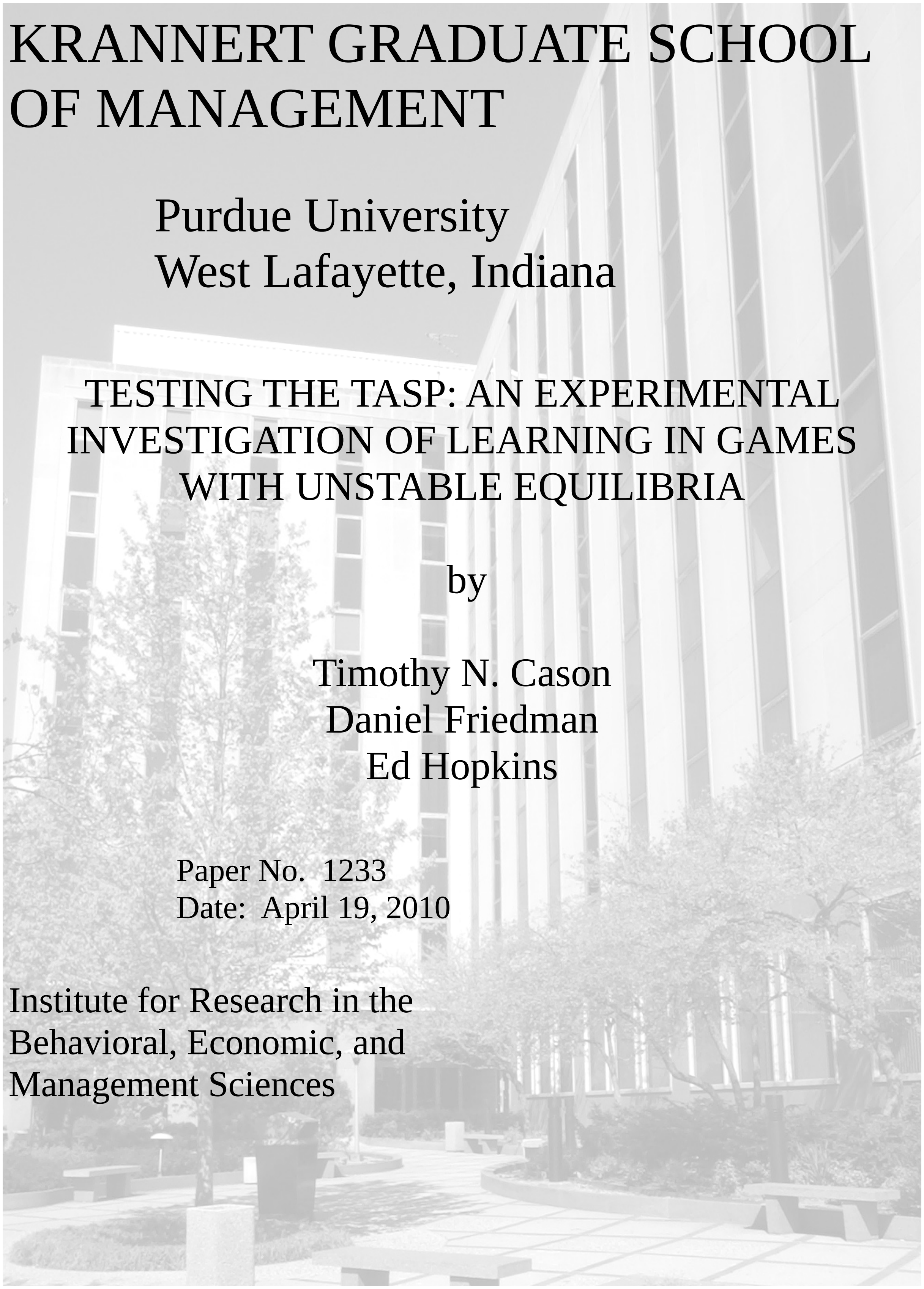


KRANNERT GRADUATE SCHOOL
OF MANAGEMENT

Purdue University
West Lafayette, Indiana

TESTING THE TASP: AN EXPERIMENTAL INVESTIGATION OF LEARNING IN GAMES WITH UNSTABLE EQUILIBRIA

by

Timothy N. Cason
Daniel Friedman
Ed Hopkins

Paper No. 1233
Date: April 19, 2010

Institute for Research in the
Behavioral, Economic, and
Management Sciences

Testing the TASP: An Experimental Investigation of Learning in Games with Unstable Equilibria[1]

Timothy N. Cason
Purdue University

Daniel Friedman
UC Santa Cruz

Ed Hopkins
University of Edinburgh

April 19, 2010

Abstract

We report experiments designed to test between Nash equilibria that are stable and unstable under learning. The "TASP" (Time Average of the Shapley Polygon) gives a precise prediction about what happens when there is divergence from equilibrium under a wide class of learning processes. We study two versions of Rock-Paper-Scissors with the addition of a fourth strategy, Dumb. The unique Nash equilibrium places a weight of 1/2 on Dumb in both games, but in one game the NE is stable, while in the other game the NE is unstable and the TASP places zero weight on Dumb. Consistent with TASP, we find that the frequency of Dumb is lower and play is further from Nash in the high payoff unstable treatment than in the other treatments. However, the frequency of Dumb is substantially greater than zero in all treatments.

JEL numbers: C72, C73, C92, D83

Keywords: games, experiments, TASP, learning, unstable, mixed equilibrium, fictitious play.

[1] cason@purdue.edu, http://www.krannert.purdue.edu/faculty/cason/; dan@ucsc.edu, http://leeps.ucsc.edu/people/Friedman/; E.Hopkins@ed.ac.uk, http://homepages.ed.ac.uk/hopkinse/. We thank Purdue's Krannert School of Management for funds used in the experiment; Steve Gjerstad, Nat Wilcox, and an anonymous associate editor and two referees for valuable comments and suggestions; and Roman Sheremeta for excellent research assistance. We received useful feedback from audiences at Amsterdam, Florida State, Purdue and Virginia Commonwealth Universities, and at Economic Science Association, Game Theory Society and Society for the Advancement of Economic Theory conferences.

1. Introduction

Economic models often only have equilibria in mixed strategies, but it is difficult to see how actual participants know how to randomize with the correct probabilities. Recent theoretical advances, in particular the development of stochastic fictitious play, demonstrate that in many games a mixed equilibrium can be achieved by agents who follow simple learning rules. Unfortunately, in other games the equilibria are not learnable—players following any one of a range of learning processes will not converge to equilibrium. However, Benaïm, Hofbauer and Hopkins (2009) show that, nonetheless, stochastic fictitious play can give a point prediction for play even when it diverges. This point is the TASP (Time Average of the Shapley Polygon) which they show can be quite distinct from any Nash equilibrium.

In this paper, we report experiments designed to test between Nash equilibria that are stable and unstable under learning. Subjects were randomly matched to play one of two 4 x 4 games each with a unique mixed Nash equilibrium. In one game, the equilibrium is predicted to be stable under learning, and in the other unstable. Both games are versions of Rock-Paper-Scissors with the addition of a fourth strategy, Dumb. The mixed equilibrium in both games is (1, 1, 1, 3)/6: Dumb is thus the most frequent strategy. In the unstable game, however, fictitious play-like learning processes are predicted to diverge from the equilibrium to a cycle, a "Shapley polygon," that places no weight upon Dumb. Thus, if fictitious play describes agents' behavior, the limiting frequency of Dumb is a ready indicator of whether we are in the stable or unstable case. It is also, therefore, a simple way to determine whether the predictions of fictitious play, and learning theory, hold in practice. Equilibrium theory suggests that the frequency of Dumb should be the same in both games. Learning theory suggests they should be quite different.

The experiment has a 2 x 2 design with four treatment conditions: unstable or stable game and high or low payoff. We find that there is a difference in play in the high payoff unstable game treatment. The frequency of Dumb is lower and play is further from Nash than in the other treatments, though the frequency of Dumb is always substantially greater than zero. That is, we find support for the idea that the stability or instability of an equilibrium under stochastic fictitious play can influence subject behavior. The data also reject Nash equilibrium, which predicts no difference between the treatments. The predictions of quantal response equilibrium fare better than Nash but its prediction that play should not vary between the stable and unstable games is not supported by the data.

Fictitious play has the underlying principle that players select a best response to their beliefs about opponents. Traditionally, these beliefs are constructed from the average past play of opponents. This we refer to as players having "classical" beliefs. However, experimental work has found greater success with generalizations of fictitious play that allow for players constructing beliefs by placing greater weight on more recent events (see Cheung and Friedman (1997), Camerer and Ho (1999) amongst many others). This is called forgetting or recency or weighted fictitious play.

Benaïm, Hofbauer and Hopkins (2009) examine weighted fictitious play in "monocyclic" games, a class of games that generalizes Rock–Paper–Scissors and that has only mixed equilibria. They prove that when learning diverges from the equilibrium, the time average of play converges to the TASP, a new concept. In the unstable game we consider, the TASP is quite distinct from the unique Nash equilibrium. Thus, an easy test of divergence is simply to see whether average play is closer to the TASP or the Nash equilibrium.

In practice, one cannot expect play to be exactly at either the Nash equilibrium or the TASP. The now extensive literature on perturbed equilibria such as quantal response equilibrium (QRE, e.g., McKelvey and Palfrey, 1995) makes clear that play in experiments can be quite distinct from Nash equilibrium. Subjects appear to behave as though their choices were subject to noise. Equally, since the stationary points of stochastic fictitious play are QRE, learning theory can make similar predictions. Thus we should expect learning to converge exactly to the TASP only in the absence of noise.

Stochastic fictitious play and standard QRE models both predict that the noise amplitude should decrease in the level of the payoffs. This effect has been found empirically by Battalio et al. (2001) and Bassi et al. (2006), although it has been challenged recently by a modified formulation advocated in Wilcox (2010). Thus, the other aspect of our design is to change the level of monetary rewards. We ran both the stable and unstable game at two different conversion rates between experimental francs and U.S. dollars, with the high conversion rate two and a half times higher than the lower.

Learning theory predicts that this change in monetary compensation will have a different comparative static effect in the two different games. Higher payoffs should make play diverge further from the equilibrium in the unstable game and make play closer to equilibrium in the stable one. By contrast, the standard QRE model predicts play should be closer to Nash equilibrium when payoffs are higher, in both the stable and unstable games. Nash equilibrium predicts no difference across the treatments. Thus we have clear and distinct comparative statics predictions to test.

In contrast, in the previous empirical and theoretical literature on mixed strategy equilibria, there has been a focus on the time series property of play. For example, Foster and Young (2003) make the distinction between convergence in *time average* and convergence in *behavior*. The first requires the overall frequencies of play to approach the mixed strategy equilibrium frequencies. The second requires the more demanding standard that players should actually come to randomize with equilibrium probabilities. To illustrate, the sequence 0,1,0,1,0,1,... converges in time average to 1/2 but clearly not to the behavior of randomizing between 0 and 1 with equal probabilities.

In the experimental literature, this distinction was first raised by Brown and Rosenthal (1990). Their analysis of the earlier experiments of O'Neill (1987) finds that while play converged in time average, it failed to do so in behavior, in that there was significant autocorrelation in play. Subsequent experiments on games with mixed strategy equilibria seem to confirm this finding. For example, Brown Kruse et al. (1994), Cason and Friedman (2003) and Cason, Friedman and Wagener (2005) find in oligopoly experiments that the average frequencies of prices approximate Nash equilibrium frequencies. However, there are persistent cycles in the prices charged, which seems to reject convergence to equilibrium in behavior.[2]

However, if play is not i.i.d. over the finite length of an experiment, is this because play is diverging, because convergence will never be better than approximate, or because convergence is coming but has not yet arrived? We avoid this problem by not measuring convergence in terms of the time series properties of play. Rather, the advantage of the game we consider is that a considerable qualitative difference in behavior is predicted between its stable and unstable versions.

Other experimental studies have tested for differences in behavior around stable and unstable mixed equilibria. Tang (2001) and Engle-Warnick and Hopkins (2006) look at stable and unstable 3 x 3 games in random matching and constant pairing set-ups respectively. Neither

[2] Exceptions appear in the literature for professional tennis players (Walker and Wooders, 2001) and soccer players (Palacios-Huerta, 2003). Palacios-Huerta and Volij (2008) find that professional sportsmen can learn equilibrium behavior in the laboratory. However, Levitt et al. (2010) report additional experiments in which professionals do no better than students.

study finds strong differences between stable and unstable games. In a quite different context, Anderson et al. (2004) find that prices diverge from competitive equilibrium that is predicted to be unstable by the theory of tatonnement.

2. RPSD Games and Theoretical Predictions

The games that were used in the experiments are, firstly, a game we call the unstable RPSD game

$RPSD_U =$

	R	P	S	D
Rock	90, 90	0, 120	120, 0	20, 90
Paper	120, 0	90, 90	0, 120	20, 90
Scissors	0, 120	120, 0	90, 90	20, 90
Dumb	90, 20	90, 20	90, 20	0, 0

and secondly, the stable RPSD game,

$RPSD_S =$

	R	P	S	D
Rock	60, 60	0, 150	150, 0	20, 90
Paper	150, 0	60, 60	0, 150	20, 90
Scissors	0, 150	150, 0	60, 60	20, 90
Dumb	90, 20	90, 20	90, 20	0, 0

Both games are constructed from the well-known Rock-Paper-Scissors game with the addition of a fourth strategy called Dumb, which is never a best response to a pure strategy. Games of this type were first introduced by Dekel and Scotchmer (1992). Both these games have the same unique Nash equilibrium which is symmetric and mixed with frequencies $p^* = (1,1,1,3)/6$. Ironically, "Dumb" is by far the most frequent strategy in equilibrium. Expected equilibrium payoffs are 45 in both games.

While these two games are apparently similar, they differ radically in terms of predicted learning behavior. To summarize our basic argument, suppose there is a population of players who are repeatedly randomly matched to play one of the two games. Then, if all use a fictitious play like learning process to update their play, in the second game there would be convergence to the Nash equilibrium. In the first game, however, there will be divergence from equilibrium and play will approach a cycle in which no weight is placed on the strategy Dumb (D).

2.1 Learning Under Fictitious Play

We state and prove results on the stability of the mixed equilibria in $RPSD_U$ and $RPSD_S$ in Appendix A, available online. Here we give a heuristic account.

Figure 1 shows the simplex of possible mixed strategies over the four available actions. The dashed triangle at the base of the simplex is the attracting cycle for the $RPSD_U$ game under fictitious play.[3] This cycle was named a Shapley triangle or polygon after the work of Shapley (1964) who was the first to produce an example of non-convergence of learning in games. See also Gaunersdorfer and Hofbauer (1995) for a detailed treatment.

More recently, Benaïm, Hofbauer and Hopkins (2009) consider "weighted" fictitious play. In classical fictitious play, beliefs are constructed by taking a simple average over all observations. In contrast, under the assumption of weighting, players construct their beliefs about the play of others by placing greater weight on more recent experience, leading to what is sometimes called constant gain learning. Then, play in the unstable game will still converge to the Shapley triangle, but the time average of play will converge to a point that they name the TASP (Time Average of the Shapley Polygon), denoted "T" on Figure 1. This is in contrast to Shapley's classical result, where in the unstable case nothing converges. For the game $RPSD_U$, the TASP places no weight on the strategy D, despite its weight of 1/2 in Nash equilibrium. That is, it is clearly distinct from the Nash equilibrium of the game, denoted "N" in Figure 1.

However, it is not the case that theory predicts that the frequency of *D* should decrease monotonically. Specifically, Proposition 2 in Appendix A identifies a region *E* in the space of mixed strategies where *D* is the best response and so its frequency will grow. This region *E* is a pyramid within the pyramid in Figure 1, with the Shapley triangle as its base and apex at the Nash equilibrium. But under fictitious play, given almost all initial conditions, play will exit *E* and the frequency of *D* will diminish.

In the second game $RPSD_S$, by contrast, the mixed equilibrium is stable under most forms of learning, including fictitious play. Hence, one would expect to see the average frequency of the fourth strategy, *D*, to be close to one half.

[3] Fictitious play is perhaps the most enduring model of learning in games. See Fudenberg and Levine (1998, Chapter 2) for an introduction. Here, we consider it in the context of a single random matching population. This is technically convenient. Further, random matching provides a motivation for the applicability of simple adaptive learning processes such as fictitious play. There are, of course, other learning models. Young (2004) gives a survey of recent developments.

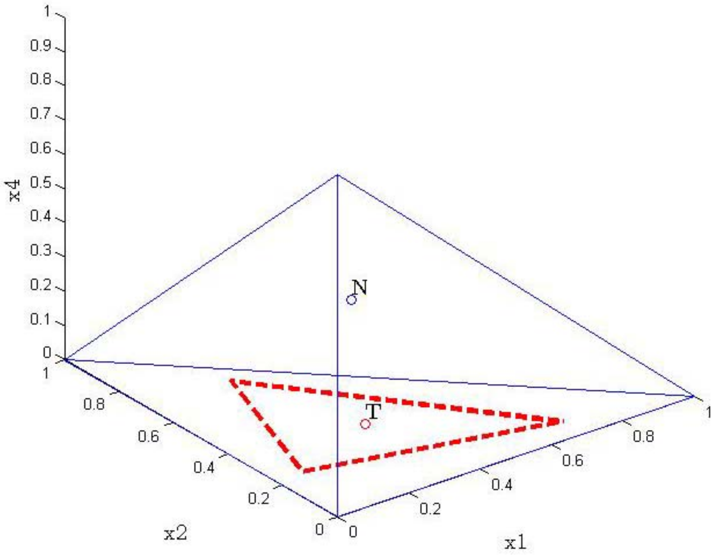


Figure 1: Nash equilibrium (N) and TASP (T) in the unstable version of the *RPSD* game. The frequencies of strategies 1 and 2 are on the horizontal axes and of strategy 4 on the vertical axis.

Thus, if fictitious play describes agents' behavior, the limiting frequency of *D* is a ready indicator of whether we are in the stable or unstable case. It is also, therefore, a simple way to determine whether the predictions of fictitious play, and learning theory, hold in practice. Equilibrium theory suggests that the frequency of *D* should be the same in both games. Learning theory suggests they should differ.

2.2 Noisy Play: SFP and QRE

This clean distinction ignores the fact that actual behavior is often noisy, as subjects often make mistakes or experiment. This behavior can be captured theoretically in two linked ways. The first is stochastic fictitious play (SFP), a modification of the learning model that allows for random choice. The second is perturbed equilibria known as quantal response equilibria (QRE). The link is that QRE are the fixed or stationary points for the SFP learning process.

The standard choice rule in SFP is logit, where the probability of player i taking action j from a menu of n possibilities at time t is given by

$$p_i^j(t) = \psi^j(A_i(t)) = \frac{e^{\lambda A_i^j(t)}}{\sum_{k=1}^n e^{\lambda A_i^k(t)}}.$$

Here $\lambda \geq 0$ is a precision parameter and in SFP the "attraction" $A_i^j(t)$ is the expected payoff to action j at time t. As λ becomes large, the probability of choosing the action with the highest expected payoff, the best response, goes to one.

In fictitious play, expectations about payoffs are derived from expectations over opponents' actions which in turn are derived from past observations of play. The now-standard approach here is the EWA (experience weighted attraction) model of Camerer and Ho (1999), which includes as special cases SFP, both the weighted form and with classical beliefs, and several other learning models. Attractions in the EWA model are set by

$$A_i^j(t) = \frac{\phi N(t-1) A_i^j(t-1) + [\delta + (1-\delta) I_j(t-1)] \pi_j(t-1)}{N(t)}$$

where $N(t) = \rho N(t-1) + 1$, and ϕ and ρ are recency parameters. For classical beliefs, $\rho = \phi = 1$; for weighted beliefs $\rho = \phi$<1. The parameter δ is an imagination factor and for all forms of fictitious play, it is set to 1, and $I_j(t-1)$ is an indicator function that is one if action j is chosen at $t-1$ and is zero otherwise.[4] Finally, π^j is the (implied) payoff to strategy j. In this context, we deal with simple strategic form games, so that given a game matrix B, we will have the payoff to strategy j being $\pi^j = B_{jk}$ given that the opponent chose action k.

Equilibrium in SFP occurs when the expected payoffs are consistent with players' actual choice frequencies. This idea for a perturbed equilibrium was proposed independently by McKelvey and Palfrey (1995) under the title of QRE. Given the logit choice rule, one finds the QRE equilibrium frequencies by solving the following system of equations

$$p^j = \psi^j(\pi(p)) = \frac{e^{\lambda \pi^j(p)}}{\sum_{k=1}^n e^{\lambda \pi^k(p)}} \text{ for } j = 1, \ldots, n$$

where $\pi(p) = B.p$ with B being the payoff matrix. QRE with the specific logit choice rule can also be called logit equilibrium.

SFP typically is analyzed using perturbed best response dynamics

$$\dot{x} = \psi(\pi(x)) - x, \qquad \textbf{(PBR)}$$

[4] The EWA model permits $\rho \neq \phi$ and/or $\delta < 1$ to capture a variety of reinforcement learning approaches.

where the function $\psi(\cdot)$ is a perturbed choice function such as the logit above and x is a vector of players' beliefs. Results from stochastic approximation theory show that a perturbed equilibrium is locally stable under SFP if it is stable under the perturbed dynamics. See Benaïm and Hirsch (1999), Hopkins (1999) and Ellison and Fudenberg (2000) for details.

One well-known property of QRE is that as the precision parameter λ increases in value, the set of QRE approaches the set of Nash equilibria. But notice that given the logit formulation above, an increase in λ is entirely equivalent to an increase in payoffs. For example, doubling payoffs would have the same effect as doubling λ.

Specific results for the logit version of the QRE in the games $RPSD_U$ and $RPSD_S$ are the following:

1) Each QRE is of the form $\hat{p} = (m, m, m, k)$ where $k = 1 - 3m$ and is unique for a given value of λ. That is, each QRE is symmetric in the first three strategies.
2) The value of k, the weight placed on the fourth strategy D, is in $[1/4, 1/2)$ and is strictly increasing in λ. That is, the QRE is always between the Nash equilibrium $(1,1,1,3)/6$ and uniform mixing $(1,1,1,1)/4$ and approaches the Nash equilibrium as λ or payoffs become large.
3) For a given value of λ, the QRE of $RPSD_U$ and of $RPSD_S$ are identical. That is, while the level of optimization affects the QRE, the stability of the equilibrium does not.

The implications of an increase of the precision parameter λ, or equivalently of an increase in payoffs, for learning outcomes are quite different. First, it is well known that the stability of mixed equilibria under the perturbed best response (PBR) dynamics depend upon the level of λ. When λ is very low, agents randomize almost uniformly independently of the payoff structure and a perturbed equilibrium close to the center of the simplex will be a global attractor. This means that even in the unstable game $RPSD_U$, the mixed equilibrium will only be unstable under SFP if λ is sufficiently large. For the specific game $RPSD_U$, it can be calculated that the critical value of λ is approximately 0.17. In contrast, in the stable game $RPSD_S$, the mixed equilibrium will be stable independent of the value of λ.

This is illustrated in Figure 2. The smooth dashed curve, labeled "Stable," gives the asymptotic level of the proportion of the fourth strategy D for game $RPSD_S$ under the perturbed best response (PBR) dynamics as a function of λ. The smooth blue curve, labeled "Unstable," gives the asymptotic level of D for game $RPSD_U$. For low values of λ, that is on the interval [0, 0.17], the perturbed best response dynamics converge to the QRE in both games. Indeed, in the stable case, the dynamics always converge to the QRE and this is why the "Stable" curve thus also gives the proportion of D in the QRE as a function of the precision parameter λ.

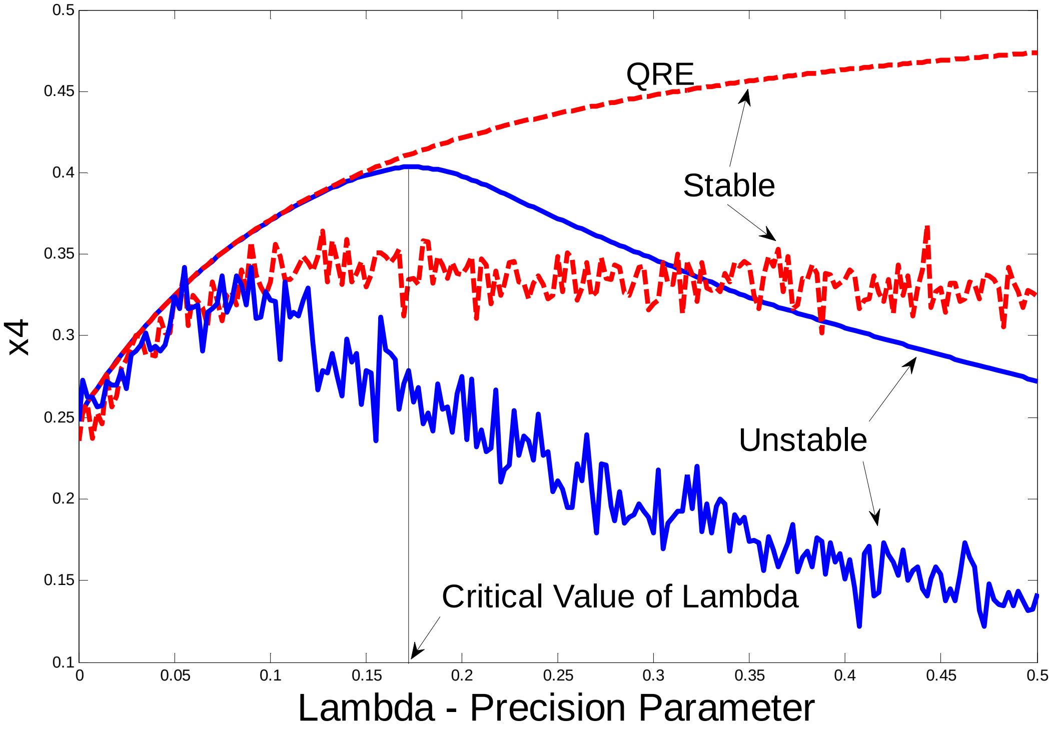


Figure 2: Frequencies of the 4th strategy D against the precision parameter λ. The smooth lines are generated by continuous time learning processes, the jagged lines by simulations of the experimental environment. Stable (dashed) refers to the *RPSD*$_S$ game, and Unstable (solid) to the *RPSD*$_U$ game. The smooth dashed line coincides with QRE for both games.

However, the behavior in $RPSD_U$ is quite different for values of λ above the critical value of 0.17. The logit form of the perturbed equilibrium is unstable above that threshold and, from almost all initial conditions, play converges to a cycle. Of course, with finite λ, the cycle is in the interior of the simplex, and only as λ increases does the proportion of D approach zero.

This leads to the very different comparative static predictions across games. An increase in λ (or payoffs) for game $RPSD_S$ leads to an increase in the frequency of D. By contrast, for λ greater than 0.17, an increase in λ (or payoffs) leads to a decrease in D. QRE predicts that an increase λ (or payoffs) lead to an increase in D in both games.

The theoretical framework assumes an infinite population of agents all who share the same beliefs, and investigates behavior in the limit as time goes to infinity and, the recency parameter ρ goes to one. In the experiments we must work with a finite population and a finite time horizon. Therefore, we also report simulations of populations of twelve agents who play 80 repetitions (both values chosen to be the same as the experiments we run). Each simulated agent learns according to weighted SFP (that is, EWA with $\phi = \rho < 1$ and $\delta = 1$). We set the recency parameter $\phi = \rho = 0.8$ and then vary the precision parameter λ. We ran one simulation for each value of λ in the sequence 0, 0.0025, 0.005, ..., 0.5 for each of the two games $RPSD_U$ and $RPSD_S$. Initial conditions in each simulation were set by taking the initial attractions to be drawn from a uniform distribution on [0, 150]. The resulting average levels of the frequency of D over the whole 80 periods and 12 simulated subjects are graphed as jagged lines in Figure 2. As learning outcomes over a finite horizon are stochastic, there is considerable variation from one simulation to the next even though the value of λ changes slowly. What is encouraging, however, is that the simulations preserve the same qualitative outcomes as the asymptotic results generated by the theoretical models.[5]

2.3 Testable Hypotheses

The experiment employed the $RPSD_U$ matrix in unstable treatments and $RPSD_S$ in stable treatments, and used a conversion rate of payoffs to cash that was 2.5 times higher in high payoff treatments than in low payoff treatments. As noted above, in standard models the payoff treatment has the same effect as an increase in λ. Empirically, as reported in Battalio et al. (2001) and elsewhere, the effect is in the right direction but is less than one-for-one.

The arguments presented above lead to the following alternative predictions.

[5] Clearly, however, they are not identical. In particular the divergence in behavior between stable and unstable cases occurs at a lower level of λ in the simulations. Further simulations, not reported here, indicate that this difference cannot be ascribed to any one of the three factors (finite horizon, finite population, ρ<1) but rather arises from a combination of the three.

1) **Nash Equilibrium (NE):** average play should be at the NE $(1,1,1,3)/6$ in all treatments.

2) **Quantal Response Equilibrium (QRE):**
   a) Average play should be between NE and $(1,1,1,1)/4$ in all treatments, with the first three strategies in equal proportions.
   b) Average play should be the same in stable as in unstable treatments.
   c) Average play should be closer to Nash equilibrium, and the proportion of D should be higher, in high payoff treatments than in low payoff treatments.

3) **TASP:**
   a) Average play should be closer to the TASP in unstable treatments, but closer to QRE in stable treatments
   b) Average play should be closer to the TASP (smaller proportion D) in the high payoff unstable treatment than in the low payoff unstable treatment, but play should be closer to Nash equilibrium (higher proportion D) in the high payoff stable treatment than in the low payoff stable treatment.
   c) Average play should converge in all treatments, but in the unstable treatments beliefs should continue to cycle.

3. Experimental Design and Procedures

The experiment featured a full factorial two-by-two design. One treatment variable was the game payoff matrix, either the unstable game *RPSD*$_U$ or the stable game *RPSD*$_S$ shown earlier. The other treatment variable was the payoff conversion rate of Experimental Francs (EF, the entries in the game matrix) to U.S. Dollars. In the High Payoffs treatment, 100 EF = \$5. In the Low Payoffs treatment, 100 EF = \$2. Subjects also received an extra, fixed "participation" payment of \$10 in the Low Payoffs treatment to ensure that their total earnings comfortably exceeded their opportunity cost.

Each period each player *i* entered her choice s_i^j = 1, 2, 3, or 4 (for Rock, Paper, Scissors, Dumb), and at the same time entered her beliefs about the opponent's choice in the form of a probability vector (p_1, p_2, p_3, p_4). When all players were done, the computer matched the players randomly into pairs and announced the payoffs in two parts. The game payoff was obtained from the matrix, and so ranged from 0 to 120 or 150 EF. The prediction payoff was $5-5\sum_{i=1}^{4} p_i^2 + 10 p_j$ when the opponent's actual choice was j, and so it ranged from 0 to 10 EF.

The payoff scheme was chosen because belief data allow diagnostic tests of the competing models, and because belief elicitation itself can help focus players on belief learning (Ruström and Wilcox, 2009). The quadratic scoring rule was calibrated so that the prediction payments were an order of magnitude smaller than the game payoffs, reducing the incentive to hedge action choices by biasing reported beliefs.[6]

In each session, 12 subjects were randomly and anonymously re-matched over a computer network for a known number of 80 periods to play the same game, either $RPSD_U$ or $RPSD_S$.[7] After each period, subjects learned the action chosen by their opponent, their own payoffs, as well as the frequency distribution of actions chosen by all 12 subjects in the session. At the conclusion of the session, 10 of the 80 periods were drawn randomly without replacement for actual cash payment using dice rolls (to control for wealth effects). Subsection 4.4 summarizes six additional 160-period sessions conducted as a robustness check for our main conclusions.

We conducted three sessions in each of the four treatment conditions, for a total of 144 subjects, plus 72 additional subjects in two treatment conditions for the longer, 160-period horizon. For the main experiment two sessions in each treatment were conducted at Purdue University, and one session in each treatment was conducted at UC-Santa Cruz. All subject interaction was computerized using z-Tree (Fischbacher, 2007). The instructions used neutral terminology, such as "the person you are paired with" rather than "opponent" or "partner." Action choices were labeled as A, B, C and D, and the instructions and decision screens never mentioned the words "game" or "play." The instructions in Appendix B, available online, provide additional details of the framing, and also show the decision and reporting screens.

[6] This potential for biased beliefs does not appear to be empirically significant in practice, at least as measured for other games (Offerman et al., 1996; Sonnemans and Offerman, 2001). Taken by itself, the quadratic scoring rule is incentive compatible (Savage, 1971), and is commonly used in experiments with matrix games (e.g., Nyarko and Schotter, 2002).

[7] Some experiments studying learning and stability in games have used a longer 100 or 150 period horizon (e.g., Tang, 2001; Engle-Warnick and Hopkins, 2006). We used the shorter 80-period length because subjects needed to input beliefs and this lengthened the time to complete each period. Including instructions and payment time, each session lasted about two hours. One of the 12 sessions was unexpectedly shortened to 70 periods due to a move by nature: a tornado warning that required an evacuation of the experimental laboratory.

4. Experiment Results

We begin with a brief summary of the overall results before turning to more detailed analysis. Figures 3 and 4 show the proportion of action choices in each 10-period interval for two of the 12 main sessions. Figure 3 displays a session with the unstable matrix and high payoffs. Paper and Scissors are initially the most common actions. Scissors appears to rise following the early frequent play of Paper, followed by a rise in the frequency of Rock. This pattern is consistent with simple best response dynamics. Dumb is played less than a quarter of the time until the second half of the session and its rate tends to rise over time. Figure 4 displays a session with the stable matrix and high payoffs. The Paper, Scissors and Rock rates again fluctuate, also in the direction expected by best response behavior. For the stable matrix, Dumb starts at a higher rate and rises closer to the Nash equilibrium prediction of 0.5 by the end of the session.

Figure 3: Choice Proportions in Session 5 (High Payoffs, Unstable Matrix)

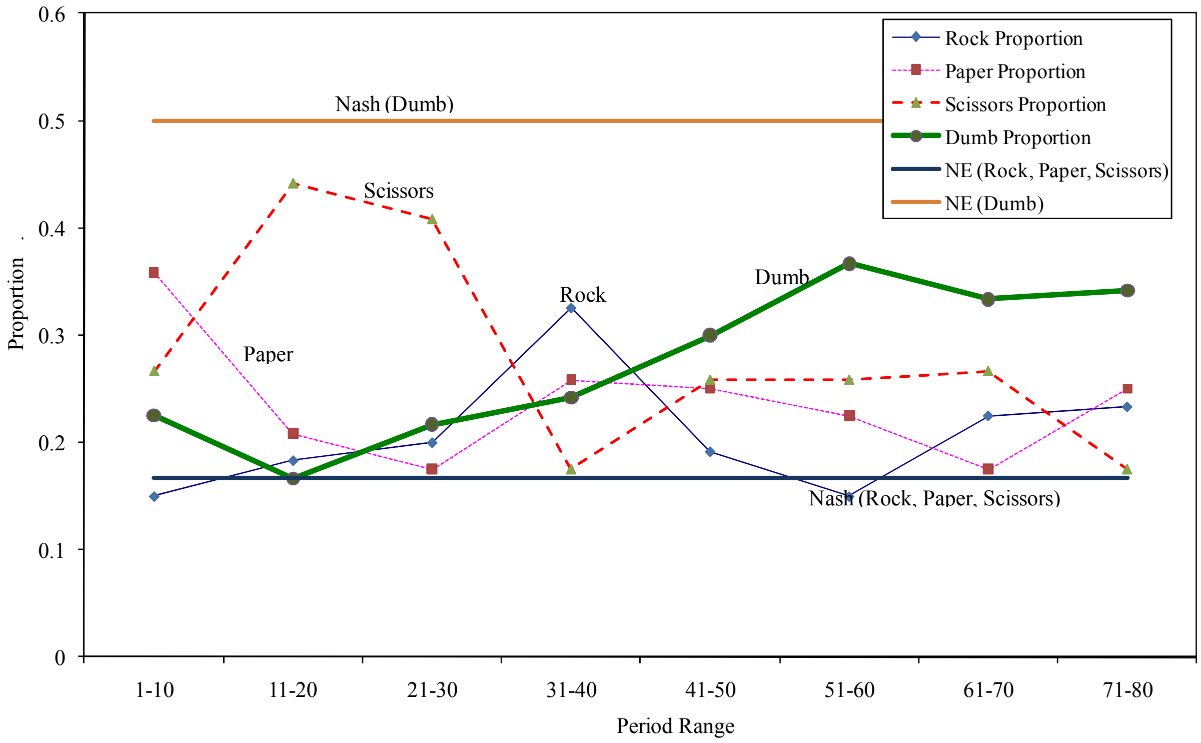


Figure 5 and Table 1 provide a pooled summary for all 12 sessions. The figure displays the frequency that subjects play the distinguishing Dumb action in each of the four treatments. This rate tends to rise over time, but is always below the Nash equilibrium frequency of 0.5. A simple reading of this would be as evidence for QRE over Nash. However, the frequency of

Figure 4: Choice Proportions in Session 11 (High Payoffs, Stable Matrix)

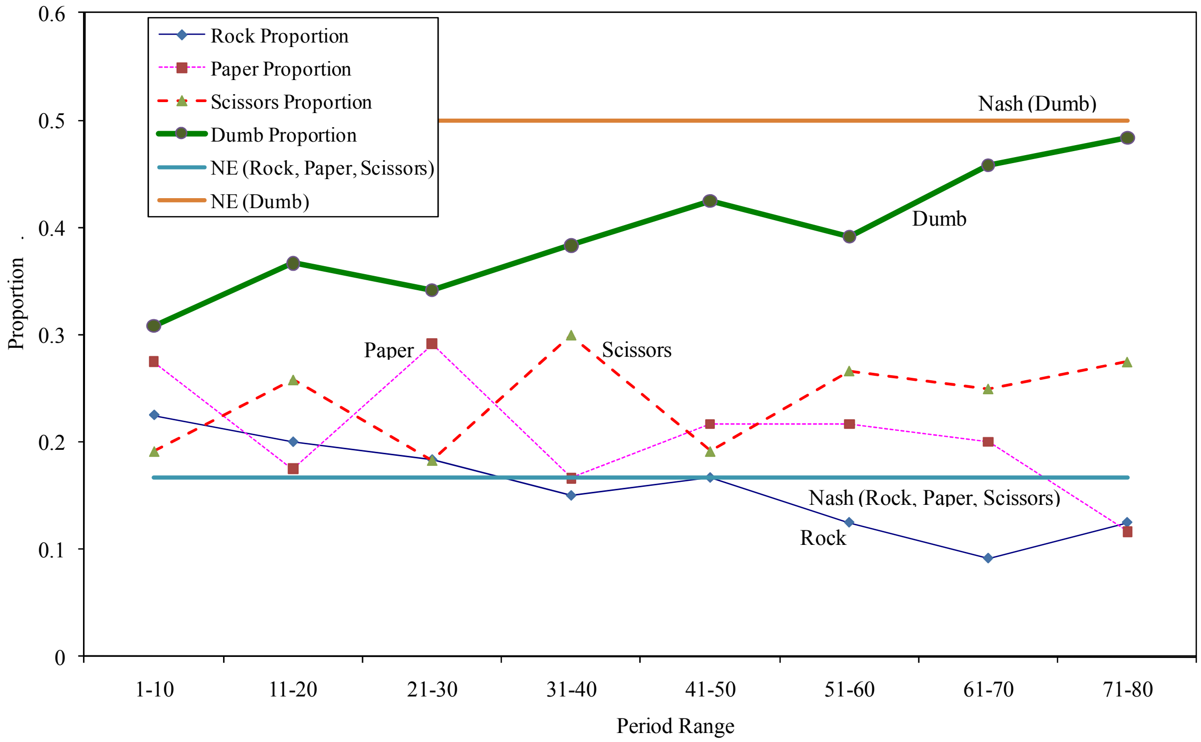
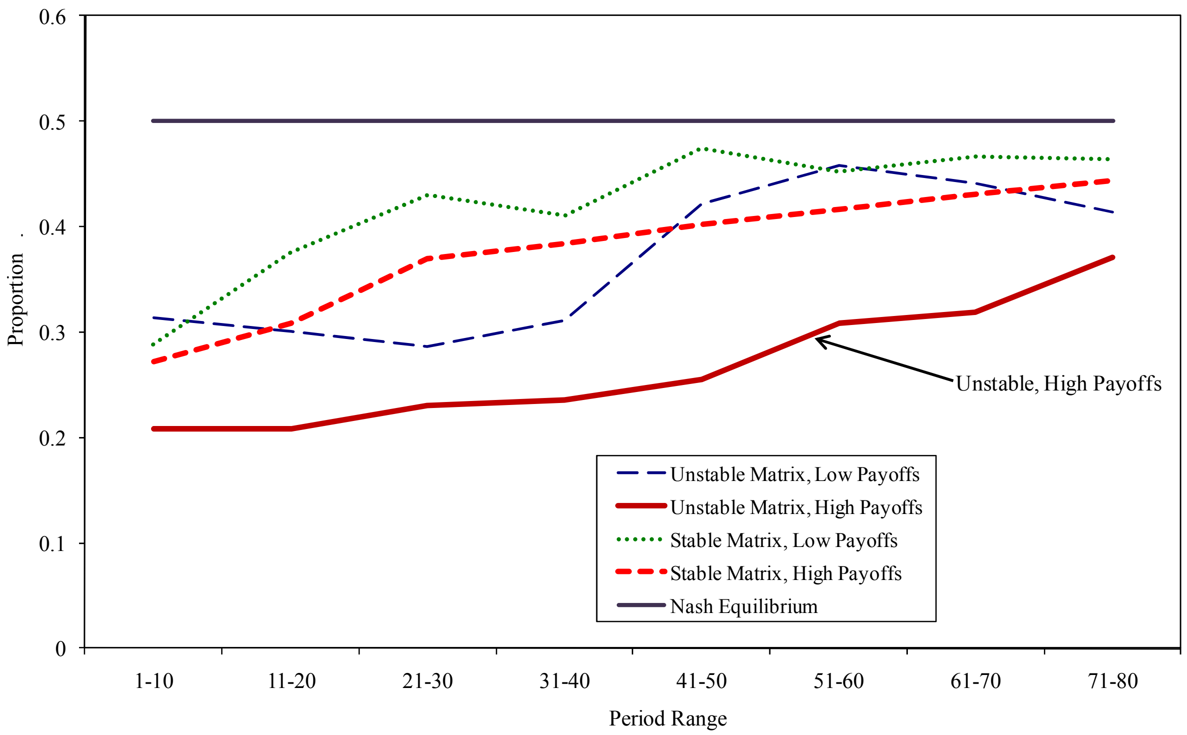


Figure 5: Proportion Choosing Action Dumb in each 10-Period Interval for All Treatments



Dumb is clearly lowest in the unstable, high payoffs condition as predicted by the TASP model. Table 1 shows that Dumb is played about 26 percent of the time overall in this treatment, compared to about 40 percent in the stable matrix treatments. Varying payoffs seems to make little difference in the stable game, which goes against the prediction of both learning and QRE that the frequencies of Dumb should be greater in the high payoff, stable treatment than in the low, stable treatment.

4.1 Tests of the Principal Hypotheses

Figure 5 indicates an upward time trend in all treatments for the rate that subjects choose the critical action Dumb. The Dumb action is played more frequently over time even in the unstable game *$RPSD_U$*. Although the sessions ran only 80 or 160 periods, they permit us to draw statistical inferences about long-run, asymptotic play.

Table 1: Theoretical Predictions and Observed Frequencies of Each Action for Each Treatment Condition

	Frequencies			
Theory	**Rock**	**Paper**	**Scissors**	**Dumb**
- Nash	0.167	0.167	0.167	0.5
- QRE	[0.167, 0.25]	[0.167, 0.25]	[0.167, 0.25]	[0.25, 0.5]
- TASP	0.333	0.333	0.333	0
Observed Time Average				
Unstable, High payoffs	0.226 (0.239)	0.231 (0.228)	0.280 (0.178)	0.263 (0.356)
Unstable, Low payoffs	0.221 (0.194)	0.203 (0.178)	0.207 (0.189)	0.368 (0.439)
Stable, High payoffs	0.176 (0.200)	0.233 (0.144)	0.212 (0.228)	0.378 (0.428)
Stable, Low payoffs	0.172 (0.161)	0.204 (0.183)	0.204 (0.228)	0.420 (0.428)

Note: Final 5 periods shown in parentheses

We focus on the following reduced form model of subjects' choice of the critical strategy Dumb,

$$y_{it}^{*} = \sum_{j=1}^{3} \beta_{1j} D_j (1/t) + \beta_2 ((t-1)/t) + u_i + v_{it},$$

$$y_{it} = 1 \text{ if } y_{it}^{*} > 0 \text{ and 0 otherwise.}$$

The *i*, *j* and *t* subscripts index the subject, session, and 10-period block, and the D_j are dummy variables that have the value of 1 for the indicated session within each treatment. We assume logistically-distributed errors $u_i + v_{it}$, including a random effect error component for subject *i* (u_i), so this can be estimated using a binary random effects logit model. [This panel data approach accounts for the repeated choices made by the same individual subjects and the resulting non-independence of actions within subjects and within sessions.] This reduced form empirical specification was developed for market experiments (e.g., Noussair et al., 1995) that converge in a small number of periods, which is too fast for our game. We therefore employ 10-period blocks for the time index to slow the fitted convergence process, so that the time index t=1 in periods 1-10. Since (t-1)/t is zero in those periods, the β_{1j} coefficient provides an estimate for the probability of choosing Dumb during the first few periods of session *j*. As $t \to \infty$ the 1/t terms approach 0 while the (t-1)/t term approaches one. Thus the β_2 coefficient provides an estimate of the asymptotic probability of choosing Dumb in the treatment.[8]

All three models discussed in Section 2 (Nash, QRE, TASP) predict stable long-run rates of Dumb play, although play of the other actions continues to cycle in TASP. The Dumb strategy is played half the time in the Nash equilibrium, which implies the null hypothesis of β_2=0 since the logit model probability $F(x)$=exp(x)/[1+exp(x)] is 0.5 at x=0. Table 2 presents the estimation results for the asymptote β_2 coefficients. Only the high payoffs, unstable game asymptotic estimate is significantly different from 0. This indicates that the Dumb strategy is not converging toward the Nash equilibrium rate of 0.5 only for the high payoffs, unstable game treatment. The data thus reject the Nash equilibrium Hypothesis 1 only for this treatment.

The average choice data in Table 1 lie in the wide interval predicted by the QRE Hypothesis 2. However, the data provide no support for the underlying qualitative QRE prediction that play will be closer to Nash in the high payoff treatments, i.e., that the coefficient

[8] We examined several alternative specifications. One alternative drops the 1/t and ((t-1)/t) terms and instead simply uses treatment dummy variables and time trends in a random effect logit model. We also specified t in 1-period blocks, 5-period blocks, and as ln(period) and also estimated the model using probit instead of logit specification. All alternatives yielded qualitatively similar conclusions to those reported here, except that the low payoff, unstable β_2 estimate is significantly less than zero for the shorter time intervals that try to fit more rapid convergence.

estimates will be closer to 0 for high payoffs than for low payoffs. Indeed, comparing columns (1) and (2), and columns (3) and (4) of Table 2, we see the opposite pattern , although this difference is not statistically significant.

The average frequency of the Dumb action in Table 1 is not close to zero, as in TASP Hypothesis 3. However, Table 2 reports estimated asymptotic rates of Dumb play that are further from the Nash equilibrium and closer to the TASP prediction for the unstable than the stable treatment for high payoffs. Column (5) indicates that these differences are statistically significant. The $\hat{\beta}_2$= -1.039 estimate for the unstable, high payoff treatment implies an asymptotic point estimate of a 26 percent rate for the Dumb strategy. While this rate is below the Nash equilibrium, it is in the QRE interval and it is also well above the rate of 0 predicted by TASP. Also contrary to TASP, play is *not* closer to the Nash equilibrium in the stable, high payoff treatment than in the stable, low payoff treatment. Thus, data are consistent with the two of the three comparative statics predictions of TASP, but clearly not with its point prediction.

4.2 Learning and Stochastic Best Responses

Average play moves closer to the Nash equilibrium frequencies over time. This is not anticipated by learning theory for the unstable game, particularly for the high payoff treatment. In order to provide insight into the possible learning process employed by the subjects, we empirically estimate the EWA learning model that was presented in Section 2.

In our application, the realized (or forgone) profit $\pi_i(t\text{-}1)$ is calculated based on the observed action chosen by the paired player in the previous period. (Similar results obtain if we use the entire vector of all 11 other players' previous actions.) Our implementation of the model incorporates stochastic best responses through a logit choice rule, the same specification as typically used in QRE applications. For our application with 80 periods, 36 subjects per treatment and 4 possible actions, the log-likelihood function is given by

$$LL(A(0), N(0), \phi, \rho, \delta, \lambda) = \sum_{t=1}^{80}\sum_{i=1}^{36} \ln\left(\sum_{j=1}^{4} I(s_i^j, s_{-i}(t)) \cdot P_i^j(t)\right),$$

where I is an indicator function for the subjects' choice and $P_i^j(t)$ is player i's probability of choosing action j.

Table 2: Random Effects Logit Model of Dumb Action Choice

Dependent Variable = 1 if Dumb Chosen; 0 otherwise

Estimation Dataset	**High Payoff× Unstable× (*t*-1)/*t* (1)**	**Low Payoff× Unstable× (*t*-1)/*t* (2)**	**High Payoff× Stable× (*t*-1)/*t* (3)**	**Low Payoff× Stable× (*t*-1)/*t* (4)**	**Probability Stable and Unstable coefficients equal[a] (5)**	**Obser-vations**	**Subjects**	**Log-L**
All 80-period Sessions	-1.039** (0.302)	-0.387 (0.299)	-0.225 (0.299)	0.106 (0.297)	0.028 (High Pay) 0.121 (Low Pay)	11400	144	-5432.4
High Payoffs 80-period Sessions	-1.039** (0.300)		-0.225 (0.297)		0.027	5640	72	-2528.9
High Payoffs 160-period Sessions (see Section 4.4)	-0.733* (0.312)		0.247 (0.440)		0.035	11520	72	-5510.6

Note: Session×(1/*t*) dummies included in each regression are not shown. The time variable t is measured in 10-period intervals, with *t*=1 corresponding to periods 1-10, *t*=2 corresponding to periods 11-20, etc. Standard errors in parentheses. * indicates coefficient significantly different from 0 at the 5-percent level; ** indicates coefficient significantly different from 0 at the 1-percent level (two-tailed tests). [a]One-tailed likelihood ratio tests, as implied by the TASP research hypotheses.

Table 3: Experience-Weighted Attraction and Stochastic Fictitious Play Learning Model Estimates

	EWA					EWA					Weighted Stochastic Fictitious Play			
						random parameter estimates					*ϕ=ρ, δ=1*			
	Unstable, Low Payoffs	Unstable, High Payoffs	Stable, Low Payoffs	Stable, High Payoffs		Unstable, Low Payoffs	Unstable, High Payoffs	Stable, Low Payoffs	Stable, High Payoffs		Unstable, Low Payoffs	Unstable, High Payoffs	Stable, Low Payoffs	Stable, High Payoffs
Decay Parameters														
ϕ	0.889	0.882	0.910	0.934	E(ϕ)	0.878	0.873	0.834	0.879	ϕ	0.997	0.828	1.000	1.000
	(0.029)	(0.024)	(0.012)	(0.009)	CV(ϕ)	0.193	0.171	0.152	0.175		(0.024)	(0.111)		
					median ϕ	0.950	0.932	0.869	0.940					
ρ	0.568	0.488	0.634	0.529	E(ρ)	0.871	0.870	0.702	0.870	ρ				
	(0.171)	(0.155)	(0.075)	(0.205)	CV(ρ)	0.192	0.171	0.152	0.175					
					median ρ	0.942	0.927	0.731	0.930					
Imagination Factor														
δ	0.000	0.000	0.000	0.000	δ	0.072	0.114	0.000	0.003	δ	1.000	1.000	1.000	1.000
	(Likelihood maximized at 0 bound)					(0.060)	(0.075)	(0.000)	(0.052)		(Constrained at 1)			
Payoff sensitivity														
λ	0.014	0.011	0.015	0.010	E(λ)	0.067	0.055	0.024	0.067	λ	0.030	0.012	0.016	0.019
	(0.004)	(0.003)	(0.003)	(0.004)	CV(λ)	0.442	0.366	0.763	0.352		(0.006)	(0.005)	(0.003)	(0.004)
					median λ	0.061	0.052	0.019	0.063					
Log-Like	-3053.5	-3091.3	-3055.6	-3007.6		-2978.0	-3074.5	-3009.2	-2985.8		-3843.4	-3732.9	-3889.6	-3863.3

Notes: Standard errors in parentheses. CV denotes the estimated coefficient of variation (=standard deviation/mean) for the parameter distribution.

Table 3 reports the maximum likelihood estimates for this model.[9] Decay parameter ϕ estimates always exceed decay parameter ρ estimates, and in all four treatments a likelihood ratio test strongly rejects the null hypothesis that they are equal. Nevertheless, the right side of that table imposes the restrictions of $\phi = \rho$ and $\delta = 1$ to implement the special case of weighted stochastic fictitious play.[10]

A drawback of the estimation results shown on the left side of Table 3 is that they pool across subjects whose learning could be heterogeneous. Wilcox (2006) shows that this heterogeneity can potentially introduce significant bias in parameter estimates for highly nonlinear learning models such as EWA. He recommends random parameter estimators to address this problem, and with his generous assistance we are able to report such estimates in the center of Table 3. The assumed distributions are lognormal for λ and a transformed normal (to range between 0 and 1) for ϕ and ρ. The table reports the mean, the coefficient of variation (standard deviation/mean) and the median to summarize the estimated distributions of these parameters. The point estimates for ϕ are similar to the central tendency of the estimated distributions, but for ρ and λ the point estimates are somewhat lower than the estimated distribution means. Although this is consistent with a statistical bias arising from imposing homogeneity, these random parameter estimates do not qualitatively change the puzzling finding that the imagination factor δ is near zero. That is, subjects' learning evolves as if they focus only on realized payoffs and actions, and not on unchosen actions, contrary to fictitious play learning. This is more consistent with simple reinforcement learning, but simulations of our games for variations of reinforcement learning models (including those specifications considered in Erev and Roth, 1998) indicate that under reinforcement learning there is no predicted difference in behavior between the stable and unstable games. The simulations indicate that the frequency of Dumb rises only to about 30-35 percent, and is similar for both games. So, while a form reinforcement learning is suggested by our EWA estimates, the fitted standard reinforcement learning models do not reproduce the main qualitative features of our data.

[9] We impose the initial conditions $A(0)=1$ and $N(0)=0$ for all four strategies, but the results are robust to alternative initial attraction and experience weights.

[10] The restriction of $\delta = 1$ implies that a subject's unchosen actions receive the same weight as chosen actions in her belief updating, which is the assumption made in fictitious play learning.

Note also that the payoff sensitivity/precision parameter estimates (λ) are always quite low, suggesting that some subjects may have simply randomized uniformly. Indeed, we cannot reject the null hypothesis that action choices are randomly allocated across the four actions for 20 of the 144 individual subjects. The estimated payoff sensitivity parameter estimates also never approach the critical level (0.17) identified in the continuous time and simulated learning models (Section 2).[11] The estimates also do not increase systematically with the treatment change from low to high payoffs. This suggests that subjects were not very sensitive to payoff levels and were not more sensitive to payoffs that were 2.5 times higher in the high payoff treatment. In other words, although as predicted by TASP subjects played Dumb less frequently in the Unstable/High payoff treatment, the structural estimates of this learning model suggest that they did not respond as expected to this treatment manipulation.[12]

4.3 Beliefs and Best Responses

Recall that average play is expected to converge in all treatments. However, if the TASP is a reasonable approximation of final outcomes then in the unstable game treatment, beliefs should continue to cycle, in contrast to equilibrium notions such as NE or QRE that predict that beliefs should converge. The difficulty in identifying a cycle is that its period depends on how quickly players discount previous beliefs and their level of payoff sensitivity. As documented in the previous subsection, these behavioral parameters are estimated rather imprecisely and the weighted stochastic fictitious play model is a poor approximation of subject learning for these games. Nevertheless, we can compare whether beliefs vary more in later periods in the unstable game than the stable game.

Table 4 summarizes this comparison using the mean absolute value of subjects' change in their reported belief from one period to the next, for each of the four actions. Although beliefs change by a smaller amount in the later periods for all treatment conditions, this increase in belief stability is insignificant in the unstable, high payoff treatment. Beliefs change on average

[11] In the random parameter estimates shown in the middle of Table 3, the 99th percentile of the estimated lognormal distribution of λ is less than 0.17 in all treatment conditions.

[12] Prompted by a referee's suggestion, we also estimated an alternative "nested" EWA model. In this model the top-level decision of whether to play Dumb or choose an action from the RPS portion of the matrix is separate from the lower-level decision regarding which RPS action to select, and we estimated separate payoff sensitivity/precision parameters for these two decisions. These estimates indicate that for all four treatment conditions the payoff sensitivity remains very low, in the 0.01 to 0.03 range for the pooled data, although the sensitivity is two to three times higher for the top-level than for the lower-level.

by 2 percent less in periods 41-80 compared to periods 1-40 in the Unstable/High treatment. By comparison, beliefs change on average by 24 percent less in periods 41-80 compared to periods 1-40 in the other three treatments. This provides evidence that belief stability improves over time *except* for the Unstable/High payoff treatment.

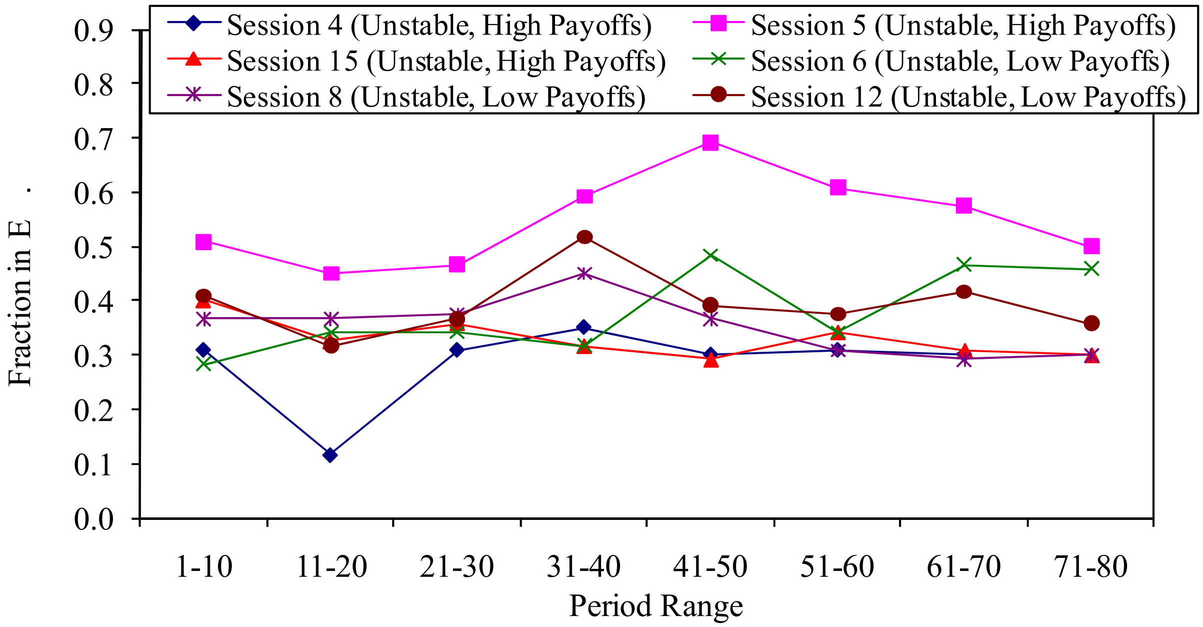
Table 4: Mean Absolute Change in Reported Beliefs

	Unstable, Low Pay		Unstable, High Pay		Stable, Low Pay		Stable, High Pay	
	Period <41	Period >40	Period <41	Period >40	Period <41	Period >40	Period <41	Period >40
Rock	0.147	0.107	0.117	0.128	0.111	0.087	0.133	0.091
Paper	0.148	0.093	0.129	0.130	0.135	0.097	0.132	0.094
Scissors	0.153	0.113	0.139	0.126	0.130	0.088	0.139	0.095
Dumb	0.133	0.146	0.092	0.083	0.135	0.115	0.137	0.114
Ave. % reduction In belief change periods 1-40 to periods 41-80	20.2%		2.2%		24.2%		27.2%	

Consider next the relationship between beliefs and best responses. As discussed in Appendix A (Proposition 2), the set of mixed strategies can be partitioned into a set E, for which the best response is Dumb, and everything else (denoted set F) where the best response is one of the other three strategies. In the unstable game the set E is a pyramid with the Shapley triangle as its base and the Nash equilibrium as its apex. Importantly, the point where all actions are chosen with equal probability is in this pyramid, and for many sessions average play begins (roughly) in this region. Therefore, we might expect the frequency of Dumb to increase initially. But as discussed in Appendix A, under fictitious play like learning, beliefs should move out of E into F and then the frequency of Dumb would begin to fall.

Since subjects report their beliefs each period when choosing their action we have a direct measure of when beliefs are in each set. Figure 6 displays the fraction of reported beliefs in set E for each of the six sessions with the unstable game. Although some variation exists across sessions, in most periods between one-third and two-thirds of subjects report beliefs in E. No session shows a substantial downward trend in the fraction of beliefs in E. At the basis of stochastic fictitious play and QRE is the idea that players do not choose the best response to their beliefs with probability one. Nevertheless, we observe subjects in the unstable game choose Dumb 893 out of the 2164 times their reported beliefs are in set E (41.3 percent), and they chose Dumb 893 out of the 3476 times their reported beliefs are in set F (25.7 percent). So, there is a correspondence between beliefs and actions, yet the general upward trend in the frequency of D is not matched by an increase in the frequency of beliefs being reported in region E.

Figure 6: Fraction of Reported Beliefs in Dumb Best Response Region (Set E)



The learning model estimates in Table 3 suggest that the belief decay parameter is close to one, particularly when imposing parameter restrictions consistent with weighted stochastic fictitious play (ϕ=ρ, δ=1). Of course, due to the low estimated payoff sensitivity λ, the likelihood function is very flat in estimating these discounting payoff parameters. Alternative estimates of the best-fitting decay parameter based directly on reported beliefs (not shown) also indicate a best pooled estimate near one. We also calculated the best-fitting decay parameter for each individual's reported beliefs based on the same procedure employed by Ehrblatt et al. (2007), which minimizes the squared prediction error between the reported belief and the belief implied by the subjects' experience for each possible decay parameter. Constraining this parameter estimate to the interval [0, 1], the best fit is on the boundary of 1 for 92 out of 144 subjects.

Thus, a large fraction of our subjects appear to update beliefs in a manner consistent with classical fictitious play.

4.4 A Robustness Test Using Long Sessions

In our main experiment subjects participated in 80 periods of play so that sessions could be completed comfortably within a two-hour time period. Even over this relatively long horizon, however, time trends are still evident in the data. Most notably, the proportion of Dumb choices tends to rise over time in all four treatments (cf Figure 5). This naturally raises the question of whether play would converge to Nash equilibrium in these games if sessions were run longer.

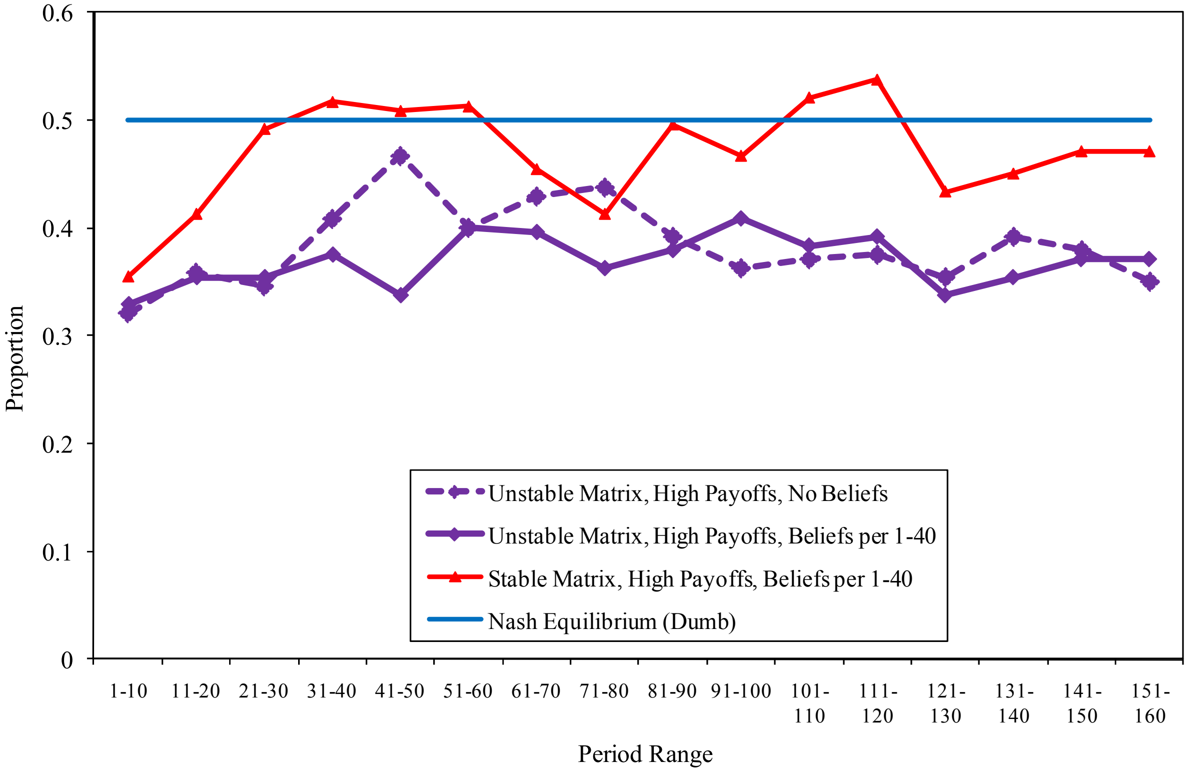
In order to address this question, we conducted six supplementary longer sessions (72 subjects) for 160 periods each, all in the high payoffs condition.[13] In order to keep the per-period expected payoffs comparable to the shorter sessions, we drew twice as many periods (20 in each session) for payment at the conclusion of each session. One of the main factors that slows down subjects is the time required to enter their beliefs each period, since they must submit four numbers that add to 100 in different input boxes. Therefore, to speed play and reduce the tedium of in these longer sessions, we made the additional design change to either (a) only solicit beliefs in periods 1-40 (four sessions), (b) not solicit beliefs in any periods (two sessions). Below we show that behavior in these additional sessions appears unaffected by whether or not beliefs were elicited in the early periods. In spite of this streamlining to move periods along faster, including the instruction and payment time these longer sessions typically required 2.5 to 3 hours to complete. We believe that subjects' attention was nevertheless maintained over these long sessions by the prospect of salient earnings that averaged $54.50 per subject.

Figure 7 shows the proportion of subjects choosing the Dumb action for these supplementary sessions. The dashed line distinguishes the proportion for the two unstable game sessions where subjects never reported beliefs, from the two unstable game sessions with belief elicitation in the early periods (solid line with diamonds). The proportion of Dumb play increases initially, but then stabilizes or falls later in these long sessions. During periods 81-160, the fraction on Dumb ranges between 33 and 40 percent for the unstable game. This is closer to the

[13] These longer sessions were quite expensive to conduct, requiring us to focus on only two of the four original treatments. As already shown in Section 4.1, the average rate that subjects play Dumb is not significantly different from Nash in the sessions with low payoffs, and the greatest deviation from Nash occurs in the high payoffs, unstable sessions. This motivated our choice of the two high payoffs treatments.

Nash prediction than the TASP prediction, but note that the average proportion of Dumb play is higher in the stable game (solid line with triangles) than in the unstable game. This is the game where play should eventually correspond to the Nash prediction of 50 percent on Dumb.

Figure 7: Proportion Choosing Strategy D, Long (160-period) Sessions (6 Sessions Total)



In order to compare the behavior of the stable and unstable game treatments statistically, the bottom row of Table 2 reports the same random effects logit models of subjects' choice of the Dumb strategy for these longer sessions. As before, each session has its own estimated starting probability (through separate β_{1j} coefficients), and separate estimated asymptotic (β_2) probabilities of choosing Dumb for the two treatments.[14] For these new longer-session data, the estimated asymptote for the unstable game is significantly different from 0. Since a coefficient estimate of 0 for this logit model implies a Dumb frequency of 50 percent, for these longer sessions play appears to stabilize at a level less than this Nash equilibrium prediction. By

[14] We first use a likelihood ratio test to determine whether these asymptotes are different for the two variants of the unstable game (i.e., with no belief elicitation, and with belief elicitation in periods 1-40). Consistent with the visual impression of Figure 7, the data do not reject the null hypothesis that the asymptotes are identical (*p*-value=0.658). We therefore pool these two treatments and estimate a common asymptote that combines all the data from these new long sessions.

contrast, the estimated asymptote for the stable game is not significantly different from 0, indicating that the play of Dumb is not inconsistent with the Nash equilibrium in the stable game. Moreover, a likelihood ratio test rejects the null hypothesis that the asymptotes are equal in the stable and unstable games. In summary, for these longer sessions we conclude that the frequency that subjects chose the critical Dumb action is further from Nash for the unstable game than the stable game. Subjects do choose this Dumb action with a higher frequency in these supplementary longer sessions than in the shorter sessions reported earlier, however.

5. Discussion

To summarize, the Nash hypothesis fared poorly in our data. The overall rate of playing Dumb ranged from 26 percent in the Unstable/High treatment to 42 percent in the Stable/Low treatment and only began to approach the NE value of 50 percent towards the end of some long Stable sessions. The performance of the QRE hypothesis was better but also was unconvincing. Although the observed rates fell into the (rather broad) QRE range, the data contradict the main prediction that there should be no difference in observed behavior between the Stable and Unstable treatments.

The TASP hypothesis also had a mixed performance. As predicted, subjects played Dumb least often in the Unstable/High treatment, and most often in the Stable treatments. This is an effect not explicable by equilibrium concepts. Thus this suggests that whether an equilibrium is stable or unstable under learning can be an important factor in subject behavior. On the other hand, the proportion of Dumb play showed no consistent tendency to decline over time, much less to zero, in either Unstable treatment.

What drives these results? First, there is evidence that the difference in behavior across treatments is a result of learning, rather than coincidence or random error. Changes in reported beliefs in all treatments, except the Unstable/High, decrease over time, which is consistent with convergence to equilibrium. The lack of a decrease in the Unstable/High indicates that the different behavior in that treatment is indeed a result of learning being qualitatively different, and convergence more difficult and protracted.

However, the theoretical prediction of complete divergence was not observed. Some clues to this can be found in a more detailed examination of the theory and the data. According to theory, learning dynamics in the Unstable treatments should increase the prevalence of Dumb when players' beliefs lie in a tetrahedral subset of the simplex labeled E, and decrease it only when they lie its complement F. The data show that subjects indeed are more likely to play Dumb when they report beliefs in E than in F. However, reported beliefs show little tendency to move (as predicted) into F. Perhaps the reason is that actual play offers little reason for beliefs to move in that direction. In several of the six Unstable sessions, average actual play (the belief proxy in the classic model of learning dynamics, fictitious play) lies in F in the first 20 periods, but it always moves back into E for the remainder of the 80 periods. Similar results obtain in the long (160-period) sessions, where average play is in E for the final 40 periods in all four unstable sessions.

Another piece of evidence concerns the payoff sensitivity parameter λ. In theory, there is a critical value, $\lambda \approx 0.17$, below which the TASP prediction fails. That is, for sufficiently low values of λ, behavior should be similar in Stable treatments as in Unstable treatments: the rate of Dumb play should remain in the 25-40 percent range and be higher in the High payoff treatments.

We estimate the EWA model using aggregate data, and obtain λ estimates far below the critical value. This can account for the overall rates of Dumb play. To account for the lower rates of Dumb play in the High payoff treatments, we can point to the tendency of the Unstable simulations in Figure 2 to have a lower proportion of Dumb than the theoretical predictions, even when values of λ are relatively low. However, it is also true that the proportion of Dumb play in the Stable treatments is higher, and play is closer to Nash equilibrium, than suggested by the estimated level of λ.

These accounts, of course, raise further questions. In particular, why do players seem to use such small values of λ, i.e., respond so weakly to estimated payoff advantages? This weak response to payoffs would appear to be the best explanation for the difference between our experimental results and the point predictions of both equilibrium and learning theory.

One can think of two potential explanations for this weak responsiveness. Choosing between them may be the key both to understanding our current results and giving directions for further research. First, payoff differences may have been not prominent enough to subjects. In which case, in future experiments, one could improve the feedback or the information provided, perhaps even showing the payoff advantages implied by forecasts and by average play. Second, in contrast, the apparent irresponsiveness of subjects to payoffs may in fact indicate that actual subject behavior is only partially captured by the EWA model, even though this model encompasses many forms of learning. In this case, the challenge is not to change the experimental design but to provide new and more refined theories of non-equilibrium behavior.

Nonetheless, our experimental design provided a simple test as to whether some form of learning matters for behavior in games with mixed strategy equilibria. We conclude that learning is important since equilibrium analysis, whether Nash or QRE, does not readily account for the observed differences between the treatments. Thus, considering the stability or instability of equilibria under learning may help explain observed behavior.

References

Anderson, C.M., C.R. Plott, K. Shimomura and S. Granat (2004). "Global instability in experimental general equilibrium: the Scarf example", *Journal of Economic Theory*, **115**, 209-249.

Bassi, A., R. Morton, K. Williams (2006). "Incentives, Complexity, and Motivations in Experiments", working paper.

Battalio, R., Samuelson, L. Van Huyck, J. (2001). "Optimization Incentives and Coordination Failure in Laboratory Stag Hunt Games," *Econometrica*, **69**, 749-764.

Benaïm, M., Hirsch, M.W. (1999). "Mixed equilibria and dynamical systems arising from fictitious play in perturbed games," *Games Econ. Behav.*, **29**, 36-72.

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

Benaïm, M., J. Hofbauer and S. Sorin (2005). "Stochastic approximation and differential inclusions," *SIAM Journal of Control and Optimization*, **44**, 328-348.

Benveniste, A., M. Métivier, P. Priouret, (1990). *Adaptive Algorithms and Stochastic Approximations*, Berlin: Springer-Verlag.

Brown, J., and R. Rosenthal (1990). "Testing the minimax hypothesis: a reexamination of O'Neill's game experiment," *Econometrica*, **58**, 1065-1081.

Brown Kruse, J., S. Rassenti, S.S. Reynolds and V.L. Smith (1994). "Bertrand-Edgeworth Competition in Experimental Markets," *Econometrica*, **62**, 343-371.

Camerer, C., Ho, T-H. (1999). "Experience-weighted attraction learning in normal form games", *Econometrica*, **67**, 827-874.

Cason, T., Friedman, D. (2003). "Buyer search and price dispersion: a laboratory study," *Journal of Economic Theory*, **112**, 232-260.

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

Cheung, Y-W., Friedman, D. (1997). "Individual learning in normal form games: some laboratory results," *Games Econ. Behavior*, **19**, 46-76.

Dekel, E., and S. Scotchmer (1992). "On the evolution of optimizing behavior", *Journal of Economic Theory*, **57**, 392-406.

Ehrblatt, W. K. Hyndman, E. Ozbay and A. Schotter (2007). "Convergence: An Experimental

Study", working paper.

Ellison, G. and D. Fudenberg (2000). "Learning purified mixed equilibria", *Journal of Economic Theory*, **90**, 84-115.

Engle-Warnick, J., and E. Hopkins (2006). "A simple test of learning theory," working paper.

Erev, I. and A. Roth (1998). "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria," *American Economic Review*, **88**, 848-881.

Fischbacher, U. (2007). "z-Tree: Zurich toolbox for ready-made economic experiments," *Experimental Economics*, **10**, 1386-4157.

Foster, D.P., Young, H.P. (2003). "Learning, hypothesis testing, and Nash equilibrium," *Games and Economic Behavior*, **45**, 73-96.

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

Fudenberg, D., Levine D., (1998). *The Theory of Learning in Games*. Cambridge, MA: MIT Press.

Gaunersdorfer, A., and J. Hofbauer (1995). "Fictitious play, Shapley Polygons, and the Replicator Equation," *Games and Economic Behavior*, **11**, 279-303.

Hofbauer, J. (2000). "From Nash and Brown to Maynard Smith: Equilibria, Dynamics and ESS", *Selection*, **1**, 81-88.

Hofbauer, J. and W. Sandholm (2002). "On the global convergence of stochastic fictitious play," *Econometrica*, **70**, 2265-2294.

Hopkins, E. (1999). "A note on best response dynamics," *Games Econ. Behavior*, **29**, 138-150.

Hopkins, E. (2002). "Two Competing Models of How People Learn in Games," *Econometrica*, **70**, 2141-2166.

Levitt, S.D., J.A. List and D.H. Reiley (2010). "What Happens in the Field Stays in the Field: Professionals Do Not Play Minimax in Laboratory Experiments," *Econometrica*, forthcoming.

McKelvey, R.D., Palfrey, T.R. (1995). "Quantal response equilibria for normal form games," *Games Econ. Behav.*, **10**, 6-38.

Noussair, C., C. Plott and R. Riezman (1995). "An Experimental Investigation of the Patterns of International Trade," *American Economic Review*, **85**, 462-491.

Nyarko, Y. and A. Schotter (2002). “An Experimental Study of Belief Learning Using Elicited Beliefs,” *Econometrica,* **70**, 971–1005.

Offerman, T., J. Sonnemans and A. Schram (1996). “Value Orientations, Expectations and Voluntary Contributions in Public Goods,” *Economic Journal*, **106**, 817-845.

O'Neill, B. (1987). “Nonmetric Test of the Minimax Theory of Two-Person Zerosum Games”, *Proceedings of the National Academy of Sciences*, **84**, 2106-09.

Palacios-Huerta, I. (2003). “Professionals Play Minimax,” *Review of Economic Studies*, **70**, 395-415.

Palacios-Huerta, I. and O. Volij (2008). “Experientia Docet: Professionals Play Minimax in Laboratory Experiments”, *Econometrica,* **76**, 71-116.

Rutström, E. and N. Wilcox (2009). “Stated beliefs versus inferred beliefs: A methodological inquiry and experimental test,” *Games and Economic Behavior*, **67**, 616-632.

Savage, L. (197). “Elicitation of Personal Probabilities and Expectations,” *Journal of the American Statistical Association*, **66**, 783-801.

Shapley, L. (1964). “Some topics in two person games,” in M. Dresher et al. eds., *Advances in Game Theory*, Princeton: Princeton University Press.

Sonnemans, J. and T. Offerman (2001). “Is the Quadratic Scoring Rule really incentive compatible?,” working paper.

Tang, F-F. (2001). “Anticipatory learning in two-person games: some experimental results,” *Journal of Economic Behavior and Organization*, **44**, 221-232.

Walker, M. and J. Wooders (2001). “Minimax Play at Wimbledon,” *American Economic Review*, **91**, 1521-1538.

Wilcox, N. (2006). “Theories of learning in games and heterogeneity bias,” *Econometrica* **74**, 1271-1292.

Wilcox, N. (2010). “‘Stochastically More Risk Averse:’ A Contextual Theory of Stochastic Discrete Choice Under Risk,” *Journal of Econometrics*, forthcoming.

Young, H.P. (2004): *Strategic Learning and its Limits*, Oxford: Oxford University Press.

Appendix A (Stability Properties of RPSD Games)

In this appendix, we state and prove some results on the behavior of the best response (BR) and perturbed best response (PBR) dynamics in the two games $RPSD_U$ and $RPSD_S$. There is already an extensive theoretical literature that shows how the PBR and BR dynamics can be used to predict the behavior of learning under stochastic fictitious play and fictitious play respectively. Specifically, Benaïm and Hirsch (1999), Hopkins (1999b, 2002), and Hofbauer and Sandholm (2002) look at the relation between the PBR dynamics and SFP, while Benaïm, Hofbauer and Sorin (2005) show the relationship between the BR dynamics and classical fictitious play. Finally, Benaïm, Hofbauer and Hopkins (2009) look at the relation between the BR dynamics and weighted fictitious play.

We have seen the perturbed best response dynamics (PBR) in section 2.2. The continuous time best response (BR) dynamics are given by

$$\dot{x} \in b(\pi(x)) - x \qquad \textbf{(BR)}$$

where $b(\cdot)$ is the best response correspondence.

When one considers stability of mixed equilibria under learning in a single, symmetric population, there is a simple criterion. Some games are positive definite with respect to the set $R_0^n = \{\xi \in R^n \ : \ \sum \xi_i = 0\}$, that is for a game matrix A, $\xi \cdot A\xi > 0$ for all non-zero $\xi \in R_0^n$. Mixed equilibria in such positive definite games are unstable, whereas mixed equilibria in games that are negative definite (with respect to R_0^n) are stable.

The game $RPSD_U$ is not positive definite. However, the RPS game that constitutes its first three strategies is positive definite. We use this to show that the mixed equilibrium of $RPSD_U$ is a saddlepoint and hence unstable with respect to the BR and PBR dynamics.

Proposition 1 *In $RPSD_U$, the perturbed equilibrium $\hat{p}$ is unstable under the logit form of the perturbed best response dynamics for all $\lambda > \lambda^* \approx 0.17$.*

Proof: This follows from results of Hopkins (1999b). The linearization of the logit PBR dynamics at $\hat{x}$ will be of the form $\lambda R(\hat{p})B - I$ where R is the replicator operator and B is the payoff matrix of $RPSD_U$. Its eigenvalues will therefore be of the form $\lambda k_i - 1$ where the k_i

are the eigenvalues of $R(\hat{p})B$. $R(\hat{p})B$ is a saddlepoint with stable manifold $x_1 = x_2 = x_3$. But for λ sufficiently small, all eigenvalues of $\lambda R(\hat{p})B - I$ will be negative. We find the critical value of 0.17 by numerical analysis. □

Next, we show that the BR dynamics converge to a cycle which places no weight on the fourth strategy D.

Proposition 2 *The Nash equilibrium* $p^* = (1,1,1,3)/6$ *of the game* $RPSD_U$ *is unstable under the best response (BR) dynamics. Further, there is an attracting limit cycle, the Shapley triangle, with vertices,* $A_1 = (0.692, 0.077, 0.231, 0)$, $A_2 = (0.231, 0.692, 0.077, 0)$ *and* $A_3 = (0.077, 0.231, 0.692, 0)$, *and time average, the TASP, of* $\tilde{x} = (1,1,1,0)/3$.

Proof: We can partition the simplex into two sets. One E is where the best response is the fourth strategy D, and F where the best response is one or more of the first three strategies. It is straightforward but tedious to confirm that the set E is a pyramid with base the Shapley triangle on the face $x_4 = 0$ and apex at the mixed strategy equilibrium p^*. In E, as D is the best response, under the BR dynamics we have $\dot{x}^4 = 1 - x_4 > 0$ and $\dot{x}_i < 0$ for $i = 1,2,3$. If the initial conditions satisfy $x_1 = x_2 = x_3 = (1 - x_4)/3$, then the dynamics converge to p^*. Otherwise, the orbit exits E and enters F. In F, the best response b to x is almost everywhere one of the first three strategies. So we have $\dot{x}_4 < 0$. Further, consider the Liapunov function $V(x) = b \cdot Ax$. We have

$$\dot{V} = b \cdot Ab - b \cdot Ax.$$

As the best response b is one of the first three strategies, we have $b \cdot Ab = 90$ and when x is close to p^*, clearly $b \cdot Ax$ is close to the equilibrium payoff of 45. So, we have $V(p^*) = 45$ and $\dot{V} > 0$ for x in F and in the neighborhood of p^*. Thus, orbits starting in F close to p^* in fact flow toward the set $b \cdot Ax = 90$, which is contained in the face of the simplex where $x_4 = 0$. The dynamics on this face are the same as for the RPS game involving the first three strategies. One can then apply the results in Benaïm, Hofbauer and Hopkins (2009) to show that the Shapley triangle attracts the whole of this face. So, as the dynamic approaches the face, it must approach the Shapley triangle. Then, the time average can be calculated directly. □

The game $RPSD_S$ is negative definite and hence its mixed equilibrium is a global attractor under both the BR and PBR dynamics. This implies it is also an attractor for (stochastic) fictitious play.

Proposition 3 *The Nash equilibrium* $p^* = (1,1,1,3)/6$ *of the game* $RPSD_S$ *is globally asymptotically stable under the best response dynamics. The corresponding perturbed equilibrium (QRE) is globally asymptotically stable under the perturbed best response dynamics for all* $\lambda \geq 0$.

Proof: It is possible to verify that in the game $RPSD_S$ is negative definite and thus its unique Nash equilibrium is an evolutionarily stable strategy or ESS. The first result then follows from Hofbauer (2000, Theorem 4.1) and the second from Hofbauer (2000, Theorem 4.2). □

What do these results imply for stochastic fictitious play? Suppose we have a large population of players who are repeatedly randomly matched to play either $RPSD_U$ or $RPSD_S$. All players use the logit choice rule and update attractions according to the EWA rule given in Section 2.2, but for the special case of SFP with the restriction that $\rho = \phi$ and that $\delta = 1$. Assume further that at all times all players have the same information and, therefore, the same attractions.

Proposition 4 *(a)* $RPSD_U$ *: for* $\lambda > \lambda^* \approx 0.17$ *, the population SFP process diverges from the perturbed Nash equilibrium. If* $\rho = \phi < 1$ *, taking the joint limit* $\rho \to 1$ *,* $\lambda \to \infty$ *and* $t \to \infty$ *, the time average of play approaches the TASP* $\tilde{p} = (1,1,1,0)/3$.

(b) $RPSD_S$ *: the population SFP process will approach the perturbed equilibrium and taking the joint limit, we have*

$$\lim_{\rho \to 1} \lim_{t \to \infty} x(t) = \hat{p}$$

players' mixed strategies will approach the perturbed equilibrium.

Proof: These results follow from our earlier results on the behavior of the BR and PBR dynamics and the application of stochastic approximation theory. For a), see Benaïm, Hofbauer and Hopkins (2009). The result b) follows from the global stability result of Proposition 3, and application of standard results, for example, Theorem 3 of Benveniste et al. (1990, p. 44). □

Appendix B (Experiment Instructions)

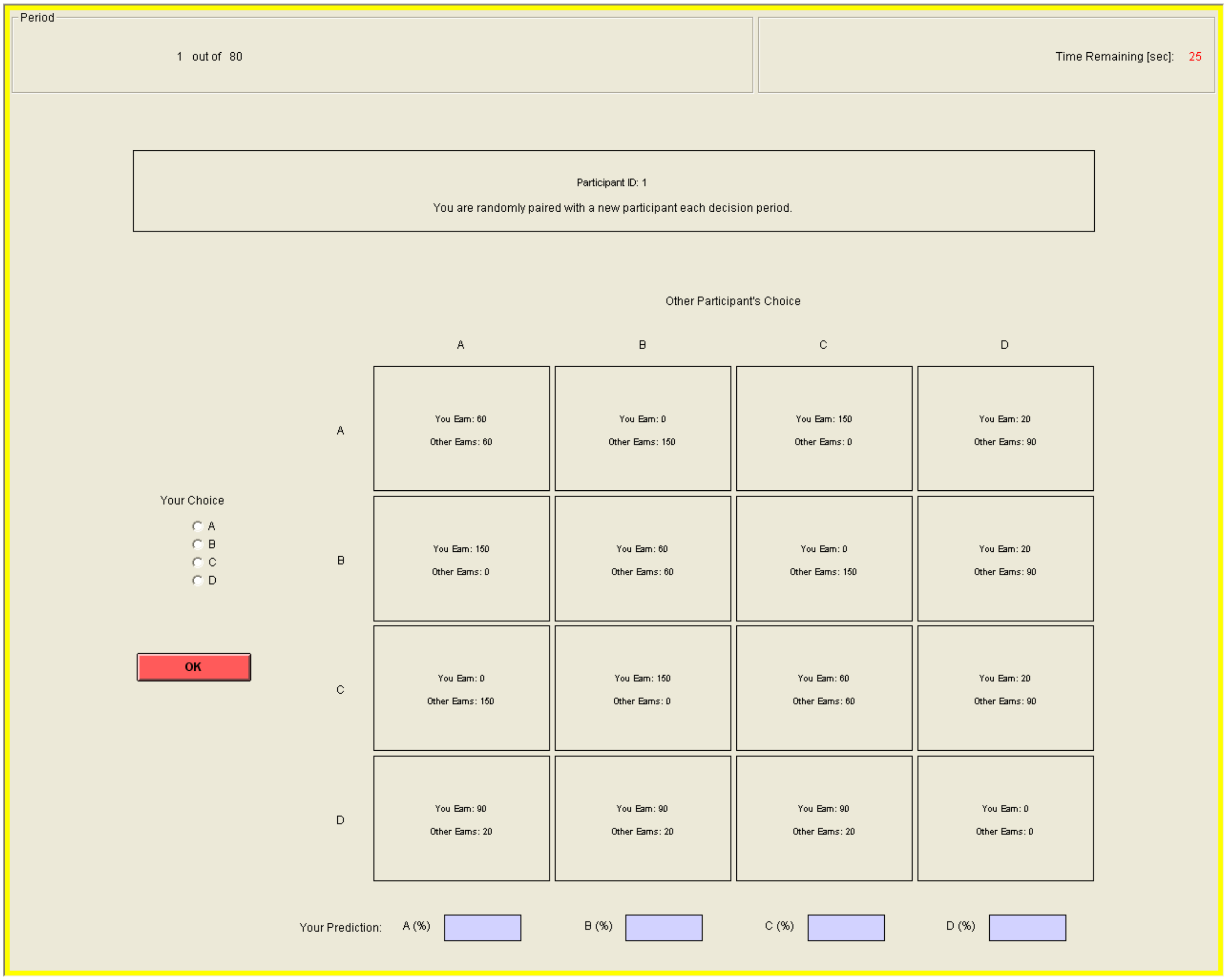
This is an experiment in the economics of strategic decision making. Various agencies have provided funds for this research. If you follow the instructions and make appropriate decisions, you can earn an appreciable amount of money. The currency used in the experiment is francs. Your francs will be converted to dollars at a rate of _____ dollars equals 100 francs. At the end of today's session, you will be paid in private and in cash for ten randomly-selected periods.

It is important that you remain silent and do not look at other people's work. If you have any questions, or need assistance of any kind, please raise your hand and an experimenter will come to you. If you talk, laugh, exclaim out loud, etc., you will be asked to leave and you will not be paid. We expect and appreciate your cooperation.

The experiment consists of 80 separate decision making periods. At the beginning of each decision making period you will be randomly re-paired with another participant. Hence, at the beginning of each decision making period, you will have a one in __11_____ chance of being matched with any one of the __12_____ other participants.

Each period, you and all other participants will choose an action, either A, B, C or D. An earnings table is provided on the decision screen that tells you the earnings you receive given the action you and your currently paired participant chose. See the decision screens on the next page. To make your decision you will use your mouse to click on the A, B, C or D buttons under *Your Choice:* and then click on the OK button.

Your earnings from the action choices each period are found in the box determined by your action and the action of the participant that you are paired with for the current decision making period. The values in the box determined by the intersection of the row and column chosen are the amounts of money (in experimental francs) that you and the other participant earn in the current period. These amounts will be converted to cash and paid at the end of the experiment if the current period is one of the ten periods that is randomly chosen for payment.



Period

1 out of 80

Time Remaining [sec]: 25

Participant ID: 1

You are randomly paired with a new participant each decision period.

Other Participant's Choice

	A	B	C	D
A	You Earn: 60 Other Earns: 60	You Earn: 0 Other Earns: 150	You Earn: 150 Other Earns: 0	You Earn: 20 Other Earns: 90
B	You Earn: 150 Other Earns: 0	You Earn: 60 Other Earns: 60	You Earn: 0 Other Earns: 150	You Earn: 20 Other Earns: 90
C	You Earn: 0 Other Earns: 150	You Earn: 150 Other Earns: 0	You Earn: 60 Other Earns: 60	You Earn: 20 Other Earns: 90
D	You Earn: 90 Other Earns: 20	You Earn: 90 Other Earns: 20	You Earn: 90 Other Earns: 20	You Earn: 0 Other Earns: 0

Your Choice

○ A
○ B
○ C
○ D

Your Prediction: A (%) B (%) C (%) D (%)

Decision Screen

To take a random example, if you choose **C** and the other participant chooses **D**, then as you can see in the square determined by the intersection of the third row (labeled C) and the fourth column (labeled D), you earn 20 francs and the other participant earns 90 francs. The 16 different boxes indicate the amounts earned for every different possible combination of A, B, C and D.

Predictions

When you make your action choice each period you will also enter your prediction about how likely the person you are paired with makes each of his or her action choices. In addition to your earnings from your action choices we will pay you an extra amount depending upon how good your prediction is.

To make this prediction you need to fill in the boxes to the right of *Your Prediction:* on the Decision Screen, indicating what the chances are that the participant you are paired with will make these choices. For example, suppose you think there is a 30% chance that this other person will choose C, and a 70% chance that he or she will choose D. This indicates that you believe that D is more than twice as likely as C, and that you do not believe that either A or B will be chosen. [The probability percentages must be whole numbers (no decimals) and sum to 100% or the computer won't accept them.]

At the end of the period, we will look at the choice actually made by the person you are paired with and compare his or her choice to your prediction. We will then pay you for your prediction as follows:

Suppose you predict that the person you are paired with will choose D with a 70% chance and C with a 30% chance (as in the example above), with 0% chances placed on A and B. Suppose further that this person actually chooses D. In that case your earnings from your prediction are

Prediction Payoff (D choice) = $5 – 5(0.7^2 + 0.3^2 + 0^2 + 0^2) + 10(0.70) = 9.1$ francs.

In other words, we will give you a fixed amount of 5 francs from which we will subtract and add different amounts. We subtract 5 times the sum of the squared probabilities you indicated for the four choices. Then we add 10 times the probability that you indicated for the choice of the person you are paired with actually made (0.7 probability in this example).

For these same example predictions, if the person you are paired with actually chooses A (which you predicted would happen with 0% probability), your prediction earnings are

Prediction Payoff (A choice) = $5 – 5(0.7^2 + 0.3^2 + 0^2 + 0^2) + 10(0) = 2.1$ francs.

Your prediction payoff is higher (9.1) in the first part of this example than in the second part of this example (2.1) because your prediction was more accurate in the first part.

Note that the lowest payoff occurs under this payoff procedure when you state that you believe that there is a 100% chance that a particular action is going to be taken when it turns out that another choice is made. In this case your prediction payoff would be 0, so you can never lose earnings from inaccurate predictions. The highest payoff occurs when you predict correctly and assign 100% to the choice that turns out to the actual choice made by the person you are paired with; in this case your prediction payoff would be 10 francs.

> **Note that since your prediction is made before you know which action is chosen by the person you are paired with, you maximize the expected size of your prediction payoff by simply stating your true beliefs about what you think this other person will do. Any other prediction will decrease the amount you can expect to earn from your prediction payoff.**

The End of the Period

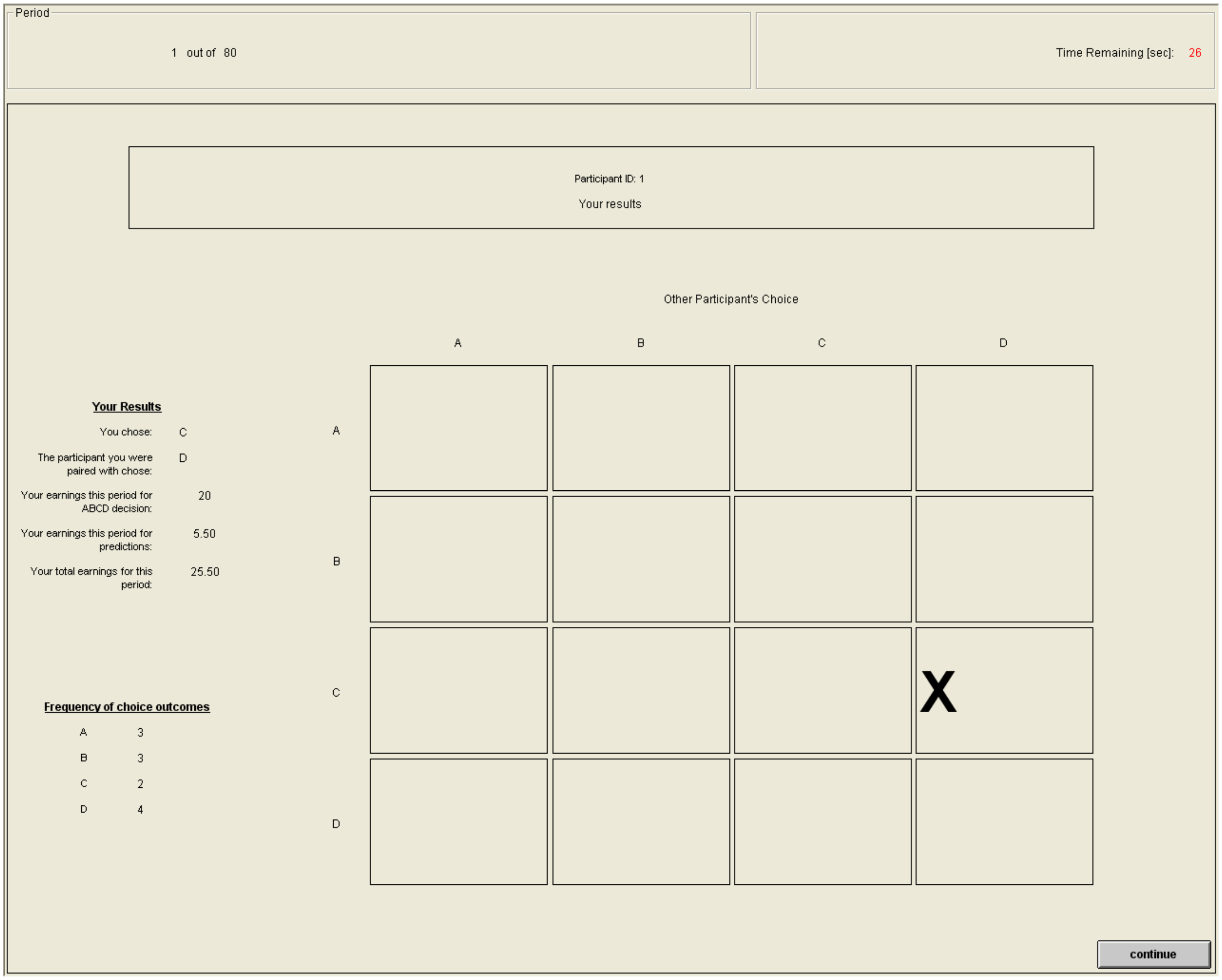
When all participants have made choices for the current period you will be automatically switched to the outcome screen, as shown on the next page. This screen displays your choice as well as the choice of the person you are paired with for the current decision making period. The chosen box is highlighted with a large **X**. It also shows your earnings for this period for your action choice (ABCD decision) and prediction, and your total earnings for the period. The outcome screen also displays the number of A, B, C and D choices made by all participants during the current period.

Once the outcome screen is displayed you should record your choice and the choice of the participant you were paired with on your Personal Record Sheet. Also record your earnings. Then click on the *continue* button on the lower right of your screen. Remember, at the start of the next period you are randomly re-paired with the other participants, and you are randomly re-paired each and every period of the experiment.

The End of the Experiment

At the end of the experiment we will randomly choose 10 of the 80 periods for actual payment using dice rolls (two ten-sided die, one with the tens digit and one with the ones digit). You will sum the total earnings for these 10 periods and convert them to a U.S. dollar payment, as shown on the last page of your record sheet.

We will now pass out a questionnaire to make sure that all participants understand how to read the earnings table and understand other important features of the instructions. Please fill it out now. Raise your hand when you are finished and we will collect it. If there are any mistakes on any questionnaire, I will go over the relevant part of the instructions again. Do not put your name on the questionnaire.



Period

1 out of 80

Time Remaining [sec]: 26

Participant ID: 1

Your results

Your Results

You chose:	C
The participant you were paired with chose:	D
Your earnings this period for ABCD decision:	20
Your earnings this period for predictions:	5.50
Your total earnings for this period:	25.50

Frequency of choice outcomes

A	3
B	3
C	2
D	4

Other Participant's Choice

	A	B	C	D
A				
B				
C				X
D				

Example Outcome Screen