante*) best response to $\hat{\pi}$ has expected payoff relative to those beliefs of at least $\tilde{\varepsilon}$ lower than that of the best response.

We say that priors are n, ε -strong 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 ε . Define $\underline{n} \equiv 2^7 \pi^2 / \bar{\varepsilon}^4$. We say that priors are strong if they are $\underline{n}, \bar{\varepsilon}$ -strong at all nodes.

Since we are free to choose any non-doctrinaire priors in order to prove the lemma, we can specify that the priors 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 the actions in $A(x)$ 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 $(\gamma \hat{\mu}_h + n \hat{p}_x) / (\gamma + n)$.

The first lemma shows that beliefs about on-path play “are close to” $\hat{\pi}$. This is useful both in showing that most players in $\bar{f}^T(\bar{\theta})$ 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 Dirichlet and strong, then for all $\bar{\theta}$ such that $\lambda(\bar{\theta} | \hat{\pi}) = \{\varepsilon_0, \varepsilon_1\}$ with $\varepsilon_0 < \bar{\varepsilon} / 2$, and all δ, 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 beliefs about on-path play at any information set at any date n are 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

beliefs will be a convex combination of the prior and the sample, and strong 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 strong priors, we know that there is no sample path of length less than \underline{n} that can make any player k 's posterior beliefs 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. 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 θ , 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 the Appendix implies that¹⁶

$$\Pr(\max_{\underline{n} \leq n \leq \bar{n}} S_{j,n} > \varepsilon) \leq \frac{8\pi^2}{3} \frac{1}{\underline{n}} \frac{\mu_j}{\varepsilon^4}.$$

Substituting $\varepsilon = \beta\mu_j^{1/4}$ we conclude that

$$\Pr(\max_{\underline{n} \leq n \leq \bar{n}} S_{j,n} > \beta\mu_j^{1/4}) \leq \frac{8\pi^2}{3} \frac{1}{\underline{n}\beta^4}.$$

If the play prescribed by $\bar{\theta}$ is within ε_0 of $\hat{\pi}$ at every information set on the path of play, then $\mu_j \leq \varepsilon_0$, and

$$\Pr(\max_{\underline{n} \leq n \leq \bar{n}} S_{j,n} > \beta\varepsilon_0^{1/4}) \leq \frac{8\pi^2}{3} \frac{1}{\underline{n}\beta^4}.$$

Hence taking $\beta = \bar{\varepsilon}/2\varepsilon_0^{1/4}$ we have

$$\Pr(\max_{\underline{n} \leq n \leq \bar{n}} S_{j,n} > \bar{\varepsilon}/2) \leq \frac{2^7 \pi^2 \varepsilon_0}{3\underline{n}\bar{\varepsilon}^4}, \text{ and}$$

¹⁶ Note that \bar{n} on the left does not matter, since it does not appear on the right.

substituting $\underline{n} \equiv 2^7 \pi^2 / \bar{\varepsilon}^4$, 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 an deviation for player i from the path of $\hat{\pi}$, we say that a player's beliefs are a, ε 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

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

generated by the player's actual beliefs about play at off-path nodes, imply a loss of no more than ε from playing a rather than the path of $\hat{\pi}$. Note that in simple games, a player's beliefs about play following some other deviation a' are irrelevant for whether the beliefs are a, ε off-path deviation inducing, as are the player's beliefs 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 Dirichlet and strong. For any $\varepsilon > \tilde{\varepsilon}$ and any state $\bar{\theta}$ with $\varepsilon_1 < \bar{\varepsilon}$, any prior and any δ , as $T \rightarrow \infty$ the fraction of agents in $f^T(\bar{\theta})$ who play a and have beliefs that are a, ε 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, ε off-path deviation-inducing is no larger than ε' . 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 θ_{-i} . Let x be the node where a is feasible. Define $\hat{p}_x(a | y_i)$ to be the frequency 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 ε' , 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 ε' . Since X is finite, for any N and ε' , 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 $\varepsilon'/3$.

Now fix an ε' and the corresponding N, N' , and divide the population of player i 's into two categories: those who have played a more than N' times, and those who have not. Then by the preceding arguments there is 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 ε' . Since priors are strong, these players' beliefs at $x \in X(a, \bar{\theta}_{-i})$ are within $\max\{\varepsilon', \varepsilon_1\} < \bar{\varepsilon}$ 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 $\hat{\mu}$.¹⁷ By definition of $\hat{\mu}$ and $\tilde{\varepsilon}$ it follows that a has an expected loss of at least $\tilde{\varepsilon}$. Since $\varepsilon \geq \tilde{\varepsilon}$ these players' beliefs are not a, ε off-path deviation inducing.

To handle the histories where a has been played fewer than N' times, note that the fraction of the population that plays a and has done so no more than N' times must go to zero as $T \rightarrow \infty$, and so is eventually smaller than $\varepsilon'/3$. So the total fraction of

¹⁷ 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.

players whose beliefs are a, ε off-path deviation-inducing is no more than ε' , and goes to 0 as $T \rightarrow \infty$.

☑

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

Lemma 7.3: *Suppose that all agents have priors that are Dirichlet and strong. For any ε_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 ε_0 .*

Proof: A player who deviates at an on-path node either (i) does not play an ε -static best-response to beliefs, (ii) has beliefs that are a, ε -off-path deviation inducing for some a , or (iii) has beliefs that are 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.¹⁸ 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 ε_1 , there exists an N such that if priors are Dirichlet and $N, 2\varepsilon_1$ -strong at all nodes one step-off the path of $\hat{\pi}$, then for all $\bar{\theta}, \varepsilon_0$ such that $\lambda(\bar{\theta} | \hat{\pi}) = \{\varepsilon_0, \varepsilon_1\}$ and all δ, 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 ε' 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 ε' . Since the prior about this node is concentrated near $\hat{\pi}$, and $\bar{\theta}$ is within ε_1 of $\hat{\pi}$ at this nodes, by choosing ε' 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 $N, 2\varepsilon_1$ -strong

¹⁸ 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].

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 most players one step off the path of play a best response to their priors.

Lemma 7.5: *Let $\varepsilon_1 \leq \bar{\varepsilon}/2$ and let $\hat{\mu}$ be Dirichlet priors that support $\hat{\pi}$ as subgame-confirmed. For any ε_1 there exists N such that if $\hat{\mu}$ is strong and is also $N, 2\varepsilon_1$ -strong one step-off the path of $\hat{\pi}$, then for all δ there is an ε_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. 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 is 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. (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 δ there is an ε_0 such that Lemma 7.5 holds with the fraction failing to play a best-response to their priors no greater than ε_1 . Fix a prior $\hat{\mu}$ that supports $\hat{\pi}$ as subgame confirmed, that is strong (relative to $\underline{n}, \bar{\varepsilon}$) and is also $N, 2\varepsilon_1$ -strong one-step off the path. We will keep this prior fixed as we vary δ, T . Fix δ . Since by Lemma 7.5 the fraction failing to play a best-response to their priors one-step off path is no greater than ε_1 , and $\hat{\mu}$ supports $\hat{\pi}$ as subgame-confirmed, this implies that all but ε_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 ε_0 of $\hat{\pi}$. Since ε_0 can be arbitrarily small, this implies that the limit for each δ as $T \rightarrow \infty$ is path equivalent to $\hat{\pi}$. As this remains true for the limit as $\delta \rightarrow 1$, this completes the proof.

☑

8. 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. When combined with the results of this paper, the conjecture would imply that patient stability is essentially equivalent to subgame-confirmed equilibrium in the games we studied here. 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.

To conclude it may be helpful to relate our analysis to past work. 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 Nash equilibria, 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?

Most work in this area has followed Fudenberg and Kreps [1988], [1995], [1996] in treating the frequency and timing of off-path “experiments” as exogenous. 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. This is similar in spirit to the work of Jehiel and Samet [2004] on the convergence of boundedly rational learning in games of perfect information, as they too assume that there is an exogenous probability of experimentation. In Noldeke and Samuelson [1993], 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. Other explanations for experimentation will lead to other equilibrium refinements; the advantage of the optimization assumption is that it provides some guidance on the amount of off-path experimentation that we might expect to see.

Appendix

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

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

Lemma A.1¹⁹: $\Pr(\max_{\underline{n} \leq n \leq \bar{n}} S_n > \varepsilon) \leq \frac{8\pi^2}{3} \frac{1}{\underline{n}} \frac{\mu}{\varepsilon^4}.$

Proof: We derive specific bounds based on the method of proof of the strong law of large numbers given by Billingsley [1986]. By Markov's inequality,

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

so

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

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 $\zeta(2) = \pi^2/6$ where ζ is the Riemann zeta function. For $\underline{n} > 1$ we have the bound

$$\sum_{n=\underline{n}}^{\infty} \frac{1}{n^2} \leq \int_{\underline{n}-1}^{\infty} \frac{1}{n^2} dn = \frac{1}{\underline{n}-1} \leq \frac{2}{\underline{n}} \leq \frac{2\pi^2}{6} \frac{1}{\underline{n}},$$

which gives the desired result.

¹⁹ The lemma is stated for the case of binomial random variables, where its strength is proportional to the mean μ , 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.

☑

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 marginal 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 $T, r, \bar{\theta}, i, a \in A(x), s_i, x \in X(s_i), x \in X_j, j \neq i$*

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

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

References: Fudenberg and Levine [1993b] Lemma B.2 and/or Fudenberg and Levine [2004] Lemma C.3.

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.

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 and/or Fudenberg and Levine [2004] Lemma C.4.

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 ε -best-response to the marginal beliefs 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 and Fudenberg and Levine [2004] Lemma D.1. The intuition for this result is that if node x has been reached

many times, the “option value” of experimenting here is likely to be low, so that with high probability the optimal rule must prescribe an ε -best response.²⁰ 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 ε -best-response to their beliefs.

Lemma A.5 : *With independent priors,*

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

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 $p_t = p(x | y_t)$ where y_t means that x was not reached during the previous t periods. Then

$$0 \leq -(1 - \delta)\Delta + \delta p_0 U + (1 - p_0)p_1 \delta^2 U + (1 - p_0)(1 - p_1)\delta^3 U \dots$$

Notice that p_t is non-increasing in t : failing to reach x must lower the posterior probability that it will be reached in the future. So in particular

$$0 \leq -(1 - \delta)\Delta + \delta p_0 U / (1 - \delta).$$

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

☑

²⁰ One might expect that we could take ε' to be 0 by taking N sufficiently large, and indeed this is possible in standard bandit problems. 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.

References

- Aumann, R. [1987] "Correlated Equilibrium as an Expression of Bayesian Rationality," *Econometrica* 55,1-18.
- Billingsley, P. [1986]: *Probability and Measure* 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.
- 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 [2004] "Rational Steady-State Learning and Subgame-Confirmed Equilibrium," in preparation.
- 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.

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.