Learning Theory and Heterogeneous Play in a Signaling-Game Experiment

Citation: Fudenberg, Drew and Emanuel Vespa. "Learning Theory and Heterogeneous Play in a Signaling-Game Experiment." AEJ: Macroeconomics, 11, 4 (November 2020): 186-215 © 2020 The Author(s)

As Published: 10.1257/MIC.20180317

Publisher: American Economic Association

Persistent URL: https://hdl.handle.net/1721.1/130346

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

American Economic Journal: Microeconomics 2019, 11(4): 186–215
https://doi.org/10.1257/mic.20180317

Learning Theory and Heterogeneous Play in a Signaling-Game Experiment†

By Drew Fudenberg and Emanuel Vespa*

We study the effect of how types are assigned to participants in a signaling-game experiment. The sender has two actions, In *and* Out, *and two types. In one treatment*, *types are i.i.d. in every period*, *and senders gather experience with both types. In the other*, *types are assigned once-and-for-all*, *and feedback is type specific. The theory of learning in games predicts that the non-Nash but self-confirming equilibrium where some fraction of types play Out can persist in the fixed-type treatment but not when types are i.i.d. Our results confirm that more senders do play Out in the fixed-type treatment.* (*JEL* C92, D82, D83)

In a Nash equilibrium (NE) and in a self-confirming equilibrium (SCE) each agent's strategy is a best response to beliefs about the play of his opponents, but while NE requires that beliefs are exactly correct, in a SCE, beliefs need only be correct along the equilibrium path of play. The notion of a SCE is grounded in the idea that equilibrium corresponds to the long-run outcome of belief-based Bayesian learning by agents who are initially uncertain about the distribution of strategies that prevails in the population, as in Fudenberg and Levine (1993b).[1] When the game has a single round of simultaneous moves, learning the opponents' moves is the same as learning their strategies, and so if agents learn to play a best response to their data and play converges, the convergence point must be a NE. However, when the game has sequential moves or some other sort of nontrivial extensive form, the convergence point may not be a NE. In such games, if agents only observe the outcomes that are reached in their own play of the game, and do not observe what opponents would have selected at off-path information sets, play can converge while the agents

* Fudenberg: Massachusetts Institute of Technology, 77 Massachusetts Avenue, Building E52-418, Cambridge, MA 02139 (email: drewf@mit.edu); Vespa: University of California at Santa Barbara, 2127 North Hall, Santa Barbara, CA 93106 (email: vespa@ucsb.edu). John Asker was coeditor for this article. We thank Chris Ackerman, Ted Bergstrom, Laura Doval, Ignacio Esponda, Kevin He, Ryan Oprea, Alistair Wilson, and Sevgi Yuksel for helpful comments. The authors gratefully acknowledge ISBER at UCSB and NSF grant 1643517 for financial support.
† Go to https://doi.org/10.1257/mic.20180317 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

[1] Kalai and Lehrer (1993) and Fudenberg and Kreps (1995) study related learning models with a single agent in each player role.

maintain incorrect off-path beliefs, even though their beliefs are consistent with their observations.[2]

One key reason why non-Nash behavior can arise in equilibrium is the possibility that incorrect beliefs are not challenged by the data that agents collect over time. In this paper, we explore whether laboratory play can indeed converge to a non-Nash outcome when such an outcome is a best response to incorrect but self-confirming beliefs. More specifically, we compare two treatments, where learning theory predicts the Nash equilibrium in one of them but not in the other. We do this in the context of a simple signaling game, in which the sender has only two types, Blue and Red, and chooses between "*In*" and "*Out*." The receiver only moves if the sender plays *In*, and does not observe the sender's type. Moreover, *In* is the optimal choice when the type is Red, regardless of the receiver's response, while the payoff to (Blue, *In*) can be higher or lower than the payoff to *Out* depending on the choice of the receiver. All equilibrium notions predict that the sender will play *In* when Red. The stage game is designed so that if the Red sender usually selects *In*, the best response for the receiver makes it optimal for a Blue sender to select *In* as well. For this reason, the unique NE of the stage game is for both types of senders to play *In*.

In our sessions, participants are assigned to one role (sender or receiver) and our focus is on the long-run behavior of senders. A sender is anonymously and randomly matched to a receiver in each of the 120 periods. Participants in each role know how their own payoffs depend on the sender's type, the sender's action, and the receiver's action, but they are not told the payoff function of the participants in the other role, so their only source of information about the play of their partners comes from their feedback. All participants received feedback at the end of each period but senders observe the behavior of receivers *only* if they selected *In*.

Our first hypothesis, Hypothesis 0, is that toward the end of the sessions most senders play *In* when Red and most receivers learn to play the corresponding best response. Our design presumes that this prediction will be borne out and focuses on the contrast in behavior in two treatments that differ in the way that the senders' type are determined. In the treatment with random types, each sender's type is i.i.d. across periods, so that participants in this role experience the "Red" and the "Blue" type many times. Thus, if they play *In* when Red, as theory predicts, their feedback will reflect that the receivers' play makes it optimal to also select *In* when the type is Blue. For this reason, in this treatment, the only SCE of this game coincides with the NE. In contrast, in the treatment with fixed types, incorrect beliefs may persist. With fixed types, nature assigns a type to senders once-and-for-all at the beginning of the session. That is, some senders are assigned the Blue type and others are assigned the Red type, and types are fixed for all periods. It is then possible that different participants in the sender's role play differently and so collect different observations, which can lead to persistent heterogeneity in their beliefs and in their play. To model such persistent heterogeneous beliefs, Fudenberg and Levine (1993a) proposed "heterogeneous self-confirming equilibrium" (or

[2] Some applications of SCE include Cho, Williams, and Sargent (2002); Sargent (2001); Fudenberg and Levine (2005); Esponda (2008a, b); and Ali (2011).

heterogenous SCE). "Heterogenous SCE" captures the fact that Blue senders who select *In* will receive different feedback about the play of receivers than Blue senders who select *Out*. A Blue sender who selects *Out* in every period may have incorrect beliefs about what would happen if she were to select *In*. Additionally, in our fixed-type game, heterogeneity of beliefs can be specific to each type. Dekel, Fudenberg, and Levine (2004) extended heterogeneous SCE to "type-heterogeneous self-confirming equilibrium" (or type-heterogenous SCE) to model heterogeneous beliefs in incomplete-information games.[3] With fixed types, it is thus possible that a Blue sender who selects *Out* never learns that receivers are selecting an action that makes it optimal to choose *In*. In fact, the game is constructed so that there is a type-heterogeneous SCE in which Blue senders select *Out* while Red senders select *In*, an outcome that cannot occur in a NE. Moreover, there are also "fully heterogenous SCE," where all Red senders select *In* and some but not all of the Blue senders play *Out*.

This leads us to our main hypothesis, Hypothesis 1, which is that in the last 20 rounds, significantly more player 1 Blues select *Out* in the treatment with fixed types than with random types. Note that observing appreciably more participants play *Out* when Blue in the random-type treatment would reject both Nash equilibrium and learning theory, as would observing similar but non-Nash behavior in both treatments, while observing the Nash outcome in both treatments would be uninformative.

In total, 604 UC Santa Barbara students participated in our experiments, and our main result is in line with the comparative statics suggested by the theory: There is a significant treatment effect of approximately 20 percentage points in the behavior of Blue senders, who settle for *Out* more often with fixed types than with random types.[4] In the treatment with fixed types, in which incorrect beliefs need not be inconsistent with collected feedback, we find that approximately half of Blue-type senders converge to *Out*, a non-Nash outcome that is part of a type-heterogenous SCE. In fact, we document that almost all Blue senders who settle for *Out* collect very little information throughout the session, thus having no data to challenge possibly incorrect beliefs. With random types, meanwhile, the majority of our participants (close to 70 percent) settle for the outcome predicted by the NE.

Moreover, in both treatments, the behavior of Red senders and receivers is as predicted by the theory. All Red senders select the dominant action *In*, and toward the end of the session receivers select the action that maximizes choosing *In* for Blue senders with frequency higher than 95 percent. Because this frequency is not 100 percent, it is possible for standard risk-averse preferences to rationalize an equilibrium selection of *Out* when the sender's type is Blue in the treatment with random types. Indeed, we observe slightly more than 30 percent of our participants

[3] The Appendix reviews the definition of this equilibrium concept.

[4] Our focus is on the long-run outcomes that are reached after several periods of play and feedback and, in particular, whether the observed difference in long-run outcomes in the fixed-type and random-type designs corresponds to the predictions derived from learning theory. In this respect, our paper differs from papers such as Roth and Erev (1995), Cheung and Friedman (1997), Camerer and Ho (1999), and Stahl (2000) that examine period-by-period models of how people learn.

making such choices. To further examine the extent to which these choices are consistent with risk aversion in an environment with experimentation, we conducted additional treatments using a bandit problem. The main findings from these treatments, reported in the online Appendix, are that many participants collect evidence and respond in a manner consistent with the SCE, and that some participants' choices to stay *Out* when Blue in the random-type treatment are consistent with risk aversion.[5]

Overall, the evidence suggests that the treatment effect in our experiments is consistent with the learning mechanism. Blue senders who select *Out* in the fixed-type treatment collect very little information on the behavior of receivers, while Blue senders in the random-type treatment are closely tracking and responding to the information that they collect. Similarly, Fudenberg and He (2019) finds that the relative experimentation rates of laboratory participants in different roles are ordered as Bayesian learning theory predicts. These findings help show the empirical relevance of the learning-in-games program, as does past work that has shown how information (or lack thereof) about other players' payoff functions changes equilibrium outcomes in the way that learning theory predicts.[6]

Our paper also contributes to the experimental literature on incomplete information games by comparing behavior between two different experimental protocols. Most experimental studies of Bayesian games use the anonymous-matching random-type protocol that corresponds to our random-type treatment. The surveys of Kagel (1995) and Kagel and Levin (2017) discuss a large number of auction experiments with such properties. Random-type protocols are also common in experiments on common-value elections (e.g., Guarnaschelli, McKelvey, and Palfrey 2000 and Esponda and Vespa 2014), informational cascades (e.g., Çelen and Kariv 2004) and cheap-talk communication (e.g., Cai and Wang 2006 and Vespa and Wilson 2016). One aspect common to experiments with random-type designs is that participants gather experience from being assigned different types as the session evolves. In contrast, there are relatively few experiments using our fixed-type treatment.[7] The fixed-type design seems a better match for some field settings, for example, when

[5] The experiments in the online Appendix provide a direct comparison between behavior in our signaling experiment and in a bandit problem where one of the players is replaced by a computer that follows a fixed strategy. As far as we know, this is the first experimental comparison of behavior in a game with behavior in a bandit problem. Early contributions to the experimental literature on bandit problems include Meyer and Shi (1995) and Banks, Olson, and Porter (1997), which reported results consistent with under-experimentation. For other experiments that involve bandit problems, see Cox and Oaxaca (2000); Anderson (2001); Brenner and Vriend (2006); Gans, Knox, and Croson (2007); Acuna and Schrater (2008); Biele, Erev, and Ert (2009); Anderson (2012); Norton and Isaac (2012); Hu, Kayaba, and Shum (2013); and Boyce, Bruner, and McKee (2016).

[6] See, for example, the Fudenberg and Levine (1997) analysis of the difference between the full-information and partial-information treatments of the best game in Prasnikar and Roth (1992), where play with partial information corresponded to a non-Nash heterogeneous SCE. For a recent example in which learning theory can rationalize findings, see Araujo, Wang, and Wilson (2018). Our experimental design goes further as learning theory can not only rationalize behavior in each treatment, but crucially predicts a specific treatment effect as well.

[7] Shachat and Walker (2004) is the only paper we know of with fixed types and random matching of partners. Fukuda et al. (2013); McLaughlin and Friedman (2016); and Che, Choi, and Kim (2017) use a design with fixed types and fixed partners.

an agent's type corresponds to a personal trait.[8,9] This paper is the first experimental study that compares behavior under the fixed-type and random-type protocols. Our results show that these two protocols lead to large differences in behavior in the way that learning theory predicts and so provide insight into how to interpret equilibrium notions in Bayesian games.

More broadly, our paper is related to the experimental literature on the difference between fixed-role and changing-role treatments in games of complete information. Güth, Schmittberger, and Schwarze (1982) finds that subjects who simultaneously make decisions as proposer and responder in an ultimatum game make more generous offers, while Binmore, Shaked, and Sutton (1985) finds that when participants play the two roles sequentially their offers are lower. As far as we know, the only such paper to have participants play repeatedly in each role is the contemporaneous work of Ponti et al. (2018). They find that changing roles leads participants to choose more efficient contracts in a principal agent game, which they explain by saying that the changing-role treatment makes subjects more aware of the "interdependence of decisions taken in different roles."

In addition, we study how the assignment of types affects long-run choices in an environment where participants are not told all of the primitives. In our game, participants are not told the payoff function of the participant in the other role. Not providing primitives is standard in market experiments (e.g., Smith 1962), where each consumer and each producer only know their own relevant information (valuations/costs), and the goal is to evaluate if decentralized markets aggregate information.[10] In games, the study of behavior when agents do not know all the primitives is relatively under-explored. A small recent literature includes Fudenberg and Peysakhovich (2014) and Esponda and Vespa (2018).

Finally, our paper contributes to the experimental literature on signaling games. The early experimental work on these games such as Miller and Plott (1985); Brandts and Holt (1992, 1993); Banks, Camerer, and Porter (1994); and Cadsby, Frank, and Maksimovic (1990, 1998) tried to evaluate when play corresponds to the predictions of refinements of Nash equilibrium. Cooper and Kagel (2003) summarizes this literature by saying "experiments have demonstrated that subjects tend to follow simple, history dependent learning processes and, with the right game structure, can be induced into violating even the simplest of equilibrium refinements."

[8] For example, consider a worker who currently works in industry X and could switch to a more lucrative job in industry Y, but believes that there is discrimination against some personal trait of hers (e.g., ethnicity, religious/political beliefs, or marital status) in industry Y. Specifically, in each period, she decides whether to retain her job (*Out*) or accept a job in industry Y (*In*). In each period, she faces different potential employers from industry Y (receiver), and the application process does not ask for the applicant to reveal her personal trait. Once in the job, a worker with her personal trait (Blue type) can be discriminated against. Given her beliefs that there currently is discrimination in industry Y, she may never accept the job opportunity. This can result even if employers do not in fact discriminate.

[9] Yet another possibility is for Nature to make a once-and-for-all choice of type profile that applies to all matches, and use that throughout the session so that all participants in a given player role in fact have the same type as in Cox, Shachat, and Walker (2001); Mitropoulos (2001); Chen (2003); Oechssler and Schipper (2003); Nicklisch (2011); and Feltovich and Oda (2014). Here, the probability distribution p used to select the type profile has only an indirect effect on the outcome of the experiment and has no direct impact on the realized complete-information game, so the design is only indirectly related to equilibrium in Bayesian games.

[10] More recently, Huck, Normann, and Oechssler (1999); Rassenti et al. (2000); and Huck, Leutgeb, and Oprea (2017) study behavior in Cournot settings in which participants are not told information regarding other players.

Similarly, papers by Cooper, Garvin, and Kagel (1997a, b); Potters and Van Winden (1996); and de Haan, Offerman, and Sloof (2011) show that behavior is history dependent in the way suggested by learning models. All of these papers used a random-type design. We use the comparison between fixed and random types to develop a testable prediction of learning theory. Our experiments also show how learning can lead to a non-Nash but self-confirming outcome. Thus, our findings suggest that predictions of learning models should receive more attention in the experimental literature.

I. Experimental Design

A. *Stage Game*

The extensive form of the game we study is presented in Figure 1. The timing is as follows. Nature moves first and with equal probability assigns a type $\theta_1 \in \Theta_1 = \{\text{Red, Blue}\}$ to player 1. Player 1 (he) is informed of his type and then makes a choice $a_1 \in A_1 = \{In, Out\}$. If player 1 selects *Out*, payoffs are implemented and the stage game is over. If player 1 selects *In*, player 2 (she) makes a choice $a_2 \in A_2 = \{\square, \triangle\}$ without observing Nature's move.[11] Player i's payoff, where $i \in \{1,2\}$, is captured by $x_i(a,\theta_1) \in X \subset \mathbb{R}_+$, and depends on the actions $a = (a_1, a_2)$ and on player 1's type.

Studying the payoff functions, we see that when player 1 is Red he wants to play *In* regardless of the play of player 2, and when player 1 is Blue, he wants to play *In* only if player 2 is very likely (Probability $> 10/13$) to play $\square$. Player 2 meanwhile wants to play $\square$ if the probability of Red is higher than one-third. Since player 1 Red always plays *In*, the only Nash equilibrium of this game is for player 2 to play $\square$ and both types to play *In*, and this is also the only SCE with independent types: since all player 1 participants should play *In* when Red, player 2 should learn to play $\square$, and player 1 participants should therefore learn to play *In* when Blue. However, when types are fixed once and for all, there is a type-heterogeneous SCE where some of the player 1 Blues always stay *Out* because they misforecast the play of the player 2's. We explore predictions in further detail as we present our experimental design.

B. *Treatments*

Participants are told their own payoff functions, the action space, and the type space. They are also told that the probability distribution over types is fixed throughout the experiment. (Section ID explains how we implemented this in more detail.) Participants are not told the payoff function of their opponent and are not told the probability p of player 1's type being Red, which is set to $1/2$.[12]

[11] We use $\square$ and $\triangle$ to describe player 2's actions to match the way we presented the game to the participants; see Section ID for details.

[12] Figures 7 and 8 in online Appendix C present the game in Figure 1 from the perspective of player 1 and player 2, respectively.

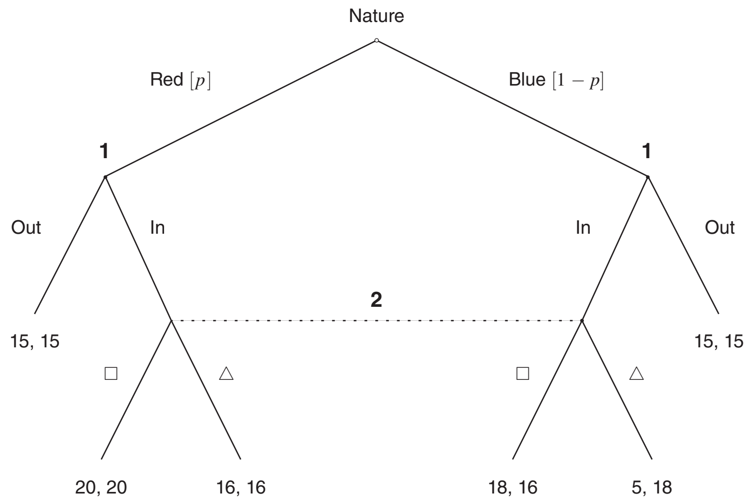

FIGURE 1. EXTENSIVE-FORM REPRESENTATION

At the end of the period, agents learn their own realized payoff. Notice that if player 1 selects *In*, own payoffs are informative in the sense that player 1 can infer what player 2's action was, and player 2 can infer what player 1's type was.

We implement two treatments that differ on how player 1's type is assigned. In the *Game with Random-types treatment* (*GR*), the baseline environment described above is repeated in every period $t = 1, \ldots, T$ of the session. At the beginning of the session ($t = 0$), half of the participants are assigned the role of player 1 and half the role of player 2.[13] In each period, one participant in the role of player 1 is randomly matched with one participant in the role of player 2 and then plays the game described above. In other words, in each period, participants are randomly rematched, and player 1's type is i.i.d. In what follows, we use the labels "temporary Red" and "temporary Blue" to denote a participant in the random-type treatment in the role of player 1 whose current type is Red or Blue, respectively.

The *Game with Fixed-types treatment* (*GF*) is identical to *GR* except that participants in the role of player 1 are allocated types once and for all; that is, types are fixed for all periods. Specifically, at the beginning of the session and with probability $1/2$ each participant in the player 1 role is assigned the Red type. In each period, one participant in the role of player 1 is matched to a participant in the role of player 2. Here, we speak of "permanent Red" and "permanent Blue" participants.[14]

[13] We use $t = 0$ to refer to the period of time before the first decision takes place.

[14] Participants in the role of player 1 are informed of payoffs for both possible types. That is, instructions regarding payoffs are distributed to player 1 before he learns his type for the session. Still, player 1 of either type is not informed of player 2's payoffs, and player 2 is not informed of the payoff function for either player 1 type.

C. *Analysis*

The game in Figure 1 is based on Example 6 in Dekel, Fudenberg, and Levine (2004). In the Appendix, we review the definitions of the equilibrium notions that we use here.

A (behavior) strategy for player 1 is a map: $\sigma_1 : \Theta_1 \rightarrow \Delta(A_1)$, where $\Delta(A_1)$ is the space of probability distributions over A_1, and a behavior strategy for player 2 is a map $\sigma_2 : A_1 \rightarrow \Delta(A_2)$. Given the payoff functions specified in Section IB, the game has a unique Nash equilibrium in which $\sigma_1 = \{In, In\}$ and $\sigma_2 = \square$. It is easy to see that this is a Nash equilibrium. To see why it is unique, note that playing *Out* when Red is strictly dominated, that $\square$ is the unique best response to any player 1 strategy that plays *In* when Red, and that $\{In, In\}$ is the unique best response to $\square$.

When player 1's type is randomly assigned in each period (*GR*), there is a unique SCE, which is *unitary* (meaning that all of the agents in the role of player 1 have the same beliefs), and coincides with the Nash equilibrium characterized above. This is because it is again dominant for player 1 to play *In* when Red, that player 2 will play $\square$, and since all agents in the role of player 1 play *In* when Red, they will learn player 2's response.

Now, consider the game with types fixed throughout the session (*GF*). In this case, it is dominant for player 1 Red to select *In*. Iterated strict dominance then requires that player 2's action is a best response to a belief about 1's type that puts at least the prior probability p on red, so it predicts that player 2 selects $\square$ provided that player 2 believes p is at least $1/3$. As in *GR*, there is an SCE in which player 1 Blue learns player 2 choices and best-responds by selecting *In*. However, the appropriate version of SCE here allows for *type heterogeneity.* That is, since the different player 1 types have different incentives, it is possible that different types persistently select different actions and maintain different beliefs about the responses of the player 2s. Type-heterogeneous SCE allows for this, and in the *GF* treatment, there is a type-heterogeneous equilibrium where the permanent Reds select *In*, the permanent Blues choose *Out*, and player 2 selects $\square$. If permanent Blues select *Out*, they will never observe what player 2 is selecting, and since they gather no information, they will not change their play. More generally, some of the permanent Blues might always play *Out* and maintain incorrect beliefs about 2's play, while others could play *In* and learn that this is their best response. This is a fully heterogenous SCE.[15]

D. *Experimental Sessions*

General Information.—All sessions in this paper were conducted between February and June of 2017 at the Experimental and Behavioral Economics

[15] A permanent Blue may decide not to experiment for several reasons. It is possible that his initial beliefs on player 2 selecting $\square$ are sufficiently pessimistic. Or it may be that his value of the future is not large enough. A combination of these features is also possible. Our experimental data will not provide enough information to determine how much each of these alternatives explains behavior, so we will not focus on exploring such differences.

Laboratory of the University of California, Santa Barbara, and participants took part in only one session. Our game treatments had a total of 300 participants.[16]

We conducted five sessions of *GR*, with a total of 90 participants (45 in each role). For the *GF* treatment, we conducted sessions until the number of player 1 Blue participants was close to the number of player 1 participants in *GR*. In total, we conducted 9 sessions with a total of 45 participants in the player 1 Blue role, 41 in the player 1 Red role, and 86 in the role of player 2.

In each session, participants were assigned to a computer terminal and general instructions were distributed and read aloud. In the general instructions to the game treatments, there is no reference to numeric payoffs, but participants were subsequently provided with instructions specific to each role that described the payoff function. Participants were then prompted to answer questions related to payoffs, and the interface would allow them to move on once they answered all questions correctly.

We used the strategy method for player 2, which means that in our implementation player 2 makes a choice at the same time as player 1, but her choice is only implemented if player 1 selects *In*. The feedback that player 2 received clearly indicated when her choice was used.[17] We did this both to get more data, and because we were concerned that if player 2s could only act if player 1 selects *In*, some player 2s might get bored.

Experimental Environment.—We tell participants that there are two decks of 100 cards each, a Blue deck and a Red deck. The interface randomly selects either the Red deck of cards or the Blue deck of cards. Participants are not informed that each deck is equally likely to be selected. The participants in the "color" role (player 1) are informed of the selected deck (once and for all in the *GF* sessions and at the beginning of each period in the *GR* sessions). Player 2, described as the "shape" role, selects whether all cards have a square or a triangle shape without knowing the color of the selected deck of cards. Player 1 selects either to get a card (*In*) or not get a card (*Out*) without knowing player 2's choice. Once choices are submitted, payoffs for the period are realized. The participants in each role observe their own payoffs, but are not told the payoffs of the participant in the other role.

Figure 2 presents screenshots of the interface at period 7. The top panel displays a screen that player 1 type Blue would see. At the top left of the screen, the participant is reminded of the period (which is referred to as a round), the color of the selected deck, and how payoffs are determined. The participant makes a choice by clicking on one of the options.[18] On the right side of the screen, the participant is presented with information on past play. The table at the top right of the screen reminds them round by round of their choices, the choices of the participant in the other role (whenever player 1 received information on that choice), and their realized payoff.

[16] Adding participants for the experiments described in online Appendices A and B, the total number of participants is 604.

[17] Each time her choice was used she can infer the type of player 1. The equilibria described earlier are not affected by the use of the strategy method.

[18] Once one alternative is selected, a submit button appears below. Participants can change their choices as long as they haven't clicked on the submit button. The experiment was conducted using zTree (Fischbacher 2007).

Panel A. Player 1-type blue in the fixed-types treatment

Panel B. Player 2

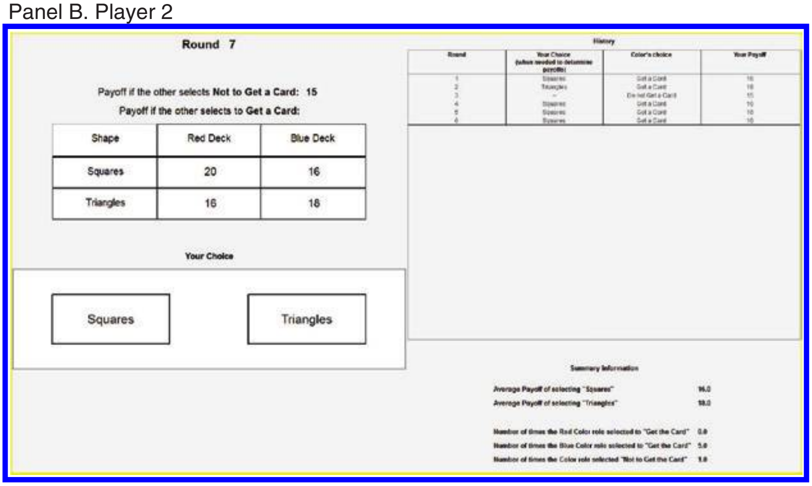

Figure 2. Example of Screenshots in Period 7

At the bottom, they are provided with summary information, which differed in the two treatments. The screenshot displayed in Figure 2 corresponds to the *GF* treatment, where the participant was only provided with the average payoffs they have received for each of the two options. In the *GR* treatment, the participant was given two averages, the average payoff of "Get a card" when the deck was Red and when it was Blue. Finally, participants are shown the number of times they have observed the shape role making each choice. The bottom panel shows the information for the case of player 2.

Note that the summary information we provided computes averages giving equal weight to all past observations. It is possible that behavior could be changed by a different presentation of feedback. For example, from the experiments in Fudenberg and Peysakhovich (2014), we know that recent periods may play a larger role in decision making than earlier periods. Our presentation of summary information may attenuate such effects in both treatments. However, since we used the same averaging procedure in both treatments, we doubt that it would drive the treatment effect.

Finally, we note that in the *GR* treatment the summary information for the payoffs of selecting *In* when Red or Blue is computed by averaging the payoff observations for each type separately. Because information collected when the type was Red is useful when deciding what to do if the type is Blue, we also conducted treatment GR_{HYP}, which again used random types but gave participants feedback that would help temporary Blues use the information that they collect when their type is Red. This treatment, which is described in Section IIA, turned out to have no effect.

The Two Parts of Each Session.—Each session consisted of 120 periods that were divided into two parts. Part 1 consisted of 60 periods, where each period corresponded to the description above. In period 61, when part 2 started, participants were presented with an alternative option to submit their choices. They could either choose to submit their choices just as they had in part 1, or they could decide to program the interface to make future choices for them. We chose this design so that we could tell whether participants who did not change their play from some point on would have done so had the session gone on longer.[19] In *GR*, participants indicated which choice they would want the interface to implement if the type were Red and which choice if the type were Blue. In *GF*, participants submitted a choice only for the type to which they were assigned. Participants in the role of player 2 indicated which choice they would like the interface to implement in future periods. If participants programmed future choices in period 61, the interface implemented the choice that they specified from that period onward, and participants did not make any additional choices in the session. If participants selected to make choices in period 61 just as in part 1, the alternative to program choices was presented to them every 10 periods as long as they had been making choices period by period.[20]

To determine the participant's payoff for the session, one period was randomly selected and the participant received the dollar amount corresponding to the payoff in the selected period. The average payoff was $17.80 and each session took approximately 75 minutes.

E. *Hypotheses*

Before describing our results, we present our hypotheses, which specify how we will test for the theoretical predictions in light of our experimental design. Our main hypotheses are concerned with the long-run behavior of the participants, once they have had the opportunity to learn from their observations. For this reason, we

[19] When reading the general instructions, participants were informed that the experiment consisted of 120 periods divided in two parts and were told that part 2 is identical to part 1 except that in part 2 the interface will provide an additional way for them to submit their choices.

[20] The last period when they are offered to program choices is period 111. All participants were paid once the 120 periods were over. Selecting to program the interface did not lead to participants leaving the laboratory earlier, and participants were informed of this.

express the hypotheses in terms of behavior in the last 20 periods of the 120-period session.[21]

HYPOTHESIS 0: *In the last* 20 *periods of GR and GF*, *most player* 1 *Reds select In and most player* 2*s select* □.

If many player 1 Reds selected *Out* in the last 20 periods, or if a substantial proportion of player 2s ended up selecting △, there would be evidence against both NE and learning theory.

If Hypothesis 0 is satisfied, we can focus on the main comparative static prediction, which involves the behavior of player 1 Blues. Learning theory predicts that some participants in the role of player 1 Blue may select *Out* in *GF*, a prediction that is not part of a NE. Both NE and learning theory predict that player 1 Blues would select *In* in *GR*. The next hypothesis presents the predictions in terms of a treatment effect in the behavior of player 1 Blue.

HYPOTHESIS 1: *In the last* 20 *periods*, *significantly more player* 1 *Blues select Out in GF than in GR.*

If both hypotheses are validated in the data, there would be evidence in support of learning theory. If the data is consistent with Hypothesis 0 but there is no difference across treatments in Hypothesis 1, our experiment would not distinguish between learning theory and NE. Finally, if more player 1 Blues selected *Out* in *GR*, the evidence would support neither NE nor learning theory.[22]

While our focus is on long-run behavior, models of learning also lead to predictions for earlier periods, when participants collect information. However, our experiment was not designed with the aim of characterizing which particular learning strategies participants may use, and there are many ways that learning can lead to equilibrium. For example, in *GR*, one participant may decide to explore the behavior of player 2 by selecting *In* for both types, while another may only explore when the type is Red. Moreover, since the collection of information during experimentation periods is endogenous, it is extremely challenging to find reliable evidence of systematic responses to feedback.[23]

Despite these complications, there are some patterns during experimentation periods that, if observed, would falsify learning theory. Assume that hypotheses 0 and 1 are verified in the data, then according to learning theory player 1 Blues who settle for *Out* in *GF* have incorrect beliefs with respect to the behavior of player 2s. That is, we would expect that player 1 Blues who settle for *Out* in *GF* collect little to no information or get misleading samples that suggest that *Out* is optimal.

[21] Our analysis will show that the "long-run" results hold throughout the second half of the session.

[22] If we observed the participant's beliefs, we would be able to make a direct test of learning theory's explanation that player 1 Blues playing *Out* in *GF* is due to incorrect beliefs. We did not elicit beliefs because we worried that this would change behavior and also because we expected that the elicitation would be very inaccurate. Moreover, what matters for rational experimentation is the participants' second-order beliefs (beliefs over probability distributions) and not just their first-order beliefs.

[23] An experiment designed to study how participants learn in an environment like ours would find a way to, for example, confront several participants with the same feedback.

If, instead, such participants collect a lot of information showing that the player 2s mostly play □, the evidence would be against learning theory.

Another pattern that would be inconsistent with learning theory concerns the behavior of player 2s: if as the session evolves, player 2s select □ with lower frequency, there would be evidence against learning theory.

After documenting the evidence with respect to hypotheses 0 and 1, we will describe the main patterns during experimentation periods, with particular attention to patterns that can falsify learning theory.

II. Results

A. *Behavior in the Last 20 Periods*

The main goal of this section is to evaluate hypotheses 0 and 1 focusing on the last 20 periods of the session.[24] With this aim, we are particularly interested in participants who do not change their choices from some period onward, as this indicates that the participants do not need to collect more information. We provide two definitions of participants whose choices are fixed from some period onward.

DEFINITION (Stable (S) participant): *A participant is said to be Stable if starting at period* $t^S \leq 111$ *the participant did not change her choices from period* t^S *onward.*[25] *We will refer to* t^S *as the first period of stable choices.*

Overall, the vast majority of participants are S participants. Starting at some $t^S \leq 111$, 97.7 and 90.8 percent of participants do not change their choices in the player 1 and player 2 role, respectively.

DEFINITION (Locked (S_L) participant): *A participant who locked in her choices at period* $t^{S_L} \in \{61, 71, 81, 91, 101, 111\}$. *That is, at period* t^{S_L} *the participant programmed the interface to make choices for her. We refer to* t^{S_L} *as the first period of locked-in choices.*

By definition, S_L participants are a subset of S participants. Overall, 89.3 and 85.5 percent of participants in the role of player 1 and player 2, respectively, are classified as S_L participants. Notice that the proportion of S participants who are not classified as S_L participants is relatively small.

According to Hypothesis 0, in both treatments, most player 1 Reds should settle for *In*, and most player 2s should settle for □.

RESULT 0: *In GR and GF*, *all player* 1 *Reds select In in the last* 20 *periods. In the last* 20 *periods of GR and GF*, *the frequency of* □ *choices by player* 2*s is higher than* 95 *percent.*

[24] In the next section, we show that the main patterns we document hold for earlier periods as well.

[25] In *GR*, we say that a player 1 is stable if the participant is not changing her choices for both types from $t^S \leq 111$ onward.

Table 1—Summary of Choices toward the End of the Sessions by Treatment

Participants who selected		Last 20 periods			S participants			S_L participants		
		GR	GR_{HYP}	GF	GR	GR_{HYP}	GF	GR	GR_{HYP}	GF
P1 Red	*In* (percent)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	Out (percent)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Mix (percent)	0.0	0.0	0.0	—	—	—	—	—	—
	Participants	45	19	41	44	18	41	40	11	37
P1 Blue	*In* (percent)	62.3	57.9	40.0	65.9	66.7	48.8	65.0	63.6	45.0
	Out (percent)	33.3	26.3	48.9	34.1	33.3	51.2	35.0	36.4	55.0
	Mix (percent)	4.4	15.8	11.1	—	—	—	—	—	—
	Participants	45	19	45	44	18	43	40	11	40
P2	□ (percent)	91.1	84.2	87.2	97.7	100.0	100.0	97.6	100.0	100.0
	△ (percent)	2.2	0.0	0.0	2.4	0.0	0.0	2.4	0.0	0.0
	Mix (percent)	6.7	15.8	12.8	—	—	—	—	—	—
	Participants	45	19	86	43	19	76	41	19	71

Notes: In "last 20 periods" columns, percentages are computed using all participants. Percentages in "S participants" compute the choices of participants classified as not changing their choices starting at $t^S \leq 111$. "S_L participants" columns compute results using only participants who at some point selected to lock choices in for all remaining periods. The actual number of participants in each case is reported in the corresponding *Participants* row. Additionally, *In* indicates that the participant selected *In* in all corresponding periods (last 20), that an S participant selected *In* from some period onward, or that an S_L participant who decided to program her choices selected to program the computer to select *In*. Similar categories apply to *Out* and to the options of player 2. Mix indicates the proportion of player 1s (player 2s) who selected *In* and *Out* (□ and △) in some periods. By definition, S participants and S_L participants do not mix. Also, GR, GR_{HYP}, and GF indicate the treatments: random types, random types with hypothetical payoffs, and fixed types, respectively; GR_{HYP} is a robustness treatment described later in this section.

Table 1 presents some summary statistics on the treatments that we have already described (GR and GF), as well as treatment GR_{HYP}. This last treatment, which we describe in detail later in this section, is identical to GR except for the way in which feedback is presented. As shown in Table 1, 100 percent of participants in the role of player 1 Red select *In*. This holds if we focus on all participants (last 20 periods), S or S_L participants, and there is no difference across treatments. Meanwhile, in both treatments close to 90 percent of player 2 participants select □ in each of the last 20 periods. The small difference across treatments is not statistically significant.[26] Among S participants and S_L participants, the proportions are close to 100 percent.

While not all player 2s eventually select □, choices of △ are relatively rare. In fact, the aggregate frequency of choices for □ in the last 20 rounds is 96.6 and 95.2 percent in GR and GF, respectively. We now turn to evaluate the comparative statics prediction of Hypothesis 1 about the long-run behavior of the player 1 Blues.

In GF, approximately 40 percent of the permanent Blues select *In* in the last 20 periods, and among S and S_L participants, the frequency is at 48.8 and 45 percent. This means that among player 1s who decide to fix their choices from some point onward, slightly less than half make choices that are consistent with the equilibrium

[26] To test for a treatment effect, we run a panel regression with random effects using the last 20 periods and clustering standard errors by session. The dependent variable takes value one if the participant selected □ and zero otherwise. The right-hand side includes a constant and a treatment dummy. The coefficient on the treatment dummy is small (0.013), and the corresponding p-value is 0.585.

in which permanent Blue selects *In*, and slightly more than half make choices that are consistent with the equilibrium in which permanent Blue selects *Out*.[27]

In *GR*, 62.5 percent of the temporary Blues select *In* in the last 20 periods. The proportion is at 65.9 and 65 percent for S and S_L participants, respectively. In other words, the treatment effect is around 20 percentage points.

To summarize our findings about the behavior of the Blue players, we have the following result.

RESULT 1:

(i) *In the last* 20 *periods of GR, approximately two-thirds of the temporary Blues select In.*

(ii) *In the last* 20 *periods of GF, approximately half of the permanent Blues select In. The treatment effect is between* 15 *and* 20 *percentage points and is significant at* $p = 0.023$.

Columns 4, 5, and 6 of Table 2 present regressions of the treatment effect using the last 20 periods, S participants, and S_L participants, respectively. The dependent variable in all regressions is a dummy that takes value 1 if the participant selected *In*. In column 4, the unit of observation is the choice of a participant in period $t \in [101, 120]$, while in columns 5 and 6 there is one observation per participant. The right-hand side includes a constant and a treatment-effect dummy that takes value 1 if the observation is from *GR*.[28] The treatment effect is statistically significant and negative, indicating that fewer participants settle for *In* when Blue in *GR*, and the magnitude is slightly below 20 percentage points. Specifically, the treatment effect coefficients in 4, 5, and 6 equal -17.8, -17.3, and -19.7 percentage points with p-values of 0.007, 0.023, and 0.032, respectively.

While the proportion of player 1 Blues who select *In* is higher in *GR* relative to *GF*, consistent with Hypothesis 1, about a third of the participants in *GR* make choices that are not consistent with equilibrium behavior. One possible reason for this deviation is related to our particular implementation of *GR*. Recall that we provide participants with three measures of summary information. First, they are reminded of the number of times they observed the participants in the shape role making each choice. This information on player 2's behavior is collected whenever player 1 selects *In*. We also provide them with: (i) the average payoff of *In* when Red and (ii) the average payoff of *In* when Blue. To see why these measures of average payoffs do not use all the information available, consider a participant in *GR* who has selected *In* when Red 20 times and has never selected *In* when the type is

[27] When the type is Blue, if the probability of □ is $10/13 \approx 0.77$, the expected payoff of selecting *In* is equal to the payoff of selecting *Out*. In *GF*, the aggregate frequency of □ in the last 20 periods is 95.2 percent. For those who select *Out*, this suggests that permanent Blues' beliefs that player 2s would select □ were incorrect by at least 18 percentage points.

[28] Because columns 1–3 of Table 2 show that there is no difference in the behavior of player 1 Blue between *GR* and GR_{HYP}, we pool these two treatments in columns 4–6 for the comparison with *GF*, but the treatment effect between *GF* and *GR* is present even if we do not pool *GR* with GR_{HYP}.

TABLE 2—PLAYER 1 BLUE: TREATMENT EFFECTS

	In_t (1)	In_{Blue} (2)	In_{Blue} (3)	In_t (4)	In_{Blue} (5)	In_{Blue} (6)
$D^{1=GR}_{0=GR_{HYP}}$	−0.008	−0.008	0.014			
	(0.077)	(0.079)	(0.087)			
$D^{1=GF}_{0=GR \text{ or } GR_{HYP}}$				−0.178	−0.173	−0.197
				(0.066)	(0.068)	(0.083)
Constant	0.666	0.667	0.636	0.660	0.661	0.647
	(0.030)	(0.029)	(0.026)	(0.049)	(0.050)	(0.062)
Observations	615	62	51	1,515	105	91

Notes: Columns 1 and 4 display results from a random effects regression. Columns 2, 3, 5, and 6 display results of a linear regression. In columns 1 and 4, the dependent variable takes value 1 if in period $t > 100$ the type assigned is Blue and the participant selected *In* and 0 if the type assigned is Blue and the participant selected *Out*. In columns 2 and 5, the dependent variable takes value 1 if the participant is an *S* participant selecting *In* when the type is Blue. In columns 3 and 6, the dependent variable takes value 1 if the participant is an S_L participant selecting *In* when the type is Blue. S (S_L) participants do not change their choice from $t^S \leq 111$ onward (lock in choices at some $t^{S_L} \leq 111$). Columns 1, 2, and 3 compare results using data for *GR* (treatment dummy equals 1 if the observation corresponds to the random-type treatment) and GR_{HYP} (treatment dummy equals 0 if the observation corresponds to the random-type treatment with hypothetical payoffs). Columns 4, 5, and 6 compare results between *GF* (the treatment with fixed types, in which case the treatment dummy equals 1) and *GR* and GR_{HYP} (in which case the treatment dummy equals 0). Standard errors are in parentheses. Standard errors are clustered by session.

Blue. Because this participant has never selected *In* when Blue, we cannot compute an average for this choice. However, the participant has collected observations on the behavior of player 2 that are informative of the payoff he would get if he were to select *In* when Blue. He could compute an expected payoff of selecting *In* when Blue by using the frequencies of observed □s and △s that he has already collected. This leads us to wonder whether the fact that the average payoffs do not take into account all collected information may have led participants to miss the connection between what they learn when the type is Red and what they should do when the type is Blue.

To evaluate this possible explanation, we conducted two additional sessions with random types (total of 38 participants, 19 in each role) in which the participants were told the hypothetical payoff of selecting *In* when Blue and the hypothetical payoff of selecting *In* when Red, where we use the frequencies of observed choices in the shape role to provide them with these computations.[29] We refer to this treatment as GR_{HYP}.[30] We use this treatment to evaluate to what extent choices in *GR* are driven by difficulties in computing expected payoffs and/or incorrect beliefs. We find that there is no difference between the outcomes in *GR* and GR_{HYP}, so the evidence does not suggest that the reason player 1s select *Out* when Blue in *GR* is incorrect beliefs or difficulties computing expected payoffs.[31] Hereafter, we merge the *GR* and GR_{HYP} sessions under the name *GR*.

[29] As an illustration, consider the case of a participant who never selected *In* when Blue but has observed a frequency of □s at 90 percent. This participant would observe the hypothetical payoff of selecting *In* when Blue computed at 16.7, which results from $18 \times 0.9 + 5 \times 0.1$.

[30] See online Appendix D (Instructions), where we provide details on how GR_{HYP} was implemented.

[31] All participants in the role of player 1 Red select *In* in the last 20 periods of both treatments. There is no difference in terms of the behavior of player 2 (e.g., 97.7 and 100 percent of S_L participants select to lock in

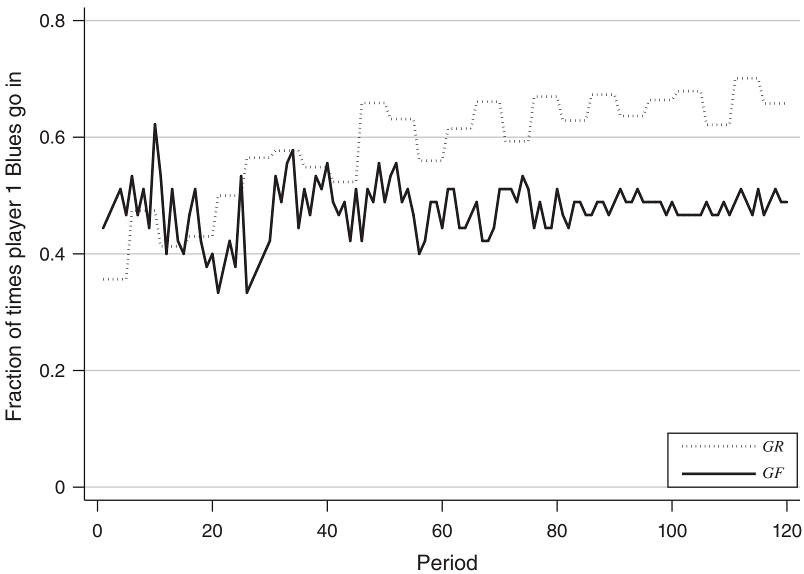

Figure 3. Player 1 Blue

Notes: *GR* and *GF* indicate the treatments: random types and fixed types, respectively; *GR* also includes participants who participated in GR_{HYP}, the random-type treatment with hypothetical payoffs. There is more volatility in the choices of player 1 in *GR* because types are randomly assigned across periods, so the *GR* line displays averages over five-period bins.

B. *The Time Path over the Course of the Sessions*

The behavior of Red player 1s is consistent with Result 0 throughout the session. Overall, player 1 Reds select *In* in 99.9 percent of the cases. The behavior as the session evolves when player 1's type is Blue is summarized in Figure 3. The figure shows for each period the proportion of participants who selected *In*, by treatment. In the case of *GF*, 47 percent of choices in the first 20 periods correspond to *In*, which is close to the 48.1 percent in the last 20 periods. Meanwhile, in *GR*, the fraction of *In* choices in the last 20 periods (67 percent) is higher than in the first 20 periods (43.3 percent).[32] The treatment effect documented in *Result* 1 starts to appear at about the fiftieth period and is consistently present from then on.[33]

Further scrutiny of the behavior of player 1 Blue, but now at the individual level, is provided in Figure 4. These graphs consider *S* participants, that is, the participants

to □). More importantly, there is no significant difference in terms of the behavior of player 1 Blue: 65.9 (65) and 66.7 (63.6) percent of S (S_L) participants select *In* in *GR* and GR_{HYP}, respectively (see Table 2). That is, we find no evidence that the presentation of summary measures of past choices is driving the one-third of participants in *GR* that are selecting *Out* when the type is Blue.

[32] There is more volatility throughout the session in choices of player 1 when assigned the Blue type in *GR* (relative to *GF*). The reason for this is that in *GR*, types are randomly assigned, and so it is possible that in any given period the composition of the participants assigned the Blue type includes a few more or a few less participants that select *In* relative to the average. For this reason, the *GR* line in Figure 3 is computed taking averages over a five-period bin. Figure 9 of online Appendix C is identical to Figure 3 except that averages for player 1 Blue in *GR* are taken period by period.

[33] Statistical support for these claims is provided in Table 9 of online Appendix C, which reproduces the regression reported in column 4 of Table 2 for several subsets of earlier periods.

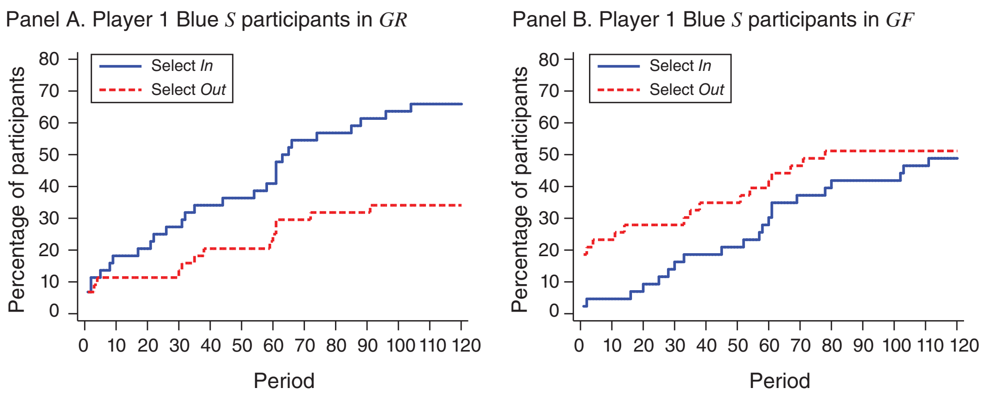

Figure 4. Choices as the Session Evolves

Notes: S participants are participants who did not change their choice from some period onward. Additionally, *GR* and *GF* indicate the treatments: random types and fixed types, respectively; *GR* also includes participants who participated in GR_{HYP}, the random-type treatment with hypothetical payoffs.

who did not change their choices from period t^S onward. For each period, the figures show the proportion of participants who did not change their choices from that period onward, depending on whether they locked on *In* or *Out*. The results are consistent with the aggregate patterns in Figure 3. In *GF* (Figure 4, panel B), the fraction of participants selecting *In* and the fraction selecting *Out* are growing at comparable rates as the session evolves. In *GR*, meanwhile, there is an increase in the difference between the two groups in the second half of the session, with relatively more participants settling on *In*. Comparing *GR* and *GF*, the treatment effect is clearly present throughout the second half of the session.

The aggregate behavior of player 2 as the session evolves is summarized in Figure 5. A clear pattern in both treatments is that the frequency of □ choices increases as the session evolves. In the first 20 periods of *GR* (*GF*), player 2s select □ with frequency 82.6 (80.4) percent, compared to 96.6 (95.2) percent in the last 20 periods.[34] This suggests that the behavior of player 2s is moving toward the equilibrium identified earlier.

We summarize our findings next.

RESULT 2:

(*i*) *Throughout the session and in both treatments, almost all player* 1 *Reds select In.*

[34] The proportions in GR_{HYP} are 79.7 and 97.6 percent for the first 20 and the last 20 periods, respectively. The frequencies change minimally if we restrict to player 2 choices that are observed by player 1s. For example, aggregating *GR* and GR_{HYP}, the frequencies are 80.9 and 97.0 percent. Table 10 of online Appendix C provides statistical support for the claims. There is a significant difference of about 15 percentage points in the frequency of □ choices when comparing the last 20 to the first 20 periods (see column 1). The difference relative to the last 20 periods diminishes as the session evolves. For example, there is only a 2 percentage point increase in the frequency of □ when comparing periods 81 to 100 against the last 20 periods.

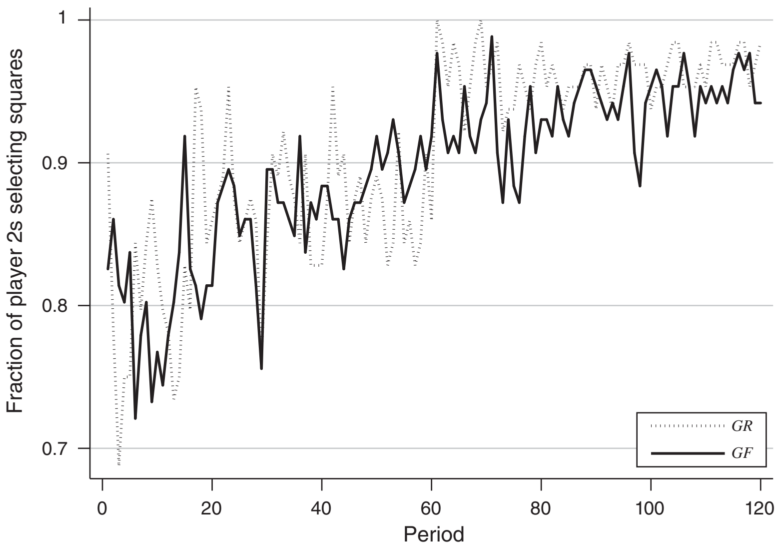

Figure 5. Player 2

Note: GR and *GF* indicate the treatments: random types and fixed types, respectively; *GR* also includes participants who participated in GR_{HYP}, the random-type treatment with hypothetical payoffs.

(ii) *Throughout the second half of the sessions, permanent Blues select Out more often than temporary Blues.*

(iii) *As the session evolves, player* 2*s select* □ *with higher frequency. While in the first* 20 *periods the frequency of* □ *in both GR and GF is close to* 80 *percent, in the last* 20 *periods, it is higher than* 95 *percent.*

Result 2 indicates that the findings documented in Result 0 and Result 1 toward the end of the session are also observed earlier in the session. We now inquire to what extent these results are in line with the information that participants collect as the session evolves. Consider first the SCE where temporary Blues (in *GR*) and permanent Blues (in *GF*) select *In*. These SCE can arise from the following steps taking place as the session evolves.

Step 1: Temporary Reds in *GR* and permanent Reds in *GF* select *In*.

Step 2: Player 2 observes that Red player 1s select *In* and so finds selecting □ more attractive.

Step 3: A player 1 who collects information on the behavior of player 2 observes that the frequency of □ increases and finds it more attractive to settle on *In*. Eventually, permanent Blues and temporary Blues select *In*.

With types fixed in *GF*, the type-heterogeneous SCE where permanent Blues select *Out* would follow from Steps 1 and 2, but Step 3 does not materialize.

Table 3—Collected Observations by Treatment and Final Choice

		In when Blue		*Out* when Blue	
		GR	*GF*	*GR*	*GF*
t^S	Mean	44.8	52	34.3	26.7
	Median	44	57	31	12.5
t^{S_L}	Mean	69.8	73.8	67.1	65.5
	Median	61	71	61	61
Number observed at t^{S_L}	Mean	60.4	54.9	40.6	14.3
	Median	61	51	35	1.5
Number observed $\triangle$ at t^{S_L}	Mean	7.3	7.7	5.9	2.1
	Median	7	7.5	5.5	1

Notes: t^S indicates the first period after which there was no change in choices. Additionally, t^{S_L} indicates the first period in which the participant locked in her choices. Number observed at t^{S_L} indicates the number of times the participant selected *In* prior to t^{S_L}. Number observed $\triangle$ at t^{S_L} indicates the number of times the participant selected *In* and observed a $\triangle$ prior to period t^{S_L}. Also, *GR* and *GF* indicate the treatments: random types (including random types with hypothetical payoffs) and fixed types, respectively.

Permanent Blues who select *Out* throughout the session do not receive feedback on the behavior of player 2 and have no incentive to revisit their choice.

The evidence in Result 2 is clearly consistent with Steps 1 and 2. The difference between the SCE and the type-heterogeneous SCE is present starting in Step 3. The time path over the course of the session would be consistent with the steps if we observed that permanent Blues who select *Out* collected substantially less information than permanent Blues who select *In*. As we summarize next, in our data, this is indeed the case.

Table 3 first presents information on t^S, the period after which stable participants did not change their choices, and t^{S_L}, the first period in which S_L participants lock in their choices. By definition, when choices are locked in, participants can no longer make any changes, so we must have that $t^{S_L} \geq t^S$.[35]

In *GF*, permanent Blues who eventually select *In* collect substantially more information than those who eventually select *Out*. Permanent Blues who select *Out* lock their choices in at the first available moment (period 61), but the median t^S is 12.5. In other words, participants who lock on *Out* in *GF* collect very few informative observations: the median number of periods in which they observed player 2s' choices is 1.5. Meanwhile, the median permanent Blue who eventually selects *In* does not lock their choices in at the first available moment and collects a median of 51 observations before locking their choices in.[36]

Do permanent Blues who select *In* in *GF* commit to a choice with less information than their counterparts in *GR*? The evidence does not show a meaningful difference.[37] On average, temporary Blues who select *In* in *GR* collect 60 observations

[35] Not all participants whose choices do not change from t^S onward decide to lock in their choices, but as described in Table 1, the vast majority of participants are classified as *S* participants, and almost all *S* participants are classified as S_L participants.

[36] Notice that, in fact, this number is not far behind the median number of 61 observations collected by those who select *In* when Blue in *GR*, even though in *GR* participants always observe what player 2 selects when the assigned type is Red.

[37] In *GR*, the median period for t^{S_L} is 61, which is the first period at which choices can be locked in. At the same time, the median t^S is 44 for participants who eventually select *In* and 31 for participants who lock on *Out*, so that by

at t^{S_L}, of which approximately 7 are $\triangle$s. Meanwhile, the average permanent Blue who locks on *In* gathers approximately 55 observations that include between 7 and 8 $\triangle$s.

So far, the data are consistent with Step 3. Permanent Blues who select *In* collect much more information than those who select *Out*, but about the same amount that is collected by temporary Blues who lock on to *In* when Blue in *GR*. However, does the information collected by permanent or temporary Blues who lock on *In* reflect the fact that player 2s increase the frequency of $\square$ choices as the session evolves? The evidence suggests that this is indeed the case.

To evaluate if player 1 participants can assess the behavior of player 2s as the session evolves, consider first a measure of *initial* player 2 behavior. We restrict the sample to permanent Blues who lock on *In* in *GF* and player 1s in *GR* who lock on *In* when the type is Blue. Out of the first 20 observations that a participant collected (the first 20 times a participant selected *In*), how many resulted in observed $\triangle$? Next, consider a *final* measure of player 2 behavior: out of the last 20 observations collected up until the participant locked her choices in, how many resulted in $\triangle$?[38] For each player 1 S_L participant who collected at least 40 observations, we have two measures of player 2 behavior: one collected at the beginning of the session and another closer to t^{S_L}. For participants in *GR*, the *initial* measure of observed $\triangle$ is on average 3.25, while the *final* measure is 1.78.[39] The 1.47 reduction is significant at the 1 percent level.[40] This effect is also present at the individual level: a majority of S_L participants in *GR* who locked on *In* when the type is Blue experienced *initial* observed $\triangle$s that at least double the *final* measure.

A similar finding holds for permanent Blues who eventually lock on *In*. The frequency of $\triangle$ in the *initial* measure is 4.0, but it is reduced to 1.9 in the *final* measure, with the difference being significant at the 1 percent level.

RESULT 3: *Consider temporary Blues in GR and permanent Blues in GF who locked their choices to In. The average of observed $\triangle$s using the first* 20 *collected observations is significantly higher than the average of observed $\triangle$s using the last* 20 *collected observations.*

the first time we allow participants to lock their choices in, most have not changed their choices for several periods. This suggests that if we had allowed for choices to be locked in prior to period 61, several participants would have taken advantage of such an option.

[38] With respect to the *initial* measure, notice that if in *GR* the player 1 participant selects *Out* when Blue and *In* when Red in the first 20 periods, it would take more than 20 periods to collect the first 20 observations. If, instead, the participant selects *In* for both types in all periods, it would take 20 periods to collect the first 20 observations. With respect to the *final* measure, note that if a participant locks choices in starting in period 61, the *final* measure looks for the last 20 periods prior to period 61 in which the participant selected *In*.

[39] Out of 51 S_L participants in *GR* and GR_{HYP}, 38 (75 percent) collected at least 40 observations. Out of 33 S_L participants who select *In* when Blue, 32 collected 40 observations or more until t^{S_L}. There are 6 participants who select *Out* when Blue and collect 40 or more observations by t^{S_L} and 12 who collect fewer than 40.

[40] To test for significance, we run a panel regression with the number of observed $\triangle$ as a dependent variable and a dummy that takes value one if the observation corresponds to the *final* measure and zero if it corresponds to the *initial* measure. The coefficient of the dummy is significant at the 1 percent level. The sample consists of all S_L participants in *GR* and GR_{HYP} who collected at least 40 observations. We would reach the same conclusions if we condition the regression on S_L participants who eventually lock in to *In* when Blue.

According to *Result* 3, participants in both treatments who lock on to *In* when they are Blue collect data that on average shows that player 2s select $\triangle$ less frequently as the session evolves. This is consistent with what Step 3 identified earlier.

An alternative approach to evaluate if behavior is consistent with learning theory is to study whether observed frequencies are predictive of choices. Indeed, we find that this is the case. We run a regression restricted to S_L participants, with one observation per participant. The left-hand side is a dummy variable that takes value 1 if the participant selected *In* when Blue. The right-hand side includes the observed frequency of $\square$ at t^{S_L}. In *GR*, for each percentage point that the frequency of $\square$ increases, the probability that the participant selects *In* increases by 3.24 percent. The coefficient is significant at the 5 percent level (p-value 0.014). Computing the same regression in *GF* is not feasible because many participants who selected *Out* when Blue very rarely played *In*. However, in Figure 10 of online Appendix C, we present the distribution of the frequency of $\square$ depending on the treatment and on whether the S_L participant selected *In* or *Out* when Blue. For both treatments, the distribution when the S_L participant selected *In* is to the right of the distribution when the participant selected *Out*. Hence, in both treatments, observing a higher frequency of $\square$ is associated with a higher likelihood of selecting *In*.

C. *Discussion*

The vast majority of our data is in line with the theoretical predictions, summarized by hypotheses 0 and 1. When the type is Red, player 1s in both treatments behave as predicted. Permanent Blues who select *Out* collect very few observations on the behavior of player 2s, which is consistent with the type-heterogenous SCE. Meanwhile, permanent Blues who select *In*, and player 1s who select *In* for both types in *GR*, collect close to 60 observations on the behavior of player 2, which on average captures the fact that player 2s reduce their selection of $\triangle$ as the session evolves. In fact, toward the end of the session, player 2s select $\square$ more than 95 percent of the time in both treatments.[41] Overall, this means that the main treatment effect documented in Result 1 matches the comparative-statics predictions well.

However, we do observe one marked deviation from our theoretically derived predictions. About a third of player 1s in *GR* select *Out* when the type is Blue, which is not predicted by the SCE or NE. One possible explanation for the one-third of temporary Blues selecting *Out* in *GR* is imperfect learning. That is, temporary Blues may be selecting *Out* when Blue because at the time when they decide to lock choices in they have not collected enough information to properly assess the long-run chances of player 2 selecting $\square$.[42]

Another explanation involves risk preferences. To see why, allow first for a utility function u_i over the payoffs in Figure 1, where u_i is a strictly increasing function of x_i. For any u_1 it is dominant for player 1 Red to select *In*, and for any u_2, it is

[41] In fact, 137 out of 138 player 2 participants who at some point make no changes in their choices in *GF*, *GR*, or GR_{HYP}, fix the choices to permanently selecting $\square$.

[42] As shown in Table 3, the median S_L participant who locks in to *In* has collected more observations (61) than has the median S_L participant who selects *Out* (35).

optimal for player 2 to select □.[43] Moreover, if player 2 selects □, it is a best response for player 1 Blue to select *In* for any u_1. But if the frequency of □ is not at 100 percent, it is no longer the case that the behavior of player 1 Blue is independent of u_1. Depending on participants' preferences for risk, it is possible that in equilibrium some participants end up selecting *In*, while others select *Out*.[44,45]

To test the robustness of this risk-aversion explanation in an environment with learning, we conducted a bandit-problem version of our signaling game in which the receiver is replaced by a computer that is programmed to mimic the behavior of experienced receivers from the signaling-game sessions. That is, the machine is programmed to select □ with probability 0.95, which is the frequency observed in games toward the end of the sessions. Participants are told that the probability is fixed throughout the session, but do know the number it is fixed to. Because the frequency is not at 100 percent, it is possible for a risk-averse sender in the bandit problem to select *Out* when their type is Blue even if they have collected substantial information on the machine's choices.

The experimental design and results are described in detail in online Appendix A. The main finding is that with random types *more* Blue senders select *Out* in the bandit relative to the game. This suggests, first, that indeed some Blue senders may be selecting *Out* in *GR* due to risk aversion. But, in addition, it also suggests that some participants who were not willing to take the risky prospect in the bandit treatment do select *In* when Blue in *GR*.

The finding that *more* Blue senders select *Out* in the bandit relative to the game suggests that some participants may respond differently in the long run depending on whether the frequency is fixed at a high level or increasing toward the high level. We leave this for future research.

III. Conclusion

In many economic environments, agents learn from past feedback about realized play and do not observe intended off-path play by their opponents. For this reason, learning theory predicts that some sorts of incorrect off-path beliefs and non-Nash

[43] The chance of player 2 facing a player 1 Red conditional on player 1 selecting *In* is between $1/2$ and 1. If player 2 only faces player 1 Red, □ is a best response. At the other extreme, selecting □ involves a lottery in which with 50-50 chance the payoffs are either 20 or 16, while selecting △ involves a 50-50 lottery with payoffs 18 or 16. Choosing □ is a best response for any u_2.

[44] Let $\hat{\sigma}_2^{GR}$ be player 1's belief that player 2 will select □ at the time when player 1 decides to fix her choices. Even if $\hat{\sigma}_2^{GR}$ is a good estimate of the actual choices of player 2, an expected-utility-maximizer player 1 will select *Out* when the type is Blue if $\hat{\sigma}_2^{GR} u_1(18) + (1 - \hat{\sigma}_2^{GR}) u_1(5) < u_1(15)$. If the participant's belief is based on observed frequencies, we can then compute the degree of risk aversion that makes a participant indifferent between *In* and *Out* under some functional form for u_1. Assume that the participant has CRRA preferences at $u_1(x) = x^{(1-\alpha)}/(1-\alpha)$ for $\alpha \neq 1$. If $\hat{\sigma}_2^{GR} = 0.8$, which is close to the frequency early in the sessions, then the α at which the participant is indifferent between *In* and *Out* is 0.33. In the low real treatment of Holt and Laury (2002), which comes closest to our incentives, 68 percent of participants make choices consistent with α of 0.33 or higher. If $\hat{\sigma}_2^{GR} = 0.95$, which is close to the frequency later in the sessions, the indifference α equals 0.82. There are 17 percent of participants in Holt and Laury (2002) who make choices consistent with this value or higher.

[45] The choice of *Out* by temporary Blues can also be rationalized by satisficing, as it is possible that participants are satisfied with the payoff of *Out* and do not think further about the consequences of selecting *In* when Blue. If a participant does not collect information when the type is Blue, we cannot identify this possible mechanism separately from risk aversion.

outcomes may persist. This paper presented an experiment in which learning theory predicts a treatment effect due to incorrect off-path beliefs depending on whether types are set once and for all or reassigned each period, a distinction that is crucial from the viewpoint of learning theory but irrelevant in standard analyses of Bayesian games.

Specifically, we studied a signaling game where the sender can be one of two types (Red or Blue) and in each period selects *In* or *Out*. A choice of *In* lets the sender learn what the receiver chooses. But a sender who selects *Out* does not observe the counterfactual payoff she could have received from selecting *In*, as the receiver is not called to play. The game is designed so that it is a strictly dominant strategy to select *In* if the sender's type is Red. Treatments differ on how types are assigned to senders. With fixed types, Blue senders who initially believe that selecting *Out* is optimal will not receive any information and will thus have no reason to change their behavior. For this reason, type-heterogeneous self-confirming equilibrium allows the outcome in which Blue senders select *Out*, even though this cannot occur in a Nash equilibrium or in a self-confirming equilibrium with unitary beliefs. With types randomly assigned after each period, senders will play *In* when they are Red and so gather information on the receivers' choices. Thus, even if the sender initially believes it is better to stay *Out* when Blue, they will eventually learn that it is better to play *In*, which is why "both types *In*" is the only self-confirming equilibrium with unitary beliefs.

We find broad support for the comparative-static predictions. When types are fixed, a majority of Blue senders select *Out*, starting from early in the session. Blue senders who select *Out* indeed collect very little information on receivers. In contrast, with random types, most senders do eventually learn to play *In* when Blue. The one significant departure of our data from the theoretical predictions is that some Blue senders select *Out* with random types. Because not all of the receivers play the response that makes *In* optimal for Blue, staying *Out* can be rationalized if participants are risk averse. In a series of treatments described in online Appendix A, we find that indeed such behavior is consistent with risk aversion.

Our design illustrates that the long-run outcome of the laboratory play of a Bayesian game depends on the time series structure of the stochastic process that governs types, and not just on their per-period marginal distribution, because whether types are independent or correlated over time can have an effect on what players learn about the strategies of others and more broadly on their long-run behavior. While most experimental studies of Bayesian games use a protocol in which types are randomly assigned in every period, our findings suggest that it may be useful to explore behavior using other protocols, especially given the seeming relevance of those protocols to many field settings.

Appendix A. Varieties of Self-Confirming Equilibrium

There are many versions of self-confirming equilibrium and related concepts in the literature, corresponding to different implicit assumptions about the corresponding learning environment. In this Appendix, we review the definitions that are used in the main text, namely unitary SCE, heterogeneous SCE, and type-heterogeneous

SCE. To simplify notation, we restrict to games like our signaling game where player 1 has private information. For more discussion of the interpretation and motivation of these notions and of the additional issues that arise in games with more than two players, see Fudenberg and Levine (1993a) and Dekel, Fudenberg, and Levine (2004); and see Fudenberg and Levine (1993b) and Fudenberg and He (2019) for explicit learning foundations for heterogeneous and type-heterogeneous SCE, respectively.

We study a learning environment in which the game is played repeatedly. Players know their own payoff functions, the sets of possible moves of all other players (A), and the set of types (Θ). Players know neither the strategies used by other players nor the distribution of Nature's move (p). They learn about these variables from their observations after each period of play.

What players learn from repeated play depends on what they observe at the end of each round of play. After each play of the game, players observe the terminal node $z = (a,\theta)$, which is their only information about Nature's and their opponents' moves. Denote (behavior) strategy profiles in the signaling game by $\sigma = (\sigma_1,\sigma_2)$, where $\sigma_1 : \Theta_1 \to \Delta(A_1)$ and $\sigma_2 : A_1 \in \Delta(A_2)$, and let $\rho(z|\sigma)$ be the probability of terminal node z under strategy profile σ.

We begin with the equilibrium concepts that are adapted to our i.i.d. types treatment (GR). Let $\hat{\sigma}_{-i} \in \Sigma_{-i}$ denote player i's conjecture about the play of player $-i$.

In some settings, agents may maintain incorrect beliefs about the distribution of Nature's move, and this can be accommodated by SCE, but we restrict the formal presentation here to the case of correct beliefs, because in the game we study allowing for incorrect beliefs about Nature's move would not enlarge the set of SCE.[46]

DEFINITION: *A strategy profile σ is a* unitary self-confirming equilibrium (*unitary SCE*) *of a signaling game if there are conjectures $\hat{\sigma}_{-1}$ and $\hat{\sigma}_{-2}$ such that:*

(*i*) $\rho(z|\sigma_i,\sigma_{-i}) = \rho(z|(\sigma_i,\hat{\sigma}_{-i}))$ *for* $i = 1,2$ *and all* z;

(*ii*) *for any pair* θ_1, a_1 *such that* $\sigma_1(a_1|\theta_1) > 0$,

$$a_1 \in \arg\max_{a_1'} \sum_{a_2} u_1(a_1', a_2, \theta)\, \hat{\sigma}_{-1}(a_2|a_1');$$

(*iii*) *for any* a_1 *such that* $\sigma_1(a_1|\theta_1) > 0$ *for at least one* θ_1 *and any* a_2 *such that* $\sigma_2(a_2|a_1) > 0$,

$$a_2 \in \arg\max_{a_2'} \sum_{a_1,\theta} u_2(a_1,a_2,\theta_1) p(\theta_1)\, \hat{\sigma}_{-2}(a_1|\theta_1).$$

Here, condition (i) says that the joint distribution of Nature's move and player i's opponent's action is consistent with what player i expected to see given their conjecture about opponent's play, and conditions (ii) and (iii) say that each player's

[46] Player 1's beliefs about the distribution of Nature's move is irrelevant, and since 1 always plays *In* when Red, player 2 will learn that □ is the best response to 1's play.

strategy maximizes its expected payoff given its conjecture. Note that SCE is weaker than Nash equilibrium as it allows players to have incorrect beliefs about how their opponent plays at unreached information sets. Specifically, in our signaling game, if player 1 always plays *Out*, then unitary SCE is consistent with any conjecture at all about how player 2 would respond to *In*. Intuitively, this is because in a learning model a player 1 who expects a very low payoff from *In* might never try it and so never learn how 2 responds.

This version of SCE is "unitary" in the sense that it implicitly assumes there is a single agent in the role of player i and so requires a single conjecture $\hat{\sigma}_{-i}$ to rationalize each action that σ_i assigns positive probability. In the anonymous random matching setting of most laboratory game experiments, it is natural to allow different participants in a given player role to play different strategies and have different conjectures. So suppose now that in the population of agents in each player role i there are several different strategies played, denoted σ_{i,j_i}, with associated conjectures $\hat{\sigma}_{-i,j_i}$, $j_i = 1,\ldots,K_i$, with distributions γ_1,γ_2. This allows, e.g., some of our player 1s to play *In* and learn 2's play while others play *Out* and maintain incorrect conjectures. As shown in Fudenberg and Levine (1997), heterogeneous beliefs are not only theoretically natural, but also lead to a better understanding of a number of laboratory experiments. Let σ_1 and σ_2 be mixed strategies that are equivalent to the aggregate play under the distributions γ_1, γ_2.

DEFINITION: *A strategy profile σ is a* heterogeneous self-confirming equilibrium *of a signaling game if there are distributions γ_1,γ_2 on the strategies and conjectures of each player i such that:*

$$(i)\quad \rho\left(z|\sigma_{i,j_i},\sigma_{-i}\right) = \rho\left(z|\left(\sigma_{i,j_i},\hat{\sigma}_{-i,j_i}\right)\right) \text{ for all } z,i,j_i;$$

(ii) *for any j_1 and for any pair θ_1, a_1 such that $\sigma_{1,j_1}\left(a_1|\theta_1\right) > 0$,*

$$a_1 \in \arg\max_{a_1'} \sum_{a_2} u_1\left(a_1', a_2, \theta\right) \hat{\sigma}_{-1,j_1}\left(a_2|a_1'\right);$$

(iii) *for any a_1 such that $\sigma_{1,j_1}\left(a_1|\theta_1\right) > 0$ for at least one θ_1, j_1, and any a_2 with $\sigma_{2,j_2}\left(a_2|a_1\right) > 0$,*

$$a_2 \in \arg\max_{a_2'} \sum_{a_1,\theta} u_2\left(a_1, a_2, \theta_1\right) p\left(\theta_1\right) \hat{\sigma}_{-2,j_2}\left(a_1|\theta_1\right).$$

In some signaling games, there are heterogeneous SCE that are not SCE. But in the game of our experiment, in any SCE, all player 1s must play *In* when Red, and SCE then requires that they have correct beliefs about how player 2s respond. Thus, in the *GR* treatment, all SCE are not only unitary but are Nash equilibria.

Now suppose that each agent's type is fixed once and for all as in *GF*. In a unitary type-heterogeneous self-confirming equilibrium, all agents in the role of a given type θ_i are required to have the same belief, and this belief is required to be consistent with how they play, but agents in the roles of different types can play differently and maintain different beliefs.

DEFINITION: A *strategy profile* σ *is a* unitary type-heterogeneous self-confirming equilibrium *of a signaling game if there are conjectures* $\hat{\sigma}_{-1,\theta_1}$ *for each* θ_1 *and a conjecture* $\hat{\sigma}_{-2}$ *for player* 2 *such that:*

(i) $\rho\big(z|\big(\sigma_1,\sigma_2\big)\big) = \rho\big(z|\big(\sigma_1,\hat{\sigma}_{-2}\big)\big)$ *for all* z, *and* $\rho\big(z|\big(\sigma_1,\sigma_2\big)\big) = \rho\big(z|\big(\sigma_2,\hat{\sigma}_{-2}\big)\big)$ *for each* θ_1;

(ii) *for* θ_1, a_1 *such that* $\sigma_1\big(a_1|\theta_1\big) > 0$,

$$a_1 \in \arg\max_{a_1'} \sum_{a_2} u_1\big(a_1', a_2, \theta_1\big)\, \hat{\sigma}_{-1,\theta_1}\big(a_2|a_1'\big);$$

(iii) *for any* a_1 *such that* $\sigma_1\big(a_1|\theta_1\big) > 0$ *for at least one* θ_1, *and any* a_2 *with* $\sigma_2\big(a_2|a_1\big) > 0$,

$$a_2 \in \arg\max_{a_2'} \sum_{a_1,\theta} u_2\big(a_1, a_2, \theta_1\big) p\big(\theta_1\big)\, \hat{\sigma}_{-2}\big(a_1|\theta_1\big).$$

In our game, "*Out* when Blue and *In* when Red" is a unitary type-heterogeneous SCE.

Note that because the different types do not play each other, unitary type-heterogeneous self-confirming equilibrium can alternatively be defined as a unitary self-confirming equilibrium where each type is treated as a separate player. We can relax SCE still more by not requiring all of the agents of a given type to have the same beliefs, and thus combining the heterogeneity of the previous two definitions. This *fully heterogeneous self-confirming equilibrium* allows different agents in the role of the same type to play differently and have different self-confirming beliefs, so it is a heterogeneous self-confirming equilibrium of the game where each type is a separate player. In our signaling game, this allows some Blue types to play *In* while others play *Out*, which is what happened in our experiment.

REFERENCES

Acuña, Daniel, and Paul Schrater. 2008. "Bayesian Modeling of Human Sequential Decision-Making Multi-armed Bandit Problem." https://core.ac.uk/download/pdf/22874996.pdf.

Ali, S. Nageeb. 2011. "Learning Self-Control." *Quarterly Journal of Economics* 126 (2): 857–93.

Anderson, Christopher M. 2001. "Behavioral Models of Strategies in Multi-armed Bandit Problems." PhD diss. California Institute of Technology.

Anderson, Christopher M. 2012. "Ambiguity Aversion in Multi-armed Bandit Problems." *Theory and Decision* 72 (1): 15–33.

Araujo, Felipe A., Stephanie W. Wang, and Alistair J. Wilson. 2018. "The Times Are A-Changing: Dynamic Adverse Selection in the Laboratory." http://www.pitt.edu/~swwang/papers/DynCurse.pdf.

Banks, Jeffrey, Colin Camerer, and David Porter. 1994. "An Experimental Analysis of Nash Refinements in Signaling Games." *Games and Economic Behavior* 6 (1): 1–31.

Banks, Jeffrey, Mark Olson, and David Porter. 1997. "An Experimental Analysis of the Bandit Problem." *Economic Theory* 10 (1): 55–77.

Biele, Guido, Ido Erev, and Eyal Ert. 2009. "Learning, Risk Attitude and Hot Stoves in Restless Bandit Problems." *Journal of Mathematical Psychology* 53 (3): 155–67.

Binmore, Ken, Avner Shaked, and John Sutton. 1985. "Testing Noncooperative Bargaining Theory: A Preliminary Study." *American Economic Review* 75 (5): 1178–80.

Boyce, John R., David M. Bruner, and Michael McKee. 2016. "Strategic Experimentation in the Lab." *Managerial and Decision Economics* 37 (6): 375–91.

Brandts, Jordi, and Charles A. Holt. 1992. "An Experimental Test of Equilibrium Dominance in Signaling Games." *American Economic Review* 82 (5): 1350–65.

Brandts, Jordi, and Charles A. Holt. 1993. "Adjustment Patterns and Equilibrium Selection in Experimental Signaling Games." *International Journal of Game Theory* 22 (3): 279–302.

Brenner, Thomas, and Nicolaas J. Vriend. 2006. "On the Behavior of Proposers in Ultimatum Games." *Journal of Economic Behavior and Organization* 61 (4): 617–31.

Cadsby, Charles B., Murray Frank, and Vojislav Maksimovic. 1990. "Pooling, Separating, and Semiseparating Equilibria in Financial Markets: Some Experimental Evidence." *Review of Financial Studies* 3 (3): 315–42.

Cadsby, Charles B., Murray Frank, and Vojislav Maksimovic. 1998. "Equilibrium Dominance in Experimental Financial Markets." *Review of Financial Studies* 11 (1): 189–232.

Cai, Hongbin, and Joseph Tao-Yi Wang. 2006. "Overcommunication in Strategic Information Transmission Games." *Games and Economic Behavior* 56 (1): 7–36.

Camerer, Colin, and Teck-Hua Ho. 1999. "Experience-Weighted Attraction Learning in Normal Form Games." *Econometrica* 67 (4): 827–74.

Çelen, Boğaçhan, and Shachar Kariv. 2004. "Distinguishing Informational Cascades from Herd Behavior in the Laboratory." *American Economic Review* 94 (3): 484–98.

Chen, Yan. 2003. "An Experimental Study of Serial and Average Cost Pricing Mechanisms." *Journal of Public Economics* 87 (9–10): 2305–35.

Cheung, Yin-Wong, and Daniel Friedman. 1997. "Individual Learning in Normal Form Games: Some Laboratory Results." *Games and Economic Behavior* 19 (1): 46–76.

Che, Yeon-Koo, Syngjoo Choi, and Jinwoo Kim. 2017. "An Experimental Study of Sponsored-Search Auctions." *Games and Economic Behavior* 102: 20–43.

Cho, In-Koo, Noah Williams, and Thomas J. Sargent. 2002. "Escaping Nash Inflation." *Review of Economic Studies* 69 (1): 1–40.

Cooper, David J., Susan Garvin, and John H. Kagel. 1997a. "Adaptive Learning vs. Equilibrium Refinements in an Entry Limit Pricing Game." *Economic Journal* 107 (442): 553–75.

Cooper, David J., Susan Garvin, and John H. Kagel. 1997b. "Signalling and Adaptive Learning in an Entry Limit Pricing Game." *RAND Journal of Economics* 28 (4): 662–83.

Cooper, David J., and John H. Kagel. 2003. "The Impact of Meaningful Context on Strategic Play in Signaling Games." *Journal of Economic Behavior and Organization* 50 (3): 311–37.

Cox, James C., and Ronald L. Oaxaca. 2000. "Good News and Bad News: Search from Unknown Wage Offer Distributions." *Experimental Economics* 2 (3): 197–225.

Cox, James C., Jason Shachat, and Mark Walker. 2001. "An Experiment to Evaluate Bayesian Learning of Nash Equilibrium Play." *Games and Economic Behavior* 34 (1): 11–33.

de Haan, Thomas, Theo Offerman, and Randolph Sloof. 2011. "Noisy Signaling: Theory and Experiment." *Games and Economic Behavior* 73 (2): 402–28.

Dekel, Eddie, Drew Fudenberg, and David K. Levine. 2004. "Learning to Play Bayesian Games." *Games and Economic Behavior* 46 (2): 282–303.

Esponda, Ignacio. 2008a. "Behavioral Equilibrium in Economies with Adverse Selection." *American Economic Review* 98 (4): 1269–91.

Esponda, Ignacio. 2008b. "Information Feedback in First Price Auctions." *RAND Journal of Economics* 39 (2): 491–508.

Esponda, Ignacio, and Emanuel Vespa. 2014. "Hypothetical Thinking and Information Extraction in the Laboratory." *American Economic Journal: Microeconomics* 6 (4): 180–202.

Esponda, Ignacio, and Emanuel Vespa. 2018. "Endogenous Sample Selection: A Laboratory Study." *Quantitative Economics* 9 (1): 183–216.

Feltovich, Nick, and Sobei H. Oda. 2014. "Effect of Matching Mechanism on Learning in Games Played under Limited Information." *Pacific Economic Review* 19 (3): 260–77.

Fischbacher, Urs. 2007. "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." *Experimental Economics* 10 (2): 171–78.

Fudenberg, Drew, and Kevin He. 2019. "Player-Compatible Equilibrium." http://economics.mit.edu/files/14570.

Fudenberg, Drew, and David M. Kreps. 1995. "Learning in Extensive-Form Games I. Self-Confirming Equilibria." *Games and Economic Behavior* 8 (1): 20–55.

Fudenberg, Drew, and David K. Levine. 1993a. "Self-Confirming Equilibrium." *Econometrica* 61 (2): 523–45.

Fudenberg, Drew, and David K. Levine. 1993b. "Steady State Learning and Nash Equilibrium." *Econometrica* 61 (3): 547–73.

Fudenberg, Drew, and David K. Levine. 1997. "Measuring Players' Losses in Experimental Games." *Quarterly Journal of Economics* 112 (2): 507–36.

Fudenberg, Drew, and David K. Levine. 2005. "Learning and Belief-Based Trade." *Latin American Journal of Economics* 42 (126): 199–208.

Fudenberg, Drew, and Alexander Peysakhovich. 2014. "Recency, Records and Recaps: Learning and Non-equilibrium Behavior in a Simple Decision Problem." *Proceedings of the 15th ACM Conference on Economics and Computation* 971–86.

Fudenberg, Drew, and Emanuel Vespa. 2019. "Learning Theory and Heterogeneous Play in a Signaling-Game Experiment: Dataset." *American Economic Journal: Microeconomics*. https://doi.org/10.1257/mic.20180317.

Fukuda, Emiko, Yoshio Kamijo, Ai Takeuchi, Michiharu Masui, and Yukihiko Funaki. 2013. "Theoretical and Experimental Investigations of the Performance of Keyword Auction Mechanisms." *RAND Journal of Economics* 44 (3): 438–61.

Gans, Noah, George Knox, and Rachel Croson. 2007. "Simple Models of Discrete Choice and Their Performance in Bandit Experiments." *Manufacturing and Service Operations Management* 9 (4): 383–408.

Guarnaschelli, Serena, Richard D. McKelvey, and Thomas R. Palfrey. 2000. "An Experimental Study of Jury Decision Rules." *American Political Science Review* 94 (2): 407–23.

Güth, Werner, Rolf Schmittberger, and Bernd Schwarze. 1982. "An Experimental Analysis of Ultimatum Bargaining." *Journal of Economic Behavior and Organization* 3 (4): 367–88.

Holt, Charles A., and Susan K. Laury. 2002. "Risk Aversion and Incentive Effects." *American Economic Review* 92 (5): 1644–55.

Huck, Steffen, Johannes Leutgeb, and Ryan Oprea. 2017. "Payoff Information Hampers the Evolution of Cooperation." *Nature Communications* 8.

Huck, Steffen, Hans-Theo Normann, and Jorg Oechssler. 1999. "Learning in Cournot Oligopoly—An Experiment." *Economic Journal* 109 (454): 80–95.

Hu, Yingyao, Yutaka Kayaba, and Matthew Shum. 2013. "Nonparametric Learning Rules from Bandit Experiments: The Eyes Have It!" *Games and Economic Behavior* 81: 215–31.

Kagel, John. 1995. "Auctions: A Survey of Experimental Research." In *The Handbook of Experimental Economics*, edited by John H. Kagel and Alvin E. Roth. Princeton: Princeton University Press.

Kagel, John, and Dan Levin. 2017. "Auctions: A Survey of Experimental Research." In *The Handbook of Experimental Economics*, Vol. 2, edited by John H. Kagel and Alvin E. Roth. Princeton: Princeton University Press.

Kalai, Ehud, and Ehud Lehrer. 1993. "Rational Learning Leads to Nash Equilibrium." *Econometrica* 61 (5): 1019–45.

McLaughlin, Kevin, and Daniel Friedman. 2016. "Online Ad Auctions: An Experiment." https://www.chapman.edu/research/institutes-and-centers/economic-science-institute/_files/WorkingPapers/McLaughlin-Friedman-Ad-Auction.pdf.

Meyer, Robert J., and Yong Shi. 1995. "Sequential Choice under Ambiguity: Intuitive Solutions to the Armed-Bandit Problem." *Management Science* 41 (5): 817–34.

Miller, Ross M., and Charles R. Plott. 1985. "Product Quality Signaling in Experimental Markets." *Econometrica* 53 (4): 837–72.

Mitropoulos, Atanasios. 2001. "Learning under Minimal Information: An Experiment on Mutual Fate Control." *Journal of Economic Psychology* 22 (4): 523–57.

Nicklisch, Andreas. 2011. "Learning Strategic Environments: An Experimental Study of Strategy Formation and Transfer." *Theory and Decision* 71 (4): 539–58.

Norton, Douglas A., and R. Mark Isaac. 2012. "Experts with a Conflict of Interest: A Source of Ambiguity?" *Experimental Economics* 15 (2): 260–77.

Oechssler, Jörg, and Burkhard Schipper. 2003. "Can You Guess the Game You Are Playing?" *Games and Economic Behavior* 43 (1): 137–52.

Ponti, Giovanni, Marcello Sartarelli, Iryna Sikora, and Vita Zhukova. 2018. "Gaining Experience as Principal or Agent. An Experimental Study." Working Paper WP-AD 2018-08.

Potters, Jan, and Frans van Winden. 1996. "Comparative Statics of a Signaling Game: An Experimental Study." *International Journal of Game Theory* 25 (3): 329–53.

Prasnikar, Vesna, and Alvin E. Roth. 1992. "Considerations of Fairness and Strategy: Experimental Data from Sequential Games." *Quarterly Journal of Economics* 107 (3): 865–88.

Rassenti, Stephen, Stanley S. Reynolds, Vernon L. Smith, and Ferenc Szidarovszky. 2000. "Adaptation and Convergence of Behavior in Repeated Experimental Cournot Games." *Journal of Economic Behavior and Organization* 41 (2): 117–46.

Roth, Alvin E., and Ido Erev. 1995. "Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term." *Games and Economic Behavior* 8 (1): 164–212.

Sargent, Thomas J. 2001. *The Conquest of American Inflation*. Princeton: Princeton University Press.

Shachat, Jason, and Mark Walker. 2004. "Unobserved Heterogeneity and Equilibrium: An Experimental Study of Bayesian and Adaptive Learning in Normal Form Games." *Journal of Economic Theory* 114 (2): 280–309.

Smith, Vernon L. 1962. "An Experimental Study of Competitive Market Behavior." *Journal of Political Economy* 70 (2): 111–37.

Stahl, Dale O. 2000. "Rule Learning in Symmetric Normal-Form Games: Theory and Evidence." *Games and Economic Behavior* 32 (1): 105–38.

Vespa, Emanuel, and Alistair J. Wilson. 2016. "Communication with Multiple Senders: An Experiment." *Quantitative Economics* 7 (1): 1–36.

This article has been cited by:

1. Boyuan Wei, Sander Claeys, Hamada Almasalma, Geert Deconinck. 2021. Stochastic distributed optimization of shapeable energy resources in low voltage distribution networks under limited communications. *International Journal of Energy Research* **45**:1, 991-1006. [Crossref]
2. Qichao Shi, James A. Dearden, Ernest Lai. 2021. Shuffling as a Sales Tactic: An Experimental Study of Selling Expert Advice. *SSRN Electronic Journal* . [Crossref]