Applying "Theory of Mind": Theory and Experiments*

Erik O. Kimbrough

Nikolaus Robalino

Arthur J. Robson

October 7, 2017

Abstract. This paper investigates our capacity to attribute preferences to others. This ability is intrinsic to game theory, and is a central component of "Theory of Mind", perhaps the capstone of social cognition. In particular, this component of theory of mind allows individuals to learn more rapidly in strategic environments with an element of novelty. We show here that the capacity to attribute preferences yields a clear advantage over less sophisticated approaches to strategic interaction (such as reinforcement learning) because it allows agents to extrapolate to novel circumstances information about counterparts' preferences that was learned previously. We report experiments investigating this capacity in simple extensive form games. We find significant learning of others' preferences, providing evidence for the presence and effectiveness of this aspect of theory of mind. Moreover, scores on survey measures of autism-spectrum tendencies are significant determinants of individual learning, so our notion of theory of mind is related to the notion as it is understood in psychology.

---

*Affiliation and mailing addresses—Kimbrough: Smith Institute for Political Economy and Philosophy, Chapman University, One University Drive, Orange, CA, USA; Robalino: Rochester Institute of Technology, 92 Lomb Memorial Drive, Rochester, NY, USA; Robson: Department of Economics, Simon Fraser University, 8888 University Drive, Burnaby, BC, Canada. Email addresses—ekimbrough@gmail.com, ndrobalin@gmail.com, robson@sfu.ca, respectively.
We thank Bertille Antoine, Glenn Ellison, Javier Espinosa, Dean Foster, Chris Frith, Ben Greiner, Leanna Mitchell, Chris Muris, Krishna Pendakur, Alex Peysakhovich, and Ricky Vohra for helpful discussions. The paper also benefited from comments by the editor and three anonymous referees as well as participants in seminars at CERGE-EI, RIT, and the University of Arkansas as well as conference attendees at the Economic Science Association North American Conference in Santa Cruz, CA, and the Formation of Beliefs Workshop at the Institute for Mathematical Behavioral Sciences, Irvine, CA. Kimbrough and Robson acknowledge support from the SSHRC SFU Small Grants Program; Kimbrough thanks the SSHRC Insight Development Grants Program; Robalino and Robson thank the Human Evolutionary Studies Program at SFU; Robson thanks the Guggenheim Foundation, the Canada Research Chair Program and the SSHRC Insight Grants Program. Some of the analysis was performed using the open source statistical software R (R Development Core Team, 2013).

1. Introduction

An individual with *theory of mind* conceives of himself, and others, as having agency, and so attributes to himself and others mental states such as belief, desire, knowledge, and intent. Psychologists generally accept that human beings beyond early infancy possess theory of mind.[1] Further, conventional game theory assumes, without much apology, that agents have the relevant components of theory of mind. In this paper we investigate a central relevant component—the ability to impute preferences to others. Following Robalino and Robson (2016), we refer to this ability as *theory of preferences*.

Robalino and Robson (2016) develop an evolutionary model to explain *why* theory of preferences might have evolved in strategic interactions—that is, why natural selection would favor it over alternative, less sophisticated, modes of learning such as reinforcement learning. Their argument is that the ability to impute preferences to others delivers a stark advantage in strategic environments with a key element of novelty. Theory of preferences allows an agent to extrapolate to novel circumstances information about other agents' preferences that was learned in prior situations. It thus results in the ability to anticipate and best respond to others' strategies in unfamiliar games.

In this paper, we design an experiment to test a subject's ability to learn others' preferences and to exploit what is learned strategically. The design is based on a simple theoretical model developed in the present paper that delivers sharp results concerning the rate of learning of agents with theory of preferences. The advantage of theory of preferences in our theoretical model corresponds closely to the evolutionary advantage established for this capacity in Robalino and Robson (2016).[2] In particular, theory of preferences allows an agent to more efficiently learn in a strategic environment.

Consider the experimental setup in greater detail. There is a fixed tree representing any two-stage game of perfect information with two choices at each decision node. In each session, each of an even number of subjects is assigned

[1] A key experiment here is the "Sally-Ann" test in Baron-Cohen, Leslie, and Frith (1985), who use a verbal technique to establish a crucial age of just over four. Onishi and Baillargeon (2005) use a non-verbal technique to push the crucial age back to 15 months.

[2] Our model is simpler than the evolutionary model of Robalino and Robson (2016), which permits us to make a more specific claim here about the learning rate of the *ToP** players. This simpler model also generates a tractable experimental framework more directly.

one of two roles, "Player 1" or "Player 2", with one role for each stage. All the players in a given role have the same (induced) preferences over a fixed set of outcomes, know that the players in the other role all have the same preferences, but do not know what these preferences might be.

Subjects in each session face the following situation iterated a fixed number of times. In each iteration Player 1s and Player 2s are paired randomly and anonymously. A stage game for the iteration is then determined by drawing four outcomes at random from the fixed set of outcomes and appending these to the end nodes of the two-stage game tree. Each pair of subjects then plays this sequential game, with Player 1s moving first, and then the Player 2s.

All the Player 1s see the complete history of the games played—the sequence of realized games and all the choices made by players in both roles. Our focus is then on the manner and the degree to which this historical information is exploited by the Player 1s.

We consider two categories of strategy for Player 1s—naive, and "theory of preferences". A naive strategy is inspired by reinforcement learning, as implicit in evolutionary game theory. Such a strategy of Player 1 entails approaching each game, when it is new, as a distinct and unfamiliar set of circumstances. In particular, learning the optimal response in a specific game by the naive strategy requires direct prior exposure to the game.

A theory of preferences strategy of Player 1, on the other hand, endeavors to learn the Player 2s' preferences. This strategy applies information from history in order to build up a detailed picture of these preferences, which can then be applied to any game, whether or not it has arisen historically.

Two types of theory of preference strategies are further considered. The first type, which we call a *ToP* type, fails to incorporate the transitivity of Player 2's preferences (fails to fully attribute rationality to Player 2, that is). This strategy pieces together the other player's preferences using only her observed pairwise choices. The second type of theory of preferences strategy, which we call a *ToP** type, incorporates the transitivity of the counterpart's preference ordering. This more sophisticated strategy is to learn Player 2's preferences by exploiting fully the revealed preference implications of rationality.

Player 1s with *ToP* and *ToP** infer Player 2 preferences from Player 2 choices and exploit this knowledge as in the following example. Suppose in some period prior to t Player 1s observed a subgame in which Player 2s chose the outcome z when y was attainable, and that in some other period prior to t, Player 1s also observed a subgame in which Player 2s chose y when x was attainable. A Player 1 with *ToP* then knows in period t that Player 2s prefer z to y and y to x. A *ToP** Player 1 also knows this, but will know, by applying transitive reasoning, that Player 2s prefer z to x. If the 2s had not been observed choosing in a subgame involving z and x, then at date t the *ToP** Player 1s will have an advantage over the *ToP* players. In particular, the *ToP** will correctly anticipate Player 2 choice in any subgame involving z and x, while there is room for error in the *ToP* Player 1's prediction in such a subgame.

In our experiments, the ability of Player 1 subjects to choose actions consistent with subgame perfect equilibrium outcomes in the random sequence of stage games reflects the extent to which they make use of historical data on Player 2 choices. Our key theoretical result here (Proposition 1 below) is that the rate of learning of theory of preference strategies outstrips the rate of learning of the naive strategies, while theory of preferences strategies that incorporate transitivity learn the fastest. Given the demonstrable advantages to *ToP* and *ToP** strategies, it is natural to ask whether and to what extent people actually employ such strategies in strategic environments. Our experimental design allows us to measure the degree to which an individual in the role of Player 1 utilizes various kinds of available information about Player 2s' preferences, as reflected in past choices. This approach allows us to measure the extent to which subjects use a theory of preference strategy, or a naive one.

We also assess the degree to which our notion of theory of preference corresponds to theory of mind as it is understood by psychologists. Specifically, at the end of each experimental session, we collect two measures of theory of mind that are commonly used in psychology and observe how they correlate with behavior in our experiment. In particular, we ask subjects to complete two short Likert scale surveys measuring the extent to which they exhibit autism spectrum behaviors. One was the Autism-Spectrum Quotient (AQ) survey due to Baron-Cohen et al. (2001); the other was the Broad Autism Phenotype Questionnaire (BAP), due to Hurley et al. (2007). Each test is composed of a number of subscales that rate lack of interest in and enjoyment of social interaction (Social Skills in the AQ, and Aloofness in the BAP), preferences for routine activities and difficulty switching attention (Attention-Switching in the AQ, Rigidity in the BAP), and other behaviors known to be more common among autistic individuals. We can then relate subjects' responses to these survey measures to our behavioral measures of naive and theory of preference learning.

There were two striking results of the experiments. First, we observed significant learning of Player 2 preferences by Player 1s, above and beyond that which could be explained via reinforcement learning. That is, support for theory of preference learning is expressed in real-world behavior. Individuals behave as if they ascribe preferences to counterparts and endeavor to learn these, given that it is advantageous, though subjects seem to overweight the information contained in their personal history, ignoring the information that could be gleaned from public information on behavior in other player pairs. Not surprisingly, this ability is present in real-world individuals to varying degrees. Second, we find evidence that our notion of theory of preferences is an aspect of theory of mind, as this term is understood in psychology: Player 1s who report fewer autism-spectrum behaviors (i.e. have lower AQ and BAP scores) are better able to learn Player 2 preferences.

1.1. **Related Literature**

How do our theory and experiments relate to previous work on bounded strategic rationality? A notion of bounded rationality in games that is attractive theoretically is "level-k" behavior. That is, an individual can account for a chain of iterated beliefs that is up to length k. Mohlin (2012) is a recent theoretical model of this, which shows that types of different levels can coexist in equilibrium. That is, a limited depth of reasoning may be optimal given a cost to greater sophistication.

Crawford and Iriberri (2007) present experiments on "hide-and-seek" games—zero-sum games analogous to matching pennies. In "non-neutral" settings, where framing matters, level-k behavior accounts well for observed behavior. On the other hand, Georganas, Healy and Weber (2015) find that the estimated level of reasoning, k, is not stable from one family of games to another and that direct tests of strategic reasoning are not predictive of strategic behavior.

A prominent strand of relevant literature compares the econometric performance of various models of low-rationality learning in games. A key paper here is Erev and Roth (1998) who show that a simple 1-parameter model of reinforcement learning outperforms the assumption that Nash equilibrium is instantly attained. Performance is enhanced if the 1-parameter model is augmented to allow for forgetting and for experimentation. In addition, allowing for behavior to be based as well on beliefs about opponents' choices, as in fictitious play, further enhances performance.

Camerer, Ho, and Chong (2002) further develop the class of low rationality learning models. The basic experience-weighted attraction (EWA) model hybridizes reinforcement learning, on the one hand, and belief learning akin to fictitious play, on the other. This basic model is augmented by the addition of sophisticated learning and strategic teaching. Sophisticated players understand that their opponent may be adaptive and take this into account in choosing an optimal strategy, while also appreciating the presence of other sophisticated agents. Nash equilibrium is obtained in the special case where there are only sophisticated players. In addition, the sophisticated players teach adaptive players to an extent that is governed by an effective discount parameter.

There is a substantial literature on how individuals categorize games so as to apply behavior that was learned in one game in another that is in some sense similar to the first. An elegant theoretical treatment of this is Steiner and Stewart (2008). They consider a simple 1-parameter class of games, where extreme values of the parameter lead to a unique Nash equilibrium, but where there are two pure strategy equilibria for an intermediate range of values. Given even a small contagion effect across adjacent games, one of the two equilibria is selected in the intermediate range, with a threshold marking the transition from one equilibrium to the other.

A recent paper on categorization of games is Huck, Jehiel, and Rutter (2011). To test the notion of an analogy-based expectation equilibrium (ABEE), as formulated by Jehiel (2005), they examine experimental games where players are given varying information about the history of play. The information available in principle is also presented in a more or less accessible fashion. They find that play converges to Nash equilibrium if the information is extensive and accessible, but may converge to an ABEE otherwise.

Another recent paper that examines learning spillovers in games is Grimm and Mengel (2012). Games are drawn randomly from a finite set, a feature that is shared with the present paper. If the number of possible games is small, or, if extensive information about previous play is provided, convergence to the unique Nash equilibrium occurs. Otherwise, convergence to a non-Nash distribution of actions may arise. Spillovers across games arise in complex settings.

A more basic question is whether subjects are even aware that their decisions are strategic, in the sense that their decision should be made to depend on beliefs about the decisions made by others. Fehr and Huck (2016) explore such "strategic awareness" in the context of the well-known beauty contest and show that not all subjects are strategically aware and that the extent of strategic awareness is related to individual cognitive ability.

Most relevantly for the present purpose, is a literature that is motivated by considering the effect of not knowing an opponent's preferences. McCabe, Rassenti, and Smith (1998) test the tradeoff between game-theoretic equilibrium outcomes and the attainment of efficient cooperation. They find that cooperation is enhanced by "complete information", whereas private information is conducive to Nash equilibrium. This is consistent with the present approach, where non-myopic incentives have been deliberately attenuated.

In a similar vein, Oechssler and Schipper (2003) consider the implications of not knowing the payoffs of an opponent in a fixed game. They ingeniously estimate the beliefs that a player has acquired concerning these payoffs. They find that games often attain an equilibrium of the derived subjective game, but that this need not coincide with the true equilibrium.

Another recent paper on the consequences of low-information environments is Friedman, Huck, Oprea, and Weidenholzer (2015). Players (oligopolists), see only realized quantities and payoffs rather than the complete payoff functions. This provides less information than does our approach concerning own payoffs, but more concerning opponents' payoffs. Initially, players seem to emulate the most successful player, driving quantities up and profits down to low levels. Eventually, however, quantities fall back down, beyond even the Cournot equilibrium, achieving a high degree of effective collusion.

Finally, Nax, Burton-Chellew, West, and Young (2016) consider agents are not merely ignorant of the other players' payoffs, as is true in the present paper, but also are ignorant of other players' actions, in contrast to the present paper. This full "black box" setting would hinder "theory of preferences" learning. They test for several characteristics of learning based only on own payoffs. These are shown to hold not only in the full black box setting, but also, to some extent, in settings with standard informational flows.

How do our model and our experiments relate to the previous literature? We introduce a new form of sophisticated play. This sophistication involves making inferences about another player's preferences from observing the history of play. For simplicity, we consider a game that facilitates such deductions, and where ignorance of an opponent's preferences is the only departure from the standard model. Experiments suggest that this new form of sophistication is empirically relevant and therefore merits further investigation.

Our theory allows for very general adaptive learning about an opponent's preferences, on the part of the naive and *ToP* player 1's. The crucial feature imposed on such adaptive learning here is only that it be initialized afresh if the game is new, for the naive types, or if the subgame is new, for the *ToP* types. (There is no corresponding restriction needed on the behavior of the *ToP** types.) There are assumed to be no spillovers from games or subgames that have arisen before, that is. However, adaptive learners can condition in a completely arbitrary way on their own payoffs that they do observe in a new game or new subgame. This initialization aspect of adaptive learning is reasonable but it is not very apparent in the previous literature since, with the few exceptions noted above, this literature considers a single fixed game.

Furthermore, our model does not restrict in any way the speed of adaptive learning. More specifically, we allow adaptation to be complete with only *one* previous exposure to the game or subgame. This may be simply impossibly fast, in general. Even if adaptive learning is given this edge, the theory shows that the *ToP* types will beat the naive types in a convincing way (in the sense of learning faster) and that the *ToP** types will, in turn, beat the *ToP* types just as convincingly.

The empirical issue that we then address with our experiments is an indirect test of the theory, namely: Are theory of mind spillovers from previous plays observable in human subjects? Although subjects with the theory of mind sophistication we describe should learn faster on that account, our current goal is not to enter the sweepstakes to find the best performing model of learning in games. Instead, perhaps, the best performing model of learning in games should ultimately be adapted to take account of such theory of mind spillovers, in relevant situations.

All of our types of players should learn from a previous occurrence of the game currently faced, and so we test whether this is true in the data. More importantly for our purposes, the theory implies that both *ToP* and *ToP** types, but not naive types, should generate spillovers for current subgames that were seen before, even when the game itself is unfamiliar. We test then whether previous exposure to current subgames improves choice. Finally, *ToP** types, but not naive or *ToP* types, should exhibit improved choice when previous observations mean the game can be solved, but only using the transitivity of the opponent's preferences. We test then whether such indirect use of transitivity improves current choice.[3]

We have rhetorically drawn a line between naive and sophisticated learners that falls between players who only condition on the entire game and *ToP* types who condition on subgames. This line might, instead, be drawn between the *ToP* types who exploit only direct preference revelations and the *ToP** types, who reason about an opponent's preferences using transitivity. That is, conditioning on the subgames could be viewed as an extended version of conditioning on the entire game. However, although location of this line is, to some extent, arbitrary, the *ToP* types do engage in limited "theory of mind" deductions about what to do given observations of an opponent—"I have seen how my opponent chooses in both subgames. Given that, I see the payoffs I would get from each choice and hence how I should choose". This contrasts with our naive types. Although the experiments drew attention to the presence of an opponent, the description of these types in the theory allows them to be unaware they are playing a game at all. Even using the more restrictive definition of sophistication, assigning that label to *ToP** types alone, our experiments provide tantalizing evidence that

[3] We do not suppose that actual subjects must be one of the three pure types we describe in the theory. An individual might be capable of *ToP* usually and *ToP** occasionally, for example. It might even be that the apparent sophistication of a player increases with repetition.

some individuals are sophisticated and that this sophistication is related to the degree of "theory of mind" as it is understood in psychology.

2. Theoretical Considerations

We now present a simple theoretical model that illustrates the key advantage of theory of preference and serves as the basis of our experiments. We begin with two players, $k = 1, 2$, and a finite set of outcomes, Z. Player $k = 1, 2$ has a utility function u_k defined over the outcomes, where u_k is one-to-one. Each agent knows his own utility function, but own preferences are private information.

There is a fixed perfect information extensive game form with two stages, and a set of actions A available at each decision node, with $|A| = 2$. The player role corresponding to stage $k = 1, 2$ of this game tree is always taken by Player k.

Given the fixed game tree, and the fixed order of moves of the players, a game in our environment can be identified with a 4-tuple of distinct outcomes from Z. The set of games available throughout is then G which consists of all such games with four distinct outcomes from Z.

There are two periods, $n = 0, 1$, as follows. The initial period, $n = 0$, consists of a *learning phase*. In the learning phase Player 1 is informed of the one-shot subgame perfect equilibrium strategy of Player 2 in $M \in \mathbb{N}$ games drawn uniformly and independently, with replacement, from G. More precisely, the information revealed to Player 1 in the learning phase is $h = ((g_1, \mu_1), \ldots, (g_M, \mu_M))$, where $\mu_m \in A \times A$ is the subgame perfect strategy for Player 2 in the game $g_m \in G$, for each $m = 1, \ldots, M$.[4] Each of these M games, paired with the corresponding subgame perfect strategy of Player 2 for the game, is referred to as an *example*.

After the learning phase follows the *playing period*, $n = 1$. In the playing period a game, g, is drawn uniformly from G. This game is referred to as the "terminal" game. Players 1 and 2 are informed of g and then play the game with Player 1 moving first followed by Player 2.

[4] Given that each u_k is one-to-one, and outcomes are distinct, each game in G has a unique subgame perfect equilibrium.

The game ends after the play of the terminal game, with payoffs determined for the players by their choices in the game. Figure 1 below is a schematic representation of the key aspects of the model described thus far.

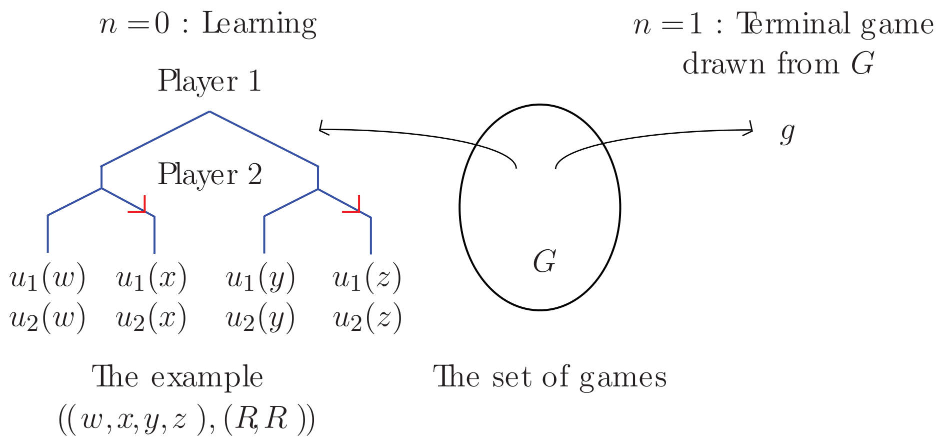

FIGURE 1: During the learning phase Player 1 is exposed to M examples drawn uniformly, with replacement, from G. An example consists a game from G, and the one-shot subgame perfect equilibrium strategy of Player 2 in the game. The figure presents the game identified with the 4-tuple of outcomes $(w, x, y, z) \in G$, paired with the strategy (R, R) of Player 2, as a particular case of an example. In the playing period the players face any terminal game g drawn from G.

2.1. **Strategies**

We finesse the issue of Player 2's rationality, by assuming that the examples provided in each history involve subgame perfect choices by Player 2. There is substantial latitude, however, in the behavior of Player 1, as he may condition his choice on the information obtained in the initial period, as well as on the terminal game. The strategy of Player 1 is denoted σ, where Player 1 with strategy σ makes the choice $\sigma(h, g) \in A$ in g, given the information h obtained during the learning phase.

Our focus is on two categories of Player 1 strategies: naive strategies, and theory of preferences strategies. Consider first the naive strategies—

DEFINITION 1: *A naive strategy of Player 1 maps own observed payoffs in each game g to an arbitrary pure choice whenever the game did not arise as an example in the learning phase. If g has arisen as an example along h, the naive strategy chooses a best response against the strategy of Player 2 that is revealed for the game in the learning phase.*

A player with a naive strategy approaches each game as a unique indivisible circumstance, that is, he fails to integrate information across games. Such a player might come to learn the optimal choice for the terminal game, but in general this will require direct exposure to that specific game during the learning phase.

Example 1.— A salient naive strategy of Player 1 is to choose $a \in A$ in a novel game if the 50-50 average of own payoffs after a exceeds the 50-50 average of own payoffs after the alternative choice, $a' \neq a$. Another salient naive strategy is a *max-min* strategy that chooses, for example, $a \in A$ in a novel game if the lowest own payoff after a exceeds the lowest own payoff after the alternative choice, $a' \neq a$. Both of the above strategies also make the subgame perfect choice in the terminal game whenever this game has occurred along h as an example. Notice that both of these naive strategies entail Player 1 making the dominant choice whenever there is one available. $||$

Players with theory of preferences strategies impute preferences to other players, and can learn these preferences by observing their counterparts' strategies. Once such a Player 1 observes 2's strategy in a game, he will have gained some information about 2's preferences, and can then exploit this knowledge in an unfamiliar game.

Consider first theory of preferences strategies that attribute preferences to player 2 but that require direct exposure to a Player 2 choice in order to learn it. Given any Player 2 subgame, say that h directly reveals Player 2's choice in this subgame, if the subgame, or its rotation along the vertical axis, appears in an example along h.

Definition 2: *A ToP strategy of Player* 1 *chooses a best response in g against a strategy attributed to Player* 2 *as follows. If h directly reveals* 2*'s choice in a subgame of g, then the strategy attributed to* 2 *for g is consistent with this revealed choice. If h does not directly reveal Player* 2*'s choice in a subgame, then Player* 1 *attributes an arbitrary strategy $\mu \in \Delta(A)$ to Player* 2 *in the subgame.*

Example 2.— Consider the following strategy of Player 1 as an example of a *ToP* strategy. If a subgame of g is unfamiliar, that is, has not arisen along h, then Player 1 assumes 2 will mix 50-50 there. If the subgame is familiar then 1 proceeds as if 2's choice in the subgame will be consistent with this revealed choice. Another example of a *ToP* strategy is one in which Player 1 assumes 2 will minimize 1's payoff in any unfamiliar subgame, while making use of revealed choices in familiar subgames. ||

Notice that the *ToP* strategy does not exploit the transitivity of Player 2's preferences. That is, if Player 1 has a *ToP* strategy, he must observe all of 2's pairwise choices in order to completely learn 2's true preferences. If 1 knew, however, that 2 has transitive preferences, he could predict a binary choice of Player 2 without having observed it in the learning phase, as in the example in the Introduction. Consider then theory of preferences strategies that avail themselves of the transitivity of 2's preferences.

DEFINITION 3: *The Player* 1 *strategy* σ *is a ToP* strategy if the following holds for every* (h, g)*. The one-shot strategy* $\sigma(h, g)$ *is a best response for player* 1 *in* g *against a Player* 2 *with choices guided by a strict transitive preference relation that is consistent with the choices made by Player* 2 *in* h*.*

That is, a Player 1 with a *ToP** strategy imputes to Player 2 a strict preference ordering that fits the data from the learning phase, and then chooses a best response as if 2's choices were guided by this imputed preference ordering. When 2's choice in a subgame of g cannot be inferred from transitivity, given the realized examples, Player 1 with a *ToP** strategy might assume, for instance, that 2 will mix 50-50 or that he will minimize Player 1s' payoff.

2.2. **Learning**

Our focus is now on learning by Player 1 in light of information revealed during the learning period. Learning by Player 1 here concerns how his strategy conditions on the data from the learning phase.

DEFINITION 4: *The Player* 1 *strategy* σ *can learn* G *after* M *examples with probability at least* $1 - \varepsilon$ *if the following is true: given any one-to-one utility function* u_2 *of Player* 2 *the choice* $\sigma(h, g)$ *is a subgame perfect choice for Player* 1 *in* g *with probability no less than* $1 - \varepsilon$*.*

That is, given M examples drawn uniformly from G, a Player 1 strategy is said to learn the set of available games with high probability if the strategy is likely to make the optimal choice for Player 1 in the terminal game g, where this is true uniformly for *all* strict preference orderings of Player 2.[5]

Our main theoretical result compares rates of learning for the types of Player 1 strategies described above. The proof is given Appendix A.

PROPOSITION 1: *Suppose* $|Z| = N \geqslant 4$. *There are functions of* ε, $L(\varepsilon)$, *and* $U(\varepsilon)$, *with* $U(\varepsilon) \geqslant L(\varepsilon)$, *and* $U(\varepsilon) < \infty$ *for all* $\varepsilon > 0$, *such that the following results are true for each given* ε*:*

1) If σ *is a naive strategy of Player* 1, *then* σ *is guaranteed to learn* G *after* M *examples with probability* $1-\varepsilon$ *if* $M \geqslant N^4 \cdot U(\varepsilon)$, *and may not learn* G *after* M *examples with probability at least* $1-\varepsilon$ *if* $M < N^4 \cdot L(\varepsilon)$.

2) If σ *is a ToP strategy of Player* 1, *then* σ *is guaranteed to learn* G *after* M *examples with probability* $1-\varepsilon$ *if* $M \geqslant N^2 \cdot U(\varepsilon)$, *and may not learn* G *after* M *examples with probability at least* $1-\varepsilon$ *if* $M < N^2 \cdot L(\varepsilon)$.

3) If σ *is a ToP* strategy of Player* 1, *then* σ *is guaranteed to learn* G *after* M *examples with probability* $1-\varepsilon$, *if* $M \geqslant N \cdot U(\varepsilon)$, *and may not learn after* M *examples with probability at least* $1-\varepsilon$ *if* $M < N \cdot L(\varepsilon)$.

The *ToP* and *ToP** strategies learn faster (i.e., require a smaller number of examples) than the naive strategies, while the *ToP** strategies learn the fastest. For example, a number of examples that is linear in the number of outcomes will suffice for the *ToP** strategy to learn G up to a given probability. The number of examples, M, must have order of N^2 in order for a *ToP* strategy to learn G with the same probability, and order of N^4 for a naive strategy to learn with this same probability.

Proposition 1 thus reveals a stark advantage maintained by the *ToP** strategies over the *ToP* ones, and by the *ToP* strategies over the naive ones. For example, for each $\varepsilon > 0$ there is a number of outcomes N, and a number of examples M such that a *ToP** strategy can learn G with probability $1-\varepsilon$, but the

[5] That is, each utility function u of Player 2 induces a joint probability distribution, say Φ_u, over the examples generated in the learning phase, and the game g. The strategy σ learns G with probability no less than $1-\varepsilon$ if

$$\Phi_u \{(h,g) : \sigma(h,g) \text{ is an } \textit{SPE} \text{ choice in } g\} \geqslant 1-\varepsilon, \tag{1}$$

for each utility function u of Player 2.

ToP strategy cannot.[6] Similarly, there is a number of outcomes and a number of examples such that the *ToP* strategy can learn with a given probability but the naive strategy cannot. On the other hand, for all sufficiently large N, if a naive strategy can learn G with a given probability, then so can a *ToP* strategy, and if a *ToP* strategy can learn, then so can a *ToP** strategy.

Our proof of Proposition 1 relies on PAC (Probably Approximately Correct) learning theory, described in Kalai (2003), where the theory is applied to the related problem of learning a rational agent's preferences. (See also Vidyasagar, 1997.)

3. Experimental Design

Our experimental design is based on our theoretical model from Section 2. The theory highlights the advantages to *ToP* and *ToP** strategies, and so we test whether subjects actually use such strategies. That is, the experiments test the ability of individuals to learn the preferences of others through repeated interaction and to use that information strategically to their advantage. Our model provides the grounds for designing the empirical tests. In particular, relying on the definitions of *ToP* and *ToP** strategies, our tests examine the extent to which subjects make use of previously revealed preferences in order to make *SPE* choices in extensive form games.

Recall that in the model there is a fixed two-stage, two-action perfect information game tree, and two player roles, 1 and 2, corresponding to each stage of this tree. In each experimental session each of an even number of subjects is assigned either the role of a Player 1 or of a Player 2, where this assignment is fixed throughout the session.

There is a fixed set of N outcomes Z, and in the experiment each outcome corresponds to a pair of monetary payoffs, one for Player 1 and one for Player 2. Each outcome yields a player a *unique* monetary payoff, resulting in an induced strict preference ordering over Z. All the subjects in a given role share the same

---

[6] This is true, in particular, for N such that $N \cdot U(\varepsilon) < N^2 \cdot L(\varepsilon)$, and M satisfying

$$N \cdot U(\varepsilon) \leqslant M < N^2 \cdot L(\varepsilon)$$

induced preferences (payoffs), and know these, but do not know the preferences (payoffs) of the other role.

In each period, $t = 1, \ldots, T$, every Player 1 participant in a session is randomly and anonymously matched with a single Player 2 participant in the same session. A two-stage extensive form game is determined for the period by drawing four outcomes at random from Z *without replacement* (one for each terminal node of the fixed game tree). Every matched pair then plays this game. In each period, this process is repeated, drawing from the same set of outcomes, Z. The set of possible games that could be created in this manner is denoted G as in the theoretical model.[7]

The situation viewed from the perspective of a Player 1 is depicted in Figure C1, in Appendix C. In each game the player 1 in each matched pair of subjects moves first, choosing one of two intermediate nodes (displayed in the figure as blue circles). Her Player 2 partner then chooses one of the two remaining terminal nodes. These choices together determine payoffs for the matched participants (displayed in the figure as a pair of boxes).

When making a decision in a game, a Player 1 participant sees her own payoff at each terminal node (in the orange boxes), but her partner's payoff at each outcome is hidden in a blue box.[8] Similarly, when a Player 2 makes his choice, he observes only his own payoffs and sees a "?" in place of his partner's payoffs (see Figure C2). We describe this situation to subjects in the instructions as follows:

[7] We employ random anonymous matching in order to diminish the likelihood of supergame effects. In any case, acting on such motives involves the 2s solving a difficult inference problem. A Player 2 who wanted to punish (or reward) Player 1 on the basis of Player 2's foregone payoffs, or to alter 1's behavior in the future, first must infer that player 1 has learned Player 2's preferences and then infer player 1's own preferences on the basis of this assumption. In practice Player 2 choices are myopically optimal, given their induced preferences, about 90% of the time, and we see no evidence of an overall trend in the optimality of Player 2s choices. When subjects were asked to explain their decisions after the experiments, responses suggested that dominated choices by Player 2s were often driven by lack of understanding rather than supergame motives.

[8] Payoff privacy has the added benefit of mitigating the effects of non-standard preferences on individual choice. Since individuals are unaware of exactly how their choices impact others' payoffs the effect of relative payoffs (as in Charness and Rabin, 2002, for example), will be reduced. Indeed, it has long been known that payoff privacy encourages the achievement of equilibrium outcomes in market settings (Smith 1982).

In this experiment, you can only see the earnings in your own box. That is, if you are Person 1 you will only see the earnings in the orange boxes, and if you are Person 2 you will only see the earnings in the blue boxes. Both boxes will be visible, but the number in the other person's box will be replaced with a "?".

However, for each amount that you earn, the amount the other person earns is fixed. In other words, for each amount that Person 1 sees, there is a corresponding, unique amount that will always be shown to Person 2.

For example, suppose Person 1 sees an earnings box containing "12" in round 1. In the same pair, suppose Person 2 sees "7". Then, at any later round, anytime Person 1 sees "12", Person 2 will see "7".

As in the theoretical model, despite their initial ignorance of the other players' preferences, Player 1s who attribute preferences to their counterparts can learn them by observing the 2s' choices. That is, given that a subject knows each outcome delivers a unique monetary own payoff, she knows that each own payoff is associated with some fixed payoff to her counterpart. Hence, by observing Player 2 choices in their subgames, a Player 1 can build up a picture of the 2s' preferences over the outcomes.

That is, subjects in the role of Player 1 encounter the scenario described by the theoretical model repeatedly, with a new game drawn at random each time from G. Each randomly drawn game faced by the 1s, say g_t, $t = 1, \ldots, T$, is then analogous in the experiments to g in the theoretical model. The list of examples given to each Player 1 arise in the experiments as the history at each iteration t, which includes the observed choices of the Player 2s.

Figure C3 in the appendix shows the feedback a subject receives at the conclusion of each game. In particular, at the end of each iteration $t = 1, \ldots, T$, the Player 1s observe all of the Player 2s' choices at the nodes that are reached, for all of the Player 2s in the lab during the session. This information is conveyed by showing a small image of the game tree (with payoffs still appropriately blinded) in the upper corner of the screen. On that image, subjects can see the empirical relative frequency with which each possible path through the game tree was taken by pairs in their session in the last period. They also see their own path through the game tree and their outcome. Using their partner's play and the choices made by the other 2s, a Player 1 with theory of preferences can make inferences about the preferences of the Player 2s.

Once a Player 1 fully learns the 2s' preferences, she can respond optimally thereafter, making the subgame perfect choice in future games. This suggests we can measure the extent of theory of preferences among Player 1s by observing how the likelihood of *SPE* choice relates to the information revealed by the history of play about Player 2 preferences.[9] By observing in our experiment how the choices of a Player 1 depend on *all* the information revealed in previous games, we can observe the degree to which a subject is capable of learning from the information that *ToP* and *ToP** types have in the theoretical environment. Our measurement approach is discussed in full detail in Section 3.2.

Existing research indicates that when subjects play a single fixed game, with repeated play, fixed matching, and private information about individual payoffs, pairs of players frequently converge to non-cooperative equilibrium outcomes over time (McCabe et al., 1998).[10] An experiment involving the repetition of the same game, however, allows a player to learn the optimal response through reinforcement learning. These one-game experiments thus cannot yield a clear distinction between players with theory of preferences and those that use reinforcement learning. Our experiment is the first (that we know of) to test for theory of preferences capacity in a dynamic setting where inferences drawn by subjects' from the play of one game might later be employed to predict play in unfamiliar games. Our setting with multiple games arising in a random fashion allows us to clearly identify games where the *SPE* can be learned from history only with theory of preferences from those where a reinforcement learner can learn from the history, and to observe heterogeneity in individuals' capabilities to learn by each method.

3.1. **Treatments**

There are two interrelated practical issues that arise from generating the games by simply drawing four outcomes without replacement from the set of

[9] Player 1 subjects in the experiments observe only the choices made by their counterparts at reached nodes of the stage games. (In contrast, in the theoretical model, Player 1 is informed of the complete strategy of Player 2 for each example.) This, however, does not preclude the measurement, among our subjects, of *ToP* and *ToP** as defined in the theoretical model.

[10] See also Fouraker and Siegel (1963) and Oechssler and Schipper (2003).

N outcomes: 1) many of the randomly generated games will have dominant actions for Player 1, and such games are not useful for inferring the capacity of Player 1s to learn the preferences of others, since the dominant choice will not depend on the strategy of Player 2, and 2) more subtly, for a significant portion of games there is a simple "highest mean" rule of thumb for Player 1 that generates *SPE* play. That is, consider a Player 1 who is initially uncertain about Player 2's preferences. From the point of view of Player 1, given independence of 2's preferences, Player 2 is equally likely to choose each terminal node, given 1's choice. The expected payoff maximizing strategy for 1 is to choose the intermediate node at which the average of potential terminal payoffs is highest. Indeed, our pilot sessions suggested that many participants followed this strategy, which was relatively successful in terms of payoffs.

For these reasons, we used a 3x1 within-subjects experimental design that, over the course of an experimental session, pares down the game set to exclude games in which choice is too simple to be informative, either because Player 1 has a dominant choice, or because the "highest mean" heuristic gives the *SPE* choice for 1. Each session included games drawn from $N = 7$ outcomes (so there are 105 possible strategically distinct games).[11] Each outcome is associated with a pair of payoffs, one for each player, and these were constructed from two random permutations of a set of possible payoffs (usually integers between 1 and 7).[12] In our experiments, we conducted sessions employing 8 different sets of randomly generated payoff pairs, with at least two sessions conducted using each set.

Each session lasted for $T = 90$ periods in which, in the first 15 periods, games were drawn randomly from G, the set of all strategically distinct games constructed using outcomes from Z. Then, starting in the 16-th period, we eliminate from G all the games in which Player 1 has a dominant choice. The next 15 periods consist of games randomly drawn from this restricted set of games (NoDominant treatment). Finally, starting in the 31st period, we also

[11] There are $N \cdot (N-1) \cdot (N-2) \cdot (N-3)$ games with four distinct outcomes, but only 1/8 as many strategically distinct such games if we consider two games equivalent whenever one is derived from the other by rotation of one or more subgames and/or swapping the subgames.
[12] In eighteen of our sessions, payoff possibilities for each participant consisted of integers between 1 and 7, and in two sessions the set was $\{1, 2, 3, 4, 8, 9, 10\}$. This variation was intended to reduce noise by more strongly discouraging Player 2 from choosing a dominated option, but observed Player 2 choices in these sessions are comparable to those in other sessions, so we pool the data for analysis below.

eliminate all games in which the "highest mean" rule of thumb corresponds to the *SPE* of the game, and our final 60 periods consist of randomly drawn games from this smaller subset (NoHeuristic treatment). Thus, our final 60 periods make it less likely a Player 1 will make the *SPE* choice without either having been exposed to the game, or to his counterpart's choices in the relevant subgame(s). The subjects were not informed of the regime changes occuring at periods 16 and 31.[13] Table F1 reports the number of unique games played in each experimental session.

3.2. **Measurement**

Our theoretical model provides a framework to test for reinforcement learning as well as *ToP* and *ToP** strategies among Player 1s. On the face of it, if a Player 1 subject makes a suboptimal choice in a game such that both pairwise choices of Player 2 in the game have been observed by the subject previously, then her suboptimal choice refutes the hypothesis she uses the *ToP* or the *ToP** strategy. Similarly, if she makes a suboptimal choice in a game such that 2's pairwise choices in the game are implied by transitivity, given the observed choices of the 2s, then this is evidence against the use of a *ToP** strategy. This approach to testing for *ToP* and *ToP** among real world subjects is too stringent, however, to be practical. We therefore carry out a more relaxed test for theory of preferences, which measures the likelihood of the *SPE* choice by Player 1s given information about the 2s' preferences implied by the history of play.

[13] Notice that during the NoHeuristic treatment, when the "highest mean" heuristic yields for player 1 a definite choice, the action with the lower mean payoff corresponds to the *SPE* of the game. One concern is that the player 1 subjects might have learned to choose the action with the lowest mean payoff, when such an action was available. It is clear, however, that our subjects did not apply this alternative heuristic. Any player 1, upon discerning that the action with the lowest mean payoff is consistent with the *SPE*, would choose the *SPE* more often when the 50-50 heuristic give a definite choice, than when both actions had the same mean payoff. However, we find that during NoHeuristic subjects chose the *SPE* 69% of the times when the "highest mean" heuristic did not give a definite choice, and 48% of the times when the rule gave a definite choice. This difference, with greater success in choosing the *SPE* when the heuristic does not give a definite choice, holds also in the later periods—in periods 45-90, 60-90, and 75-90 (See Figure F2 in Appendix F).

One reason why the Player 1s might not make appropriate inferences from Player 2's choices is that the Player 2s do not always make the dominant choice.[14] We control for non-dominant choices by the 2s as follows—

Controlling for 2's Non Dominant Choices: Hereafter, we enumerate the Player 1 subjects $i = 1, \ldots, 87$. To control for the Player 2s' use of non-dominant choices we define a variable that is the proportion of dominant choices made by all the 2s in the previous periods, within a given session, j. In particular, let W_{ijt} be an indicator variable that takes the value of 1 if the Player 2 partner of subject i in session j chose the dominant action in the subgame she reached in iteration t, and 0 otherwise. At iteration $t > 1$, the proportion of Player 2 dominant choices observed by Player 1s in session j is then

$$D_{jt} = \frac{\sum_{i \in I_j} \sum_{s=1}^{t-1} W_{ijs}}{|I_j| \cdot (t-1)},$$

where I_j is the set of Player 1 subjects in the session. (Since no choices have been observed by the subjects in the first period, we have $D_{j1} = 0$.)

If Player 1s were Bayesian rational, they would treat the observations on all Player 2s as equally informative. It is plausible psychologically, however, that they pay particular attention to their partnered Player 2, especially since this partner's behavior affects the current payoff for Player 1.[15] We therefore also track the proportion of dominant choices made by the partners of subject i in session j and denote this $D_{ijt} = \frac{\sum_{s=1}^{t-1} W_{is}}{t-1}$, for each $t > 1$. Again, nothing has been observed in the first period and so $D_{ij1} = 0$.

Measuring Reinforcement Learning: In the experiments, measurement of reinforcement learning is straightforward. Since all matched pairs play the same game in a given period, we can measure reinforcement learning in regression analysis using the indicator variable K_{jt} which takes the value of 1 if the iteration

[14] To avoid this source of noise, we considered automating the role of the Player 2s. However, on reflection, this design choice seems untenable. In the instructions, we would need to explain that algorithmic Players 2 maximize their payoffs in each stage, which would finesse a key part of the inference problem faced by the Player 1s—in essence the instructions would be providing a key part of the theory of preferences. It is also conceivable that individuals would behave differently towards a computer program than they would towards a human agent.

[15] Indeed, this is consistent with evidence that individuals overweight private information; see, for example, Goeree et al., (2007).

t game arose previously in session j, and 0 otherwise.[16] When $K_{jt} = 1$, all Player 1's have been exposed to Player 2 behavior in the current game, but perhaps not the 2s' complete strategy for the game. Nonetheless we expect naive Player 1s to more often make the *SPE* choice when $K_{jt} = 1$.

Measuring *ToP* Learning: Recall that, at each iteration, each Player 1 subject observes the choice of his matched Player 2 partner, as well as the empirical distribution of choices made by all the other Player 2s in his session. Although the personal history of each Player 1 in a session might differ, each such subject will have been exposed to the same historical data concerning the choices of all Player 2s in the session. This common history will be referred to simply as the *history* for brevity.

We say that the history in a given session *directly reveals* 2's preferences over the outcomes z and z' whenever this history involves any Player 2 in the session making the optimal choice in a subgame over the outcomes z and z' and no Player 2s making non-optimal choices in a subgame over z and z'. That is, direct revelation of a pairwise choice of Player 2 is said here to occur if and only if the history involves at least one *correct* choice by any Player 2 in the session over a given pair of outcomes and no *incorrect* choices by any Player 2 in the session.[17]

We then say that a history in a session *directly implies* Player 1's *SPE* (subgame perfect equilibrium) choice in a game if a Player 1 can infer this *SPE* choice using *only* the direct revelations of Player 2 preferences along the history.

---

[16] We consider two games equivalent if one can be obtained by reflection of one or more subgames of the other. Our definition of K_{jt} is then charitable to reinforcement learning since it assumes that what is learned by a naive player in one game is correctly applied to any game obtained by reflecting one or more subgames of the original game.

[17] We impose this restriction since non-dominant choices by Player 2s mean that a fully rational Player 1 may well believe that a non-*SPE* choice is optimal. Nevertheless, in most cases, Player 1s never see any confounding evidence about Player 2 preferences. In particular, 76% of the times a Player 1's history reveals a correct binary choice of 2, say over z and z', there have been no incorrect choices by the 2s over z and z' along the history, so our definition of 'direct revelation' is not overly restrictive. Moreover, this implies that most of the times history directly reveals a pairwise preference of Player 2, a *ToP* Player 1 should, in principle, correctly infer the specific pairwise choice of Player 2.

That is, Player 1 does not need to use the transitivity of Player 2's preferences.[18]

Two key variables constructed using this notion of 'direct implication' reflect the extent of *ToP* learning in subjects. The first of these, $S_{jt} \in \{0, 1\}$, is a dummy variable that is 1 if and only if the iteration t history of session j directly implies Player 1's *SPE* choice in iteration t of the game in session j, and, further, $K_{jt} = 0$, so the game is unfamiliar. That is, S_{jt} is a conservative approach to showing that *ToP* play is present. We presume that *ToP* players will learn if $K_{jt} = 1$, but we discount this effect in defining S_{jt}. If the coefficient of S_{jt} is significant, this provides strong evidence of strategic sophistication that exceeds that of the naive types.

Player 1s should learn from the entire history of choices made by Player 2s, and this is the assumption in the model described above. Behaviorally, however, it is plausible that individual Player 1s might find the history of choices made by their actual partners more compelling evidence concerning Player 2s' preferences. To consider this, we define a dummy variable $S_{ijt} \in \{0, 1\}$ that is 1 if and only if it is possible for Player i in session j to infer his *SPE* choice at iteration t using only the direct revelations of preferences made by his matched partners, that is along his personal history. Here, direct revelation of a binary preference of Player 2 along i's personal history occurs when i has observed a partner making the correct choice over the pair of outcomes, and never observed a partner making the incorrect choice over the same outcomes. Again, we allow $S_{ijt} = 1$ only if $K_{jt} = 0$ to take a conservative approach to establishing evidence of *ToP* play.

---

[18] The *SPE* choice of 1 is implied by history if it can be inferred by backward induction using only the directly revealed choices by Player 2s. If a history directly reveals both of Player 2's binary choices in a game then it directly implies 1's *SPE* choice in the game. In some cases, however, it suffices to know only one of 2's binary choices in a game in order to infer 1's *SPE* choice. For example, if the *SPE* payoff to Player 1 in a Player 2 subgame, say L, is strictly better (worse) for Player 1 than the two outcomes in the other Player 2 subgame, then, knowing only 2's preferences over the outcomes in L, we can infer that 1's *SPE* choice is to choose into (away from) L. In other cases, however, it will be necessary for both of Player 2's pairwise choices in a game to have been directly revealed for the *SPE* choice of 1 to be directly implied. This is true when the *SPE* outcome in subgame L is neither better or worse, from 1's perspective, than the two outcomes in subgame R, and, at the same time, the *SPE* outcome in subgame R is neither better or worse, from 1's perspective, than the two outcomes in subgame L.

Measuring *ToP Learning:** As is relevant for considering *ToP** Player 1s, we say that a history in a session *indirectly reveals* 2's preferences over z and z' if no Player 2 in the session has been observed making a sub-optimal choice over these outcomes, 2's preferences over the outcomes have not been directly revealed by the history, and there is an outcome $x \in Z$ such that i's history indirectly reveals Player 2 prefers z to x, and also x to z'.[19] We then say that a history *indirectly implies* Player 1's *SPE* choice in a game if a Player 1 can only infer this *SPE* choice if he uses the indirect revelations of Player 2 preferences along the history. (That is, the notion is as in direct implication of the *SPE* choice but for indirect revelation instead.) Note that this includes cases where Player 2s' preferences in one subgame have been directly revealed but indirect revelations are necessary to determine the preferred outcome in the other subgame and thereby to solve the game.

To test for the ability of Player 1s to make inferences about Player 2 based on the transitivity of Player 2's preferences, as in *ToP**, we define two additional variables—S^*_{jt} and S^*_{ijt}—that reflect the information revealed to any Player 1 on the basis of indirect revelations of Player 2 preferences along history. The first of these variables, $S^*_{jt} \in \{0,1\}$, is a dummy variable that takes the value 1 if and only if the iteration t history of session j indirectly implies Player 1's *SPE* choice in the period t game of the session, and, moreover, the game is unfamiliar so that $K_{jt} = 0$. It also follows, by the definition of indirect revelation of preferences, that $S^*_{jt} = 1$ only if $S_{jt} = 0$. This will generate a conservative test of the significance of *ToP** sophistication over and above *ToP* sophistication.

Given previous evidence that individuals overweight private information, we define $S^*_{ijt} \in \{0,1\}$ as a dummy variable that takes the value 1 if and only if it is possible for Player 1 subject i to infer his *SPE* choice in the current game at iteration t using indirect revelations of preferences by his previous partners, and the game is unfamiliar, so that $K_{jt} = 0$. It follows that $S^*_{ijt} = 1$ only if $S_{ijt} = 0$, again leading to a conservative test of *ToP** over *ToP* .

[19] This usage is non-standard in that indirect revelation rules out direct revelation, and we do not allow chains of arbitrary length. This is done for tractability and since shorter chains are more likely to arise in the data.

Measuring Theory of Mind: We relate our experimental results directly to theory of mind as it is understood in psychology, using two short survey instruments as theory of mind measures. At the conclusion of each experimental session, the participants completed the Autism-Spectrum Quotient (AQ) survey designed by Baron-Cohen et al. (2001), since autism spectrum reflects varying degrees of inability to "read" others' minds. This short survey has been shown to correlate with clinical diagnoses of autism spectrum disorders, but it is not used for clinical purposes. The instrument was designed for use on adults of normal intelligence to identify the extent of autism spectrum behaviors in that population. Participants also completed the Broad Autism Phenotype Questionnaire (BAP) due to Hurley et al. (2007), which provides a similar measure of autism spectrum behavior and is highly correlated with the AQ. Using this survey data we evaluate how a subject's ability to learn preferences in our experiments correlates with two well-known metrics for theory of mind.[20] Copies of the questionnaires are available in Appendices D and E.[21]

3.3. Procedures

We report data from 20 experimental sessions with a total of 174 participants (87 in the role of Player k, for each $k = 1, 2$). Each experimental session consisted of 6, 8 or 10 participants, recruited from the students of Simon Fraser University between April and October 2013. Participants entered the lab and were seated at visually isolated computer terminals where they privately read self-paced instructions. A researcher was available to privately answer any questions about the instructions. After reading the instructions, if there were no additional questions, the experiment began. Subjects had a single blank sheet

[20] One might be concerned that any differences we observe in behavior that are correlated with AQ are actually driven by differences in intelligence. Indeed, it is well-known that extreme autistics tend to have low IQs. Crucially, however, within the normal range of AQ scores (for those who have not been diagnosed with an autism spectrum disorder), the survey measure is uncorrelated with intelligence (Baron-Cohen et al., 2001). Our sample consists of undergraduates none of whom (to our knowledge) had been diagnosed with any autism spectrum disorder. Thus any relationship we observe between AQ and performance is unlikely to be due to differences in intelligence.

[21] In the first wave of these experiments performed in April and June 2013 (76 subjects total), we conducted the AQ questionnaire with a 5-point Likert scale that allowed for indifference rather than the standard 4-point scale which requires participants to either agree or disagree with each statement. The AQ questionnaire is scored by assigning 1 or 0 to each response and summing the result. In our data analysis below, we assign indifferent responses a score of 0.5.

of paper on which they could record information throughout the experiment. Instructions are available in Appendix B.

Each experimental session took between 90 and 120 minutes. At the conclusion of each session, participants were paid privately an amount of cash based on their payoffs from two randomly chosen periods. Hence each game was payoff-relevant. For each chosen period, the payoff from that period was multiplied by 2 or 3 (depending on the session) and converted to CAD. Average salient experimental earnings were $25.00, with a maximum of $42.00 and a minimum of $6.00. In addition to their earnings from the two randomly chosen periods, participants also received $7 for arriving to the experiment on time. Upon receiving payment, participants were dismissed.

4. Experimental Findings

Since the decision problem is trivial for the player 2s, our analysis is of decisions by subjects in the role of Player 1. In particular, we focus on the probability that a Player 1 chooses an action consistent with the subgame perfect equilibrium of the game.

First, we describe the overall trend, and then we test whether the likelihood of choosing the *SPE* action as Player 1 depends on how much information a subject has acquired about the preferences of Player 2s. In particular, we can use each individual's history of play to construct the variables K, S and S^* as described above to determine whether subjects make use of past choices by player 2s like *ToP* and *ToP** types in the model. The results suggest that our Player 1s exhibit theory of preferences in the sense that they are more likely to choose the *SPE* action when a *ToP* or *ToP** type would do so, on average. However, the effect on average is of moderate size suggesting that not all subjects learn quickly or successfully from the implications of Player 2s' choices. We then unpack this heterogeneity across individuals, asking how the ability to infer the *SPE* action from others' past choices varies across the observed distribution of our survey measures of theory of mind from the psychology literature. In particular, we examine how the marginal effect of our variables K, S and S^* on the probability of choosing the *SPE* action varies with a subject's AQ score. Our results suggest that our theoretical concept of theory of preferences corresponds, at least to some extent, with theory of mind as understood by psychologists. For instance, we find no evidence that the probability of choosing the *SPE* action when there is a dominant action varies with AQ, but we see considerable variation across the AQ distribution in the ability to learn from indirect preference revelations (i.e. the coefficient on S^* varies considerably with AQ score).

4.1. **Learning Others' Preferences**

Figure 2 displays a time series of the fraction of Player 1s that made the *SPE* choice in each of the 90 periods of the experiment. After 15 periods, the game set no longer included games in which Player 1 has a dominant choice. After 30 periods, the game set no longer included games where Player 1 can make the *SPE* choice by following the "highest mean" rule of thumb. At period 31, when subjects enter the NoDominant/NoHeuristic treatment, there is a sizable downtick in the fraction of optimal choices by Player 1s, but afterwards there is a notable upward trend in the fraction of 1s choosing optimally.[22] Despite the fact that individuals tend to learn Player 2's preferences over time on average, we observe substantial heterogeneity, which we exploit in the next section.

To provide statistical support for these observations, Table 1 reports OLS regressions where the dependent variable takes the value of 1 if Player 1 chooses an action consistent with the *SPE* of the game and 0 otherwise. We include dummies that take a value of 1 when Player 1 had a dominant action or when following the "highest mean" rule of thumb, despite the absence of a dominant action, would lead to an *SPE*-consistent choice (0 otherwise), as well as individual fixed effects, and we cluster the standard errors at the session level. We report OLS specifications so that the coefficients give marginal effects, but our results are qualitatively unchanged if we employ logistic regression instead. We report one-tailed hypothesis tests since we expect higher values of each RHS variable to increase the likelihood of *SPE* choices by player 1s.

In order to control for the impact of feedback quality from Player 2 choices on the likelihood of *SPE* choices by player 1s, column (2) introduces a variable that controls for the proportion of dominant choices made by all session j Player 2s

[22] Table F1 in Appendix F reports summary statistics for each experimental session. Figure F1 displays a histogram of the individual rates of *SPE* consistent choices over all of the most informative games—those in which Player 1 has no dominant choice and in which the rule of thumb did not lead to the *SPE* consistent choice.

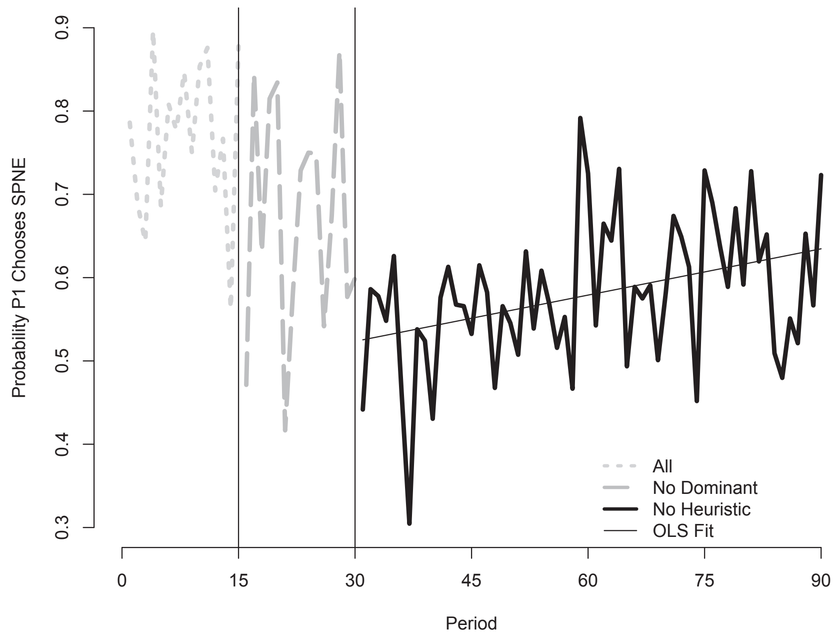

Figure 2: **Time Series of Learning Counterpart's Preferences.**

in previous periods (D_{jt}). Column (4) presents an analogous specification using only the history of choices by Player 2 partners of subject i in session j (D_{ijt}).

Columns (3) and (5) report specifications including additional variables that allow us to disentangle reinforcement learning, *ToP* and *ToP** strategies by Player 1s, as in the model. To test for naive reinforcement learning, both specifications include the variable K_{jt} which takes the value of 1 if subjects in session j have seen the game before and 0 otherwise, as described above. A positive and significant coefficient on K_{jt} would provide evidence that subjects are capable of learning from previous plays of a game, at least. To test whether subjects are capable of learning via theory of preferences, column (3) includes two additional variables—S_{jt} and S^*_{jt}—(from Section 3.2)—that reflect the information revealed to Player 1 subjects about Player 2s' preferences from the public history of choices by all player 2s in the session—whether the *SPE* is directly or indirectly implied by player 2s' choices, respectively.

To reiterate, these variables capture whether Player 1 can compute the *SPE* of the game using only direct revelations of Player 2 preferences (what we call direct implication) or a combination of direct and indirect revelations (indirect implication) of Player 2 preferences that occur along the session history. Column

(5) includes analogous variables—S_{ijt} and S^*_{ijt}—that reflect whether the *SPE* choice has been—directly or indirectly—implied by subject i's *counterpart's* choices along his personal history. Positive and significant coefficients on the S and S^* variables will provide evidence that subjects learn from the kind of information exploited in *ToP* and *ToP** strategies, respectively.

Regression Results: In Table 1, the positive and significant estimated coefficient on Period in column (1) indicates that Player 1 participants are increasingly likely to choose optimally over time. This is consistent with the evidence in Figure 2. A positive, large and significant coefficient on the dummy indicator for dominant strategy suggests that subjects generally play dominant strategies when available, and a positive and significant coefficient on the "highest mean" heuristic dummy suggests that our decision to eliminate those choices in later games was sensible. Subjects chose the *SPE* action 92% of the time when they had a dominant action and 84% of the time when the "highest mean" heuristic led to the *SPE* action. In columns (2)-(3), when we include the variable measuring the fraction of previous dominant choices by all Player 2s in a session (i.e., D_{jt}), and in columns (4)-(5), when we include the variable measuring the fraction of dominant choices by Player 2s partnered with i in session j (D_{ijt}), we find positive and significant coefficients, indicating that high quality Player 2 decision-making is important for Player 1 learning to choose the SPE.

In column (3), we find no evidence of significant naive reinforcement learning, nor any evidence that Player 1s learn from the public history of choices by Player 2s in a manner consistent with *ToP* or *ToP** as defined above. However, in column (5), when direct and indirect revelations are considered only in the *personal* history of each Player 1, we find substantial evidence of *ToP* and *ToP** strategies, and not just reinforcement learning. In particular, having seen the game before increases the probability of choosing optimally by 0.05. Moreover, a highly significant and positive coefficient on S_{ijt} indicates that Player 1s improve their performance by applying to novel games what they have learned by personal experience about the preferences of Player 2s through past direct revelations. In particular, when history directly reveals enough information about 2s' preferences so that a Player 1 with *ToP* can solve for own optimal choice, the probability the Player 1 chooses optimally increases by 0.09. We also see some evidence for learning of the kind performed by *ToP** types in the model;

a positive and marginally significant coefficient on S^*_{ijt} suggests that Player 1s also exploit the indirectly revealed preferences of their partnered Player 2s; here when the *SPE* has been indirectly implied by history (which is defined to mean *not* directly implied), the probability of a Player 1 choosing the *SPE* action increases by approximately 0.05.

P1 Chose SPE	(1)	(2)	(3)	(4)	(5)
Period	0.002***	0.002***	0.002***	0.002***	0.002***
	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)
Has Dominant Strategy	0.463***	0.484***	0.482***	0.481***	0.483***
	(0.024)	(0.025)	(0.026)	(0.025)	(0.024)
"Highest Mean" Heuristic = SPE	0.370***	0.376***	0.371***	0.374***	0.370***
	(0.022)	(0.023)	(0.025)	(0.023)	(0.023)
Fraction All P2s Chose Dominant$_{t-1}$ (D_{jt})		0.475**	0.451**		
		(0.243)	(0.249)		
Game Played Previously (K_{jt})			0.007		0.050**
			(0.018)		(0.020)
SPE Directly Implied in Session j (S_{jt})			0.025		
			(0.019)		
SPE Indirectly Implied in Session j (S^*_{jt})			0.012		
			(0.033)		
Fraction i's Counterparts Chose Dominant$_{t-1}$ (D_{ijt})				0.308***	0.274***
				(0.084)	(0.087)
SPE Directly Implied (by i's personal history) (S_{ijt})					0.089***
					(0.020)
SPE Indirectly Implied (by i's personal history) (S^*_{ijt})					0.045*
					(0.034)
Constant	0.374***	-0.058	-0.045	0.116*	0.115*
	(0.025)	(0.222)	(0.228)	(0.075)	(0.078)
Observations	7830	7743	7743	7743	7743
R-Sq.	0.234	0.235	0.235	0.236	0.240

Clustered standard errors in parentheses.
One-tailed hypothesis tests * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1: **OLS Regression Analysis of Learning.**

In other words, our subjects exhibit characteristics of both *ToP* and *ToP** strategies in the sense of our model. That is, even in this complex setting, individuals are able to learn the preferences of others and use this information in a strategic setting. However, our evidence also suggests a few important caveats: first, subjects seem to condition their behavior disproportionately on the payoff-salient interactions with their previous counterparts, ignoring the information that could be gleaned from the public history. This is inconsistent with the theory, which assumes *ToP* and *ToP** types use all available information to learn.

However, it is at least psychologically plausible that individuals learn disproportionately from their personal experiences. Second, we do not see evidence that, once the *SPE* action is either directly or indirectly implied by past choices by P2s, subjects *always* choose correctly, as if they were *ToP* or *ToP** types in the model. In that sense, we do not view our theory as providing a best fit model of learning in games - rather, we would suggest that existing learning models might be augmented by considering the implications of our notion of "theory of mind", which assumes that individuals exploit the implications of goal-driven behavior by others. Finally, though our subjects do show evidence of learning according to *ToP* and *ToP** strategies, on average, they nevertheless vary considerably in their ability to do so. We summarize these observations below:

Finding 1: On average, there is a significant increase in the fraction of Player 1s making the *SPE* choice over time, despite individual variation.

Finding 2: The increase is driven by observation of Player 2's preferences (theory of preferences) and not only by naive reinforcement learning, although subjects seem to learn primarily from their personal histories of interaction.

4.2. **Theory of Preferences and Theory of Mind**

Table 1 provides evidence that increases in the overall fraction of *SPE* choices by Player 1s over time result from theory of preferences. However, our data reveal clear heterogeneity across individuals. Thus, we unpack this heterogeneity to ask whether our measure of theory of preferences varies at the individual level with survey measures of theory of mind from psychology. Specifically, we examine how subjects' AQ scores vary with the rate at which subjects learn to choose the *SPE* action by drawing inferences from the past actions of others.[23]

We use the regression specification reported in column (5) of Table 1 as our starting point. This regression identified the average marginal effect of the variables K_{jt}, S_{ijt} and S^*_{ijt} on the probability of choosing the *SPE* action. We augment this regression specification by also including on the right hand side a subject's AQ score and interactions between the AQ score and the K_{jt}, S_{ijt} and S^*_{ijt} variables. These interactions allow us to estimate how the ability to

[23] Using BAP scores provides qualitatively similar results, though the results are not as sharp, possibly due to the fact that there is less variation in BAP scores than in AQ scores. Estimates available upon request.

exploit the information used by *ToP* and *ToP** types in the model varies with the subject's AQ score. As control tests, we also include interactions between AQ scores and the dummy variables for the presence of a dominant strategy and for those games in which the "highest mean" heuristic led to the *SPE* action.[24] As before, we cluster standard errors at the session level, and we employ OLS so that coefficients can be directly interpreted as marginal effects; full estimates are reported in the Appendix in Table F2. Also, recall that lower AQ scores reflect fewer self-reported autism spectrum behaviors, so that a lower AQ score is associated with better "theory of mind".

Columns (1) - (3) of Table 2 report linear combinations of the coefficients that allow us to assess how the probability of choosing the *SPE* action varies across the AQ distribution. For each of the variables shown (in column (5) of Table 1) to increase the average probability of choosing the *SPE* action, we compute the marginal effect at the lower decile, the median and the upper decile of the AQ distribution. For instance, the estimated marginal effect reported in the first row of column (1) is the sum of the coefficient on AQ and 13.5 (the 1st decile of the observed AQ distribution) times the coefficient on the interaction between AQ and the dummy variable indicating that Player 1 had a dominant action. The remaining estimates are constructed analogously. Each row of column (4) reports the results of an F-test of the hypothesis that the estimates in columns (1) and (3) are equal, i.e. a significant test reveals differences in the marginal effects of the relevant dummy variable at the upper and lower decile of the AQ distribution. This allows us to test whether lower AQ scores (stronger theory of mind) are associated with improved choices.

First, note that our data pass the control tests; that is, we find no evidence that the marginal increase in the probability of choosing the *SPE* action varies across the AQ distribution, when there is a dominant action or when the "highest mean" heuristic gives the *SPE* action. Second, while we find no evidence that the marginal effect of direct implications (S_{ijt}) varies significantly with AQ score, we find that lower AQ individuals are substantially more likely to choose the *SPE* action when it has been *indirectly* implied (S^*_{ijt}). This is consistent with our interpretation of theory of preferences as an aspect of theory of mind, as it is

[24] Note that in order to perform this analysis, we can no longer include individual fixed effects as in Table 1 since the AQ score does not vary at the individual level.

	(1) Lower Decile AQ = 13.5	(2) Median AQ = 20	(3) Upper Decile AQ = 26	(4) F-stat H_0: (1) = (3)
Has Dominant Strategy	0.49	0.49	0.49	0.00
	(0.04)	(0.02)	(0.05)	
"Highest Mean" Heuristic = SPE	0.36	0.38	0.39	0.18
	(0.04)	(0.02)	(0.04)	
Game Played Previously (K_{jt})	0.12	0.06	0.01	4.65**
	(0.04)	(0.02)	(0.03)	
SPE Directly Implied (by i's private history) (S_{ijt})	0.11	0.10	0.08	0.29
	(0.03)	(0.02)	(0.03)	
SPE Indirectly Implied (by i's private history) (S^{*}_{ijt})	0.13	0.03	-0.06	6.53**
	(0.06)	(0.04)	(0.04)	

Standard errors as computed by the lincom command in Stata in parentheses.
F-test p-values, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 2: **Marginal Effects across the AQ Distribution.**

understood by psychologists. Perhaps surprisingly, we also find that the ability to learn from past observations of the complete game varies with AQ score in the same manner.

Overall, this table provides tantalizing evidence that the ability to learn from preference revelation in our games is associated with theory of mind as it is conceived by psychologists. Reading through the questionnaires, the observed correlations agree with intuition. For example, consider questions underlying the AQ subscale that emphasizes *social skills*: AQ_Social. We highlight this subscale because it is explicitly designed to measure capacity for and enjoyment of social interaction, which is particularly reliant on theory of mind. One particularly telling item on the AQ_Social subscale asks individuals how strongly they agree with the statement:

"I find it difficult to work out people's intentions."

This is consistent with our notion of theory of preferences in a strategic setting.

It is worth reiterating that our survey data exhibit AQ scores in the normal range. Thus, differences in the strategic aspects of theory of mind vary significantly across individuals in the normal range of social intelligence.[25]

[25] Figure F3 in the online appendix displays histograms of our participants' AQ and BAP scores over the range of feasible scores. There we report additional analyses showing that

Finding 3: Our dynamic measure of theory of preferences based on the observed rate of learning of preferences from past interactions varies significantly with survey measures of theory of mind.

5. Conclusions

This paper develops a theoretical model to study theory of preferences, a central component of theory of mind. The model demonstrates the advantages to attributing preferences to others in simple games of perfect information. We show that sophisticated types that recognize agency in others can build up a picture of others' preferences and thus learn more rapidly than naive types that learn each game separately (Proposition 1).

We then perform experiments measuring the ability of human subjects to learn the preferences of others in a strategic setting. The experiments exploit the insights of the theoretical model to distinguish those games in which naive types should be able to infer the *SPE* action from the history of play and those games in which only sophisticated types should be able to infer to *SPE* action from the history of play. We find evidence of 1) significant learning of counterparts' preferences over time on average, consistent with the presence of sophistication, and of 2) a relationship between sophisticated (theory of preferences) learning in these experiments and responses to two well-known survey instruments measuring theory of mind from psychology. This validates the use of the term "theory of mind" in the present context. Indeed, the experiments here raise the interesting possibility of incorporating behavioral measures into tests of autism to complement other measures based on surveys.

In economics, theory of mind is implicated, in particular, as driving behavior in social settings involving reciprocity and mutualistic gains from exchange (see, for example, McCabe et al., 2003, and Izuma et al. 2011). Theory of mind is crucial here because individuals condition their behavior on others' beliefs and intentions, presuming preferences in those others.

We show that the essential capacity to attribute preferences to others provides a substantial strategic advantage theoretically and is actually present in our

---

individual-level estimated coefficients on S^*_{ijt} are correlated with AQ as well as some BAP and AQ subscales, providing further evidence for a relationship between our notion of theory of preferences and psychologists' notion of theory of mind.

sample to a varying degree. Other social phenomena that assume the presence of theory of mind then gain firmer footing, and so an indirect contribution of our work is to set the stage for future research on such phenomena.

References

[1] **Simon Baron-Cohen, Alan M. Leslie & Uta Frith** (**1985**): Does the Autistic Child Have a 'Theory of Mind?'. *Cognition*, 21(1), 37–46.

[2] **Simon Baron-Cohen, Sally Wheelwright, Richard Skinner, Joanne Martin & Emma Clubley** (**2001**): The Autism-S pectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Males and Females, Scientists and Mathematicians. *Journal of Autism and Developmental Disorders*, 31(1), 5–17.

[3] **Colin F. Camerer, Teck-Hua Ho & Juin-Kuan Chong** (**2002**): Sophisticated Experience-Weighted Attraction Learning and Strategic Teaching in Repeated Games. *Journal of Economic Theory*, 104(1), 137–188.

[4] **Gary B. Charness & Matthew Rabin** (**2002**): Understanding Social Preferences With Simple Tests. *Quarterly Journal of Economics*, 117(3), 817–869.

[5] **Vincent P. Crawford & Nagore Iriberri** (**2007**): Level-k Auctions: Can a Non-Equilibrium Model of Strategic Thinking Explain the Winner's Curse and Overbidding in Private-Value Auctions? *Econometrica*, 75(6), 1721–1770.

[6] **Ido Erev & Alvin E. Roth** (**1998**): Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria. *American Economic Review*, 88(4), 848–881.

[7] **Dietmar Fehr & Steffen Huck** (**2016**): Who Knows It Is a Game? On Strategic Awareness and Cognitive Ability. *Experimental Economics*, 19(4), 713–726.

[8] **Lawrence E. Fouraker & Sidney Siegel** (**1963**): *Bargaining Behavior*. McGraw Hill.

[9] **Daniel Friedman, Steffen Huck, Ryan Oprea & Simon Weidenholzer** (**2015**): From Imitation to Collusion: Long-run Learning in a Low-Information Environment. *Journal of Economic Theory*, 155, 185–205.

[10] **Sotiris Georganas, Paul Healy & Roberto Weber** (**2015**): On the Persistence of Strategic Sophistication. *Journal of Economic Theory*, 159, 369–400.

[11] **Jacob K. Goeree, Thomas R. Palfrey, Brian W. Rogers & Richard D. McKelvey** (**2007**): Self-Correcting Information Cascades. *The Review of Economic Studies*, 74(3), 733–762, URL `http://restud.oxfordjournals.org/content/74/3/733.abstract`.

[12] **Veronika Grimm & and Friederike Mengel** (**2012**): An Experiment on Learning in a Multiple Games Environment. *Journal Economic Theory*, 147(6), 2220–2259.

[13] **Steffen Huck, Philippe Jehiel & Tom Rutter** (**2001**): Feedback Spillover and Analogy-Based Expectations: A Multi-Game Experiment. *Games and Economic Behavior*, 71(2), 351–365.

[14] **Robert S.E. Hurley, Molly Losh, Morgan Parlier, Stephen J. Reznick & Joseph Piven** (**2007**): The Broad Autism Phenotype Questionnaire. *Journal of Autism and Developmental Disorders*, 37(9), 1679–1690.

[15] **Keise Izuma, Kenji Matsumoto, Colin F. Camerer & Ralph Adolphs** (**2011**): Insensitivity to Social Reputation in Autism. *Proceedings of the National Academy of Sciences*, 108(42), 17302–17307, URL `http://www.pnas.org/content/108/42/17302.abstract`.

[16] **Philippe Jehiel** (**2005**): Analogy-Based Expectation Equilibrium. *Journal of Economic Theory*, 123(2), 81–104.

[17] **Gil Kalai** (**2003**): Learnability and Rationality of Choice. *Journal of Economic Theory*, 113, 104–117.

[18] **Kevin A. McCabe, Stephen J. Rassenti & Vernon L. Smith** (**1998**): Reciprocity, Trust, and Payoff Privacy in Extensive Form Bargaining. *Games and Economic Behavior*, 24(1), 10–24.

[19] **Kevin A. McCabe, Mary L. Rigdon & Vernon L. Smith** (**2003**): Positive Reciprocity and Intentions in Trust Games. *Journal of Economic Behavior and Organization*, 52(2), 267–275.

[20] **Erik Mohlin** (**2012**): Evolution of Theories of Mind. *Games and Economic Behavior*, 75(1), 299–318.

[21] **Heinrich H. Nax, Maxwell N. Burton-Chellew, Stuart A. West, H. Peyton Young** (**2016**): Learning in a Black Box. *Journal of Economic Behavior & Organization*, 127(1), 1–15.

[22] **Jörg Oechssler & Burkhard Schipper** (**2003**): Can You Guess the Game You are Playing? *Games and Economic Behavior*, 43(1), 137–152.

[23] **Kristine H. Onishi & René Baillargeon** (**2005**): Do 15-Month-Old Infants Understand False Beliefs? *Science*, 308(8), 255–258.

[24] **R Development Core Team** (**2013**): *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, URL `http://www.R-project.org/`, ISBN 3-900051-07-0.

[25] **Nikolaus Robalino & Arthur J. Robson** (**2016**): The Evolution of Strategic Sophistication. *American Economic Review*, 106, 1046–72.

[26] **Vernon L. Smith** (**1982**): Microeconomic Systems as an Experimental Science. *American Economic Review*, 72(5), 923–955.

[27] **Jakub Steiner & Colin Stewart** (**2008**): Contagion Through Learning. *Theoretical Economics*, 3(4), 431-458.

[28] **M. Vidyasagar** (**1997**): *A Theory of Learning and Generalization.* London: Springer.

Appendices

A. Proof of Proposition 1

We begin by establishing the upper bounds from Proposition 1. The result concerning these is restated here for convenience—

Lemma 1: *Suppose* $|Z| = N \geqslant 4$. *There is a function* $U(\varepsilon)$, *with* $U(\varepsilon) < \infty$ *for all* $\varepsilon > 0$, *such that the following is true for each given* ε, *and every number of outcomes* N. *The number of examples needed for a player* 1 *strategy* σ *to learn* G *with probability no less than* $1 - \varepsilon$ *is at most* $N^{\alpha} \cdot U(\varepsilon)$, *where* $\alpha = 4$ *when* σ *is a naive strategy,* $\alpha = 2$ *when it is a ToP strategy, and* $\alpha = 1$ *when it is a ToP* strategy.*

Our proof of Lemma 1 relies on PAC (Probably Approximately Correct) learning theory. The required elements from the theory are described in Kalai (2003), where they are applied to the related problem of learning a rational agent's preferences, and also in Vidyasagar (1997).[26]

The PAC learning problem is as follows. Consider a family of functions F from a set X to a set Y, and a probability measure P on X. Given an unknown function $f \in F$ we wish to learn f but can only observe its values on a sequence of examples $x_1, \ldots, x_M$ drawn independently from X according to P.

Say that F is PAC-learnable from M examples with probability no less than $1-\varepsilon$ with respect to P if the following is true for every $f \in F$. With probability no less than $1 - \varepsilon$, a sequence of $M + 1$ examples drawn independently from P, say $(x_1, \ldots, x_{M+1})$, is such that $f'(x_{M+1}) = f(x_{M+1})$ for every $f' \in F$ that coincides with f on the first M terms of the sequence.[27] Then, say that F is PAC-learnable from M examples with probability no less than $1-\varepsilon$ if the above holds uniformly for every probability measure P on X.

The number of examples needed to PAC-learn a family of functions depends on the desired probability of successful learning, $1-\varepsilon$, but also on the complexity of the family of candidate functions F. The P-dimension of F gives a suitable

[26] We present here the aspects of the theory as they are needed given the current context. In particular, we do not require the full generality of available PAC learning results.

[27] That is, if we know that f is contained in F, then with probability no less than $1 - \varepsilon$ we can correctly guess the value of $f(x)$ after having observed the value of f on M examples drawn independently according to P.

measure of this complexity. In particular, say that F *shatters* $X' \subseteq X$ if there is a function $f \in F$ such that for every partition of X' into two sets, X_1, X_2, there is an $f' \in F$ such that $f'(x) = f(x)$ if $x \in X_1$, and $f'(x) \neq f(x)$ if $x \in X_2$. The P-dimension of F is the size of the largest set that is shattered by F. The key result from PAC learning is then (for a proof see Vidyasagar (1997))—

THEOREM 1: *Suppose the family of functions F has finite P-dimension $d \geqslant 2$. Then there exists a function $U(\varepsilon)$ (finite valued for $\varepsilon > 0$) such that F is PAC-learnable from M examples with probability no less than $1 - \varepsilon$ whenever $M \geqslant d \cdot U(\varepsilon)$.*

The following claim, which follows immediately from the definition of the P-dimension, will be used.

CLAIM 1: *Suppose that for an integer d the following holds for every set $X' \subseteq X$ with $|X'| > d$. For each $f \in F$ there is an $x \in X'$, and a set $X'' \subseteq X' \setminus \{x\}$ such that if $f'(x) \neq f(x)$, then $f'(x') \neq f(x')$ for some $x' \in X''$. Then, the P-dimension of F is less than d.*

In the remainder we label the actions at each node, L and R, for both the players, 1 and 2. Each game is thus identified with a 4-tuple of distinct outcomes, say (w, x, y, z), where, in order from left to right, the outcomes correspond to the LL, LR, RL, and RR terminal nodes of the fixed game tree, respectively.

In order to relate PAC learning to learning in our environment as in Definition 4 we proceed as follows. Let Σ be the set of all mappings from the set of games G to the set of Player 2 pure strategies, $A \times A$, where $A = \{L, R\}$. Recall that a naive strategy makes the *SPE* choice in the realized game in period $n = 1$ whenever the game appears as an example along h. It follows that a naive strategy can learn G after M examples with probability no less than $1 - \varepsilon$ (as in our Definition 4) if Σ is PAC-learnable from M examples with probability no less than $1 - \varepsilon$. For the theory of preferences strategies we proceed in a similar manner. With this in mind, consider choice functions, C, that map each set of outcomes to a a singleton outcome contained in the set, for example, $C(\{w, x\}) \in \{w, x\}$ for all $\{w, x\} \subset Z$. Let $\mathscr{R}$ denote the set of all mappings from G to pairs of outcomes that are consistent with such a choice function (not necessarily consistent with a transitive preference relation) for player 2. Specifically, $R \in \mathscr{R}$ if and only if $R(w, x, y, z) = (C(\{w, x\}), C(\{y, z\}))$

for some choice function C over the outcomes. Let $\mathscr{R}^*$ denote the subset of $\mathscr{R}$ where the underlying choice function is generated by a transitive strict preference relation over the outcomes. Notice that a *ToP* strategy can learn G after M examples with probability no less than $1-\varepsilon$ whenever $\mathscr{R}$ is PAC-learnable from M examples with probability no less than $1-\varepsilon$. This holds true also for the *ToP** strategy when $\mathscr{R}$ is replaced with $\mathscr{R}^*$.

In the light of Theorem 1, Lemma 1 is now immediate given the following result.

LEMMA 2: *Given N outcomes the P-dimension of Σ is no greater than N^4. For $\mathscr{R}$ it is no greater than N^2, and for $\mathscr{R}^*$ it is no greater than N.*

Proof. In particular, the P-dimension of Σ is $|G| < N^4$, i.e., Σ is the set of *all* the mappings from G to $A \times A$. The P-dimension of $\mathscr{R}$ is less than $Q = N \cdot (N-1)/2$. To see this, notice that given N outcomes there are Q binary choices. If G' is such that $|G'| \geqslant Q$, then for some $(w,x,y,z) \in G'$ the binary choices $\{w,x\}$, and $\{y,z\}$ appear again in $G' \setminus \{(w,x,y,z)\}$ for Player 2, in, say, g, and g'. But then, given any $R, R' \in \mathscr{R}$ if $R(w,x,y,z) \neq R'(w,x,y,z)$, then either $R(g) \neq R'(g)$, or $R(g') \neq R'(g')$, and thus Claim 1 gives the desired result.

The P-dimension of $\mathscr{R}^*$ is no greater than $N-1$. To see this consider the following. First, say that the game (w,x,y,z) is *linked* in G' by $R \in \mathscr{R}^*$ if for each binary choice of Player 2 in the game, $\{w',x'\} \in \{\{w,x\},\{y,z\}\}$, either 1) the choice appears for player 2 in a game in $G' \setminus \{(w,x,y,z)\}$, or 2) there is sequence of choices $\{w',y_1\},\{y_1,y_2\},\ldots,\{y_{s-1},y_s\},\{y_s,x'\}$ that all arise for Player 2 in $G' \setminus \{(w,x,y,z)\}$ such that either $w' \succ_R y_1 \succ_R \cdots \succ_R y_s \succ_R x'$, or $x' \succ_R y_s \succ_R \cdots \succ_R y_1 \succ_R w'$, where $\succ_R$ is the transitive preference relation that is consistent with R. Notice that if (w,x,y,z) is linked in G' by R, then $R'(w,x,y,z) \neq R(w,x,y,z)$ implies $R'(g) \neq R(g)$ for some $g \in G' \setminus \{(w,x,y,z)\}$. The desired result then follows from Claim 1 upon showing that if G' contains more than $N-1$ games, then for each $R \in \mathscr{R}^*$ there is a game in G' that is linked in G' by R. Suppose by way of contradiction, then, that no game in G' is linked by R, where $|G'| > N-1$. It follows that for each game in G' at least one Player 2 choice fails to meet both the criteria 1) and 2) in the defininion of "linked" above. It must be possible then to take one Player 2 binary choice from each game in G' in order to obtain a set of binary choices, say $C = \{\{x_1, y_2\}, \ldots, \{x_K, y_K\}\}$, such that C is shattered by $\mathscr{P}(X)$ the set of transitive preference relations on X. Kalai (2003), however, shows that the P-dimension of $\mathscr{P}(X)$ is $N - 1$. We obtain a contradiction then if $K > N - 1$, that is if $|G'| > N - 1$. ∎

We now complete the proof of Proposition 1 by establishing the lower bounds from the result, which are restated here, again, for convenience—

Lemma 3: *There is a function $L(\varepsilon)$ such that the following is true for each given ε, and every number of outcomes N. The number of examples needed for a Player 1 strategy σ to learn g with probability no less than $1 - \varepsilon$ is at least $N^\alpha \cdot L(\varepsilon)$, where $\alpha = 4$ when σ is a naive strategy, $\alpha = 2$ when it is a ToP strategy, and $\alpha = 1$ when it is a ToP* strategy.*

Proof. Recall that our notion of learning by a strategy requires that it make the *SPE* choice with some given probability in the terminal game g, for any possible preferences of Player 2. In order to prove Lemma 3 it then suffices to show that for each strategy of Player 1 there is a utility function of Player 2 for which learning by the strategy requires at least the claimed number of examples.

Recall that for each game (w, x, y, z) the outcomes in order from left to right correspond to the LL, LR, RL, and RR end nodes of the fixed game tree, respectively. Then, consider the set of games, $F \subset G$, such that $(w, z, y, z) \in F$ if and only if one of the Player 2 subgames of the game includes both the best and the worst outcome of the four available ones, from 1's perspective.

Say that a game is *unfamiliar* to a naive strategy of Player 1 if it has not arisen along h. For a *ToP* strategy of Player 1, a game is unfamiliar if neither of the pairwise choices of Player 2 in the game have arisen along h. For a *ToP** strategy, a game is unfamiliar if neither of Player 2's choices in the game can be inferred from the historical data given that 2 has transitive preferences.

Now consider a Player 1 strategy σ of a given type (naive, *ToP* or *ToP**). Let $F_1 \subseteq F$ be the games in F such that whenever the game is unfamiliar to σ it chooses into the Player 2 subgame that includes the best and worst outcomes from 1's perspective (of those four that are available in the game). Similarly, F_2 are the games in F in which σ chooses away from the best and worst outcomes, when the game is unfamiliar. Notice that for at least one of $k = 1, 2$ it must be that $|F_k|/|F| \geqslant 1/2$. Consider first the case in which $|F_1|/|F| \geqslant 1/2$.

Consider for Player 2 a utility function u_2 that reverses, for every binary choice, Player 1's preference as given by u_1. The strategy σ will then make a suboptimal choice at a game in F_1, whenever it is unfamiliar. It follows that, given any sequence of examples, Player 1 with strategy σ will make the *SPE* choice in period $n = 1$ with probability that is bounded above by $(|F_1| - K)/|G|$, where K is the number of distinct games in F_1 that are *not* unfamiliar to σ. That is, for any given h the strategy σ makes the *SPE* choice in g with probability no less than $1 - \varepsilon$ only if $(|F_1| - K)/|G| < \varepsilon$, which implies

$$K \;>\; |F_1| - |G| \cdot \varepsilon = N^4 \cdot \left[\frac{|G|}{N^4} \cdot \left(\frac{|F_1|}{|F|} \cdot \frac{|F|}{|G|} - \varepsilon \right) \right]. \tag{2}$$

A straightforward calculation gives $|F|/|G| = 1/3$. Recall that $|G| = N \cdot (N-1) \cdot (N-2) \cdot (N-3)$. The assumption that $N \geqslant 4$ then ensures that $|G|/N^4 \geqslant 3/32$. The bracketed term in (2) is then bounded below by $\frac{3}{32}\left(\frac{1}{6} - \varepsilon\right)$, since we have supposed that $|F_1|/|F| \geqslant 1/2$. Equation (2) then implies

$$K \;> N^4 \cdot \frac{3}{32}\left(\frac{1}{6} - \varepsilon \right). \tag{3}$$

Notice next that for the naive strategy $K \leqslant M$ since only M examples are provided in period $n = 0$. The number of distinct pairwise choices of Player 2 that are observed along h is at most $2 \cdot M$. For the *ToP* strategy it thus follows that $K \leqslant (2 \cdot M)^2$, since this exceeds the number of games in G that can be generated from $2 \cdot M$ binary choices of Player 2. It also follows that the longest chain of Player 2 preferences that can be inferred from h, given the implications of transitivity, is $2 \cdot M$. Hence, at most $2 \cdot M \cdot (2 \cdot M - 1)/2 < (2 \cdot M)^2$ binary choices can be inferred from h by the *ToP** strategy. Clearly, for the *ToP** strategy $K < (2 \cdot M)^4$, since this exceeds the number of games in G that can be generated given $(2 \cdot M)^2$ binary choices of Player 2. Combining these facts with equation (3) gives that σ can learn g from M examples with probability no less than $1 - \varepsilon$ only if

$$
\begin{aligned}
M &> N^4 \cdot \left[\left(\frac{3}{32} \cdot \left(\frac{1}{6} - \varepsilon\right)\right)\right], && \text{for naive } \sigma, \\
M &> N^2 \cdot \left[\frac{1}{2} \cdot \sqrt{\left(\frac{3}{32} \cdot \left(\frac{1}{6} - \varepsilon\right)\right)}\right], && \text{for } \mathit{ToP}\ \sigma, \\
M &> N \cdot \left[\frac{1}{2} \cdot \sqrt[4]{\left(\frac{3}{32} \cdot \left(\frac{1}{6} - \varepsilon\right)\right)}\right], && \text{for } \mathit{ToP}^*\ \sigma.
\end{aligned}
\tag{4}
$$

For σ such that $|F_2|/|F| > 1/2$ identical bounds can be obtained by choosing u_2 equal to u_1, so that the choice of σ is suboptimal for every unfamiliar game in F_2. We thus set $L(\varepsilon) = \frac{3}{32} \cdot (1/6 - \varepsilon)/2$,[28] to complete the proof in view of equation (4). ■

[28] Notice that $L(\varepsilon)$ thus defined is no greater than any of the three bracketed expressions in the component equations of (4). Thus, for instance, if $M < N^2 \cdot L(\varepsilon)$, then the *ToP* strategy cannot learn.

B. Experiment Instructions

Page 1

In this experiment you will participate in a series of two person decision problems. The experiment will last for a number of rounds. Each round you will be randomly paired with another individual. The joint decisions made by you and the other person will determine how much money you will earn in that round.

Your earnings will be paid to you in cash at the end of the experiment. We will not tell anyone else your earnings. We ask that you do not discuss your earnings with anyone else.

If you have a question at any time, please raise your hand.

Page 2

You will see a diagram similar to one on your screen at the beginning of the experiment. You and another person will participate in a decision problem shown in the diagram.

One of you will be Person 1 (orange). The other person will be Person 2 (blue). In the upper left corner, you will see whether you are Person 1 or Person 2.

You will be either a Person 1 or a Person 2 for the entire experiment.

Page 3

Notice the four pairs of squares with numbers in them; each pair consists of two earnings boxes. The earnings boxes show the different earnings you and the other person will make, denoted in Experimental Dollars. There are two numbers, Person 1 will earn what is in the orange box, and Person 2 will earn what is in the blue box if that decision is reached.

In this experiment, you can only see the earnings in your own box. That is, if you are Person 1 you will only see the earnings in the orange boxes, and if you are Person 2 you will only see the earnings in the blue boxes. Both boxes will be visible, but the number in the other person's box will be replaced with a "?".

However, for each amount that you earn, the amount the other person earns is fixed. In other words, for each amount that Person 1 sees, there is a corresponding, unique amount that will always be shown to Person 2.

For example, suppose Person 1 sees an earnings box containing "12" in round 1. In the same pair, suppose Person 2 sees "7". Then, at any later round, anytime Person 1 sees "12", Person 2 will see "7".

Together, you and the other person will choose a path through the diagram to an earnings box. We will describe how you make choices next.

Page 4

A node, displayed as a circle and identified by a letter, is a point at which a person makes a decision. Notice that the nodes are color coded to indicate whether Person 1 or Person 2 will be making that decision. You will always have two options.

If you are you Person 1 you will always choose either "Right" or "Down", which will select a node at which Person 2 will make a decision.

If you are Person 2 you will also choose either "Right" or "Down" which will select a pair of earnings boxes for you and Person 1.

Once a pair of earnings boxes is chosen, the round ends, and each of you will be able to review the decisions made in that round.

Page 5

In each round all pairs will choose a path through the same set of nodes and earnings boxes. This is important because at the end of each round, in addition to your own outcome, you will be able to see how many pairs ended up at each other possible outcome.

While you review your own results from a round, a miniature figure showing all possible paths through nodes and to earnings boxes will be displayed on the right hand side of the screen.

The figure will show how many pairs chose a path to each set of earnings boxes.

The Payoff History table will update to display your payoff from the current period.

Page 6

We have provided you with a pencil and a piece of paper on which you may write down any information you deem relevant for your decisions. At the end of the experiment, please return the paper and pencil to the experimenter.

At the end of the experiment, we will randomly choose 2 rounds for payment, and your earnings from those rounds will be summed and converted to $CAD at a rate of 1 Experimental Dollar = $2.

Important points:

1. You will be either a Person 1 or a Person 2 for the entire experiment.
2. Each round you will be randomly paired with another person for that round.
3. Person 1 always makes the first decision in a round.
4. Person 1's payoff is in the orange earnings box and Person 2's in the blue earnings box.
5. Each person will only be able to see the numbers in their own earnings box.
6. Earnings always come in unique pairs so that for each amount observed by Person 1, the number observed by Person 2 will be fixed.
7. In a given round, all pairs will choose a path through the same set of nodes and earnings boxes.
8. After each round you will be able to see how many pairs ended up at each outcome.

9. We will choose 2 randomly selected periods for payment at the end of the experiment. Any questions?

C. Screenshots

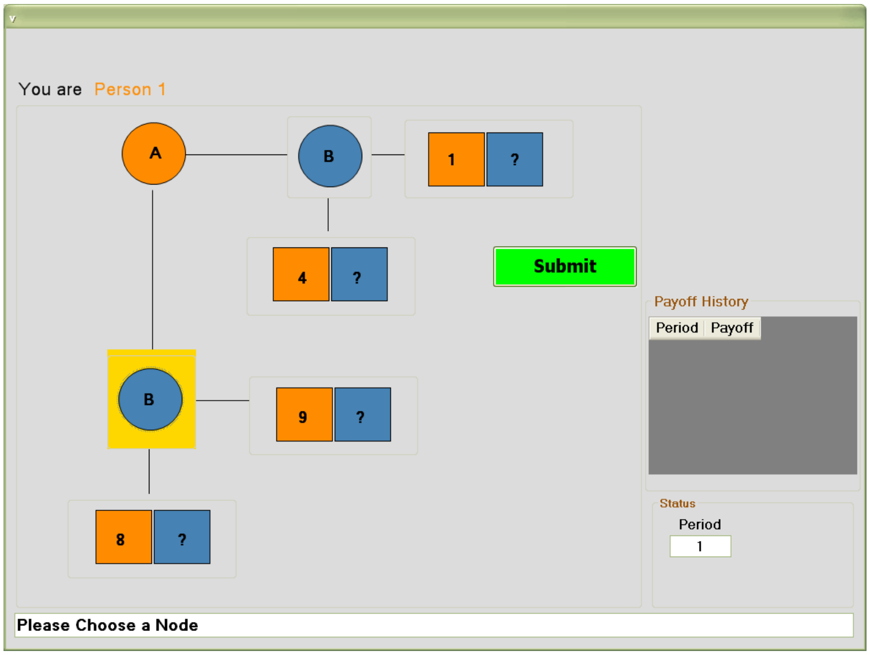

Figure C1: **Screenshot for Player 1.** This figure shows the screen as player 1 sees it prior to submitting his choice of action. The yellow highlighted node indicates that player 1 has provisionally chosen the corresponding action, but the decision is not final until the submit button is clicked. While waiting for player 1 to choose, player 2 sees the same screen except that she is unable to make a decision, provisional choices by player 1 are not observable, and the "Submit" button is invisible.

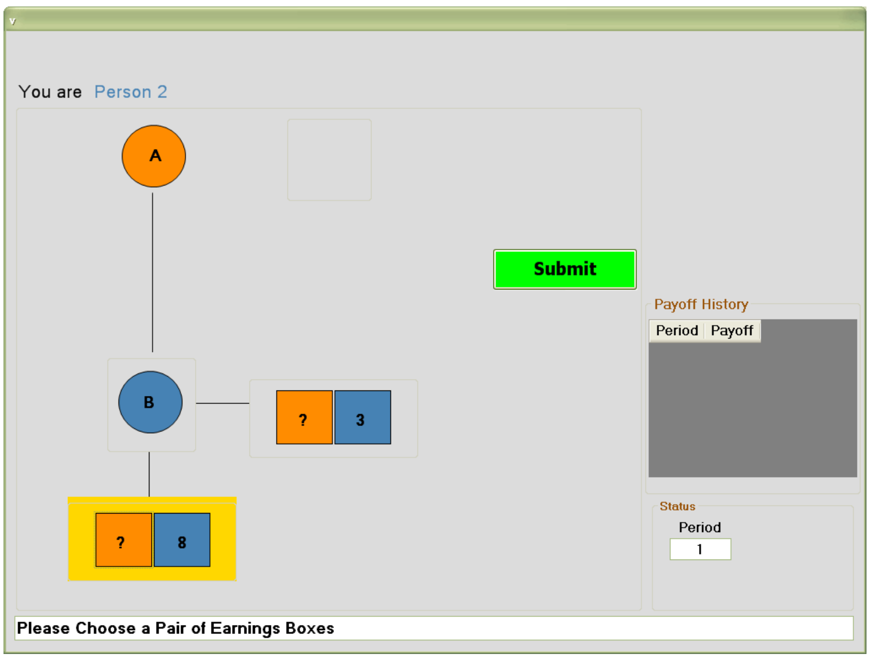

FIGURE C2: **Screenshot for Player 2.** This figure shows the screen as player 2 sees it after player 1 has chosen an action. Here, player 1 chose to move down, so the upper right portion of the game tree is no longer visible. While player 2 is making a decision, player 1 sees an identical screen except that he is unable to make a decision and the "Submit" button is invisible.

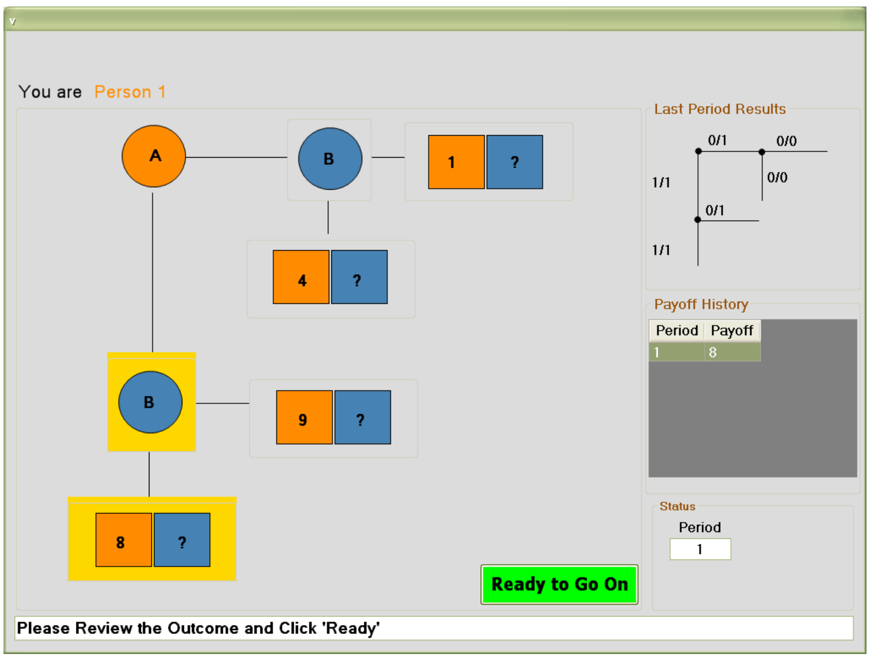

FIGURE C3: **Screenshot of Post-Decision Review.** This figure shows the final screen subjects see in each period after both player 1 and player 2 have made their decisions. The smaller game tree in the upper right portion of the figure displays information about how many pairs ended up at each outcome. For the purposes of the screenshot, the software was run with only one pair, but in a typical experiment, subjects learned about the decisions of 4 pairs (3 other than their own).

D. Autism-Spectrum Quotient Questionnaire

1.	I prefer to do things with others rather than on my own.	[1]	[2]	[3]	[4]
2.	I prefer to do things the same way over and over again.	[1]	[2]	[3]	[4]
3.	If I try to imagine something, I find it very easy to create a picture in my mind.	[1]	[2]	[3]	[4]
4.	I frequently get so absorbed in one thing that I lose sight of other things.	[1]	[2]	[3]	[4]
5.	I often notice small sounds when others do not.	[1]	[2]	[3]	[4]
6.	I usually notice car number plates of similar strings of information.	[1]	[2]	[3]	[4]
7.	Other people frequently tell me that what I've said is impolite, even though I think it is polite.	[1]	[2]	[3]	[4]
8.	When I'm reading a story, I can easily imagine what the characters might look like.	[1]	[2]	[3]	[4]
9.	I am fascinated by dates.	[1]	[2]	[3]	[4]
10.	In a social group, I can easily keep track of several different people's conversations.	[1]	[2]	[3]	[4]
11.	I find social situations easy.	[1]	[2]	[3]	[4]
12.	I tend to notice details that others do not.	[1]	[2]	[3]	[4]
13.	I would rather go to a library than a party.	[1]	[2]	[3]	[4]
14.	I find making up stories easy.	[1]	[2]	[3]	[4]
15.	I find myself drawn more strongly to people than to things.	[1]	[2]	[3]	[4]
16.	I tend to have very strong interests, which I get upset about if I can't pursue.	[1]	[2]	[3]	[4]
17.	I enjoy social chit-chat.	[1]	[2]	[3]	[4]
18.	When I talk, it isn't always easy for others to get a word in edgeways.	[1]	[2]	[3]	[4]
19.	I am fascinated by numbers.	[1]	[2]	[3]	[4]
20.	When I'm reading a story I find it difficult to work out the characters' intentions.	[1]	[2]	[3]	[4]
21.	I don't particularly enjoy reading fiction.	[1]	[2]	[3]	[4]
22.	I find it hard to make new friends.	[1]	[2]	[3]	[4]
23.	I notice patterns in things all the time.	[1]	[2]	[3]	[4]
24.	I would rather go to the theatre than a museum.	[1]	[2]	[3]	[4]
25.	It does not upset me if my daily routine is disturbed.	[1]	[2]	[3]	[4]
26.	I frequently find that I don't know how to keep a conversation going.	[1]	[2]	[3]	[4]
27.	I find it easy to "read between the lines" when someone is talking to me.	[1]	[2]	[3]	[4]
28.	I usually concentrate more on the whole picture, rather than the small details.	[1]	[2]	[3]	[4]
29.	I am not very good at remembering phone numbers.	[1]	[2]	[3]	[4]
30.	I don't usually notice small changes in a situation, or a person's appearance.	[1]	[2]	[3]	[4]
31.	I know how to tell if someone listening to me is getting bored.	[1]	[2]	[3]	[4]
32.	I find it easy to do more than one thing at once.	[1]	[2]	[3]	[4]
33.	When I talk on the phone, I'm not sure when it's my turn to speak.	[1]	[2]	[3]	[4]
34.	I enjoy doing things spontaneously.	[1]	[2]	[3]	[4]
35.	I am often the last to understand the point of a joke.	[1]	[2]	[3]	[4]
36.	I find it easy to work out what someone else is thinking or feeling just by looking at their face.	[1]	[2]	[3]	[4]
37.	If there is an interruption, I can switch back to what I was doing very quickly.	[1]	[2]	[3]	[4]
38.	I am good at social chit-chat.	[1]	[2]	[3]	[4]
39.	People often tell me that I keep going on and on about the same thing.	[1]	[2]	[3]	[4]
40.	When I was young, I used to enjoy playing games involving pretending with other children.	[1]	[2]	[3]	[4]
41.	I like to collect information about categories of things (e.g. types of car, types of bird, types of train, types of plant, etc.	[1]	[2]	[3]	[4]
42.	I find it difficult to imagine what it would be like to be someone else.	[1]	[2]	[3]	[4]
43.	I like to plan any activities I participate in carefully.	[1]	[2]	[3]	[4]
44.	I enjoy social occasions.	[1]	[2]	[3]	[4]
45.	I find it difficult to work out people's intentions.	[1]	[2]	[3]	[4]
46.	New situations make me anxious.	[1]	[2]	[3]	[4]
47.	I enjoy meeting new people.	[1]	[2]	[3]	[4]
48.	I am a good diplomat.	[1]	[2]	[3]	[4]
49.	I am not very good at remembering people's date of birth.	[1]	[2]	[3]	[4]
50.	I find it very easy to play games with children that involve pretending.	[1]	[2]	[3]	[4]

1 = definitely agree, 2 = slightly agree, 3 = slightly disagree, 4 = definitely disagree

E. Broad Autism Phenotype Questionnaire

You are about to fill out a series of statements related to personality and lifestyle. For each question, circle that answer that best describes how often that statement applies to you. Many of these questions ask about your interactions with other people. Please think about the way you are with most people, rather than special relationships you may have with spouses or significant others, children, siblings, and parents. Everyone changes over time, which can make it hard to fill out questions about personality. Think about the way you have been the majority of your adult life, rather than the way you were as a teenager, or times you may have felt different than normal. You must answer each question, and give only one answer per question. If you are confused, please give it your best guess.

1	I like being around other people.	1	2	3	4	5	6
2	I find it hard to get my words out smoothly.	1	2	3	4	5	6
3	I am comfortable with unexpected changes in plans.	1	2	3	4	5	6
4	It's hard for me to avoid getting sidetracked in conversation.	1	2	3	4	5	6
5	I would rather talk to people to get information than to socialize.	1	2	3	4	5	6
6	People have to talk me into trying something new.	1	2	3	4	5	6
7	I am "in-tune" with the other person during conversation.*	1	2	3	4	5	6
8	I have to warm myself up to the idea of visiting an unfamiliar place.	1	2	3	4	5	6
9	I enjoy being in social situations.	1	2	3	4	5	6
10	My voice has a flat or monotone sound to it.	1	2	3	4	5	6
11	I feel disconnected or "out of sync" in conversations with others.*	1	2	3	4	5	6
12	People find it easy to approach me.*	1	2	3	4	5	6
13	I feel a strong need for sameness from day to day.	1	2	3	4	5	6
14	People ask me to repeat things I've said because they don't understand.	1	2	3	4	5	6
15	I am flexible about how things should be done.	1	2	3	4	5	6
16	I look forward to situations where I can meet new people.	1	2	3	4	5	6
17	I have been told that I talk too much about certain topics.	1	2	3	4	5	6
18	When I make conversation it is just to be polite.*	1	2	3	4	5	6
19	I look forward to trying new things.	1	2	3	4	5	6
20	I speak too loudly or softly.	1	2	3	4	5	6
21	I can tell when someone is not interested in what I am saying.*	1	2	3	4	5	6
22	I have a hard time dealing with changes in my routine.	1	2	3	4	5	6
23	I am good at making small talk.*	1	2	3	4	5	6
24	I act very set in my ways.	1	2	3	4	5	6
25	I feel like I am really connecting with other people.	1	2	3	4	5	6
26	People get frustrated by my unwillingness to bend.	1	2	3	4	5	6
27	Conversation bores me.*	1	2	3	4	5	6
28	I am warm and friendly in my interactions with others.*	1	2	3	4	5	6
29	I leave long pauses in conversation.	1	2	3	4	5	6
30	I alter my daily routine by trying something different.	1	2	3	4	5	6
31	I prefer to be alone rather than with others.	1	2	3	4	5	6
32	I lose track of my original point when talking to people.	1	2	3	4	5	6
33	I like to closely follow a routine while working.	1	2	3	4	5	6
34	I can tell when it is time to change topics in conversation.*	1	2	3	4	5	6
35	I keep doing things the way I know, even if another way might be better.	1	2	3	4	5	6
36	I enjoy chatting with people.	1	2	3	4	5	6

*Casual interaction with acquaintances, rather than special relationships such as with close friends and family members

1—very rarely
2—rarely
3—occasionally
4—somewhat often
5—often
6—very often

	BAPQ

F. Additional Tables and Figures

means	8-1	8-2	10-1	10-2	8-3	6-1	8-4	10-3	8-5*	10-3*	10-4	10-5	8-6	10-6	8-7	8-8	8-9	8-10	10-7	8-11
Pairs SPE	0.68	0.54	0.62	0.61	0.63	0.44	0.51	0.46	0.68	0.50	0.32	0.66	0.53	0.38	0.54	0.52	0.57	0.64	0.25	0.71
P1 SPE	0.72	0.59	0.64	0.71	0.67	0.54	0.61	0.54	0.70	0.59	0.50	0.72	0.64	0.50	0.67	0.61	0.63	0.70	0.36	0.72
P2 SPE | P1 SPE	0.94	0.91	0.98	0.87	0.94	0.83	0.84	0.86	0.96	0.85	0.63	0.92	0.83	0.77	0.80	0.85	0.91	0.92	0.71	0.98
P1 Heur	0.36	0.40	0.42	0.38	0.42	0.45	0.34	0.55	0.43	0.43	0.55	0.38	0.46	0.47	0.44	0.52	0.49	0.45	0.59	0.34
P1 SPE | Heur $\neq$ SPE	0.64	0.52	0.55	0.58	0.55	0.36	0.49	0.37	0.64	0.45	0.25	0.64	0.44	0.31	0.46	0.43	0.51	0.58	0.21	0.68
P1 SPE | Heur = SPE	0.81	0.60	0.86	0.74	0.88	0.71	0.58	0.75	0.78	0.73	0.57	0.70	0.73	0.59	0.75	0.77	0.78	0.82	0.52	0.88
P1 SPE | No Heur	0.71	0.69	0.65	0.72	0.70	0.49	0.56	0.57	0.74	0.66	0.63	0.79	0.60	0.51	0.73	0.78	0.73	0.78	0.52	0.81
P2 Dom	0.94	0.90	0.98	0.86	0.94	0.84	0.85	0.82	0.97	0.87	0.63	0.93	0.86	0.78	0.80	0.83	0.91	0.93	0.66	0.98
# Unique Games	55	55	55	55	46	46	46	54	54	50	50	46	46	47	47	47	53	53	57	57
# Unique Games (Pd. 31-90)	27	27	27	27	17	17	17	26	26	23	23	20	20	20	20	20	25	25	27	27

Table F1: **Observed Probability of Outcomes by Session.**

Each entry is a probability that we observed a particular outcome in a particular session. Pairs SPE refers to the probability that a pair ended at the *SPE*. P1 SPE is the probability that player 1's choice was consistent with the *SPE*. P1 Heur is the probability that Player 1 chose in a manner consistent with the "highest mean" rule of thumb (heuristic). P1 SPE | heur $\neq$ SPE is the probability that player one followed the *SPE* when it did not correspond to the "highest mean" rule of thumb. P1 SPE | heur = SPE is the probability that player one followed the *SPE* when it *did* correspond to the "highest mean" rule of thumb. P1 SPE | No Heur is the probability that player 1 followed the *SPE* when the rule of thumb was inapplicable (i.e. equal means). P2 Dom is the probability that player 2 chose the dominant strategy. P2 SPE | P1 SPE is the conditional probability of player 2 choosing the dominant strategy given that player 1 followed the *SPE*. Sessions are labeled in the format # of Subjects – Session ID so that 10-2 corresponds to the 2nd session with 10 subjects. * indicates sessions in which the payoff set was {1,2,3,4,8,9,10}, rather than {1,2,3,4,5,6,7}. The final row indicates the number of unique games played by subjects in each session. We observe no significant correlation between the rate at which subjects end up at the SPE and the number of unique games played ($r = 0.15$, p-value $= 0.52$).

As another robustness check, we note that there is no trend in the rate of Player 2s choosing dominant actions over time. In the first 30 periods, they do so 85% of the time, and in the last 30 periods they do so 88% of the time, with an overall average of 86%. We also see no correlation between AQ scores or BAPQ scores and the share of dominant choices by a person in the role of Player 2 (both p-values > 0.35).

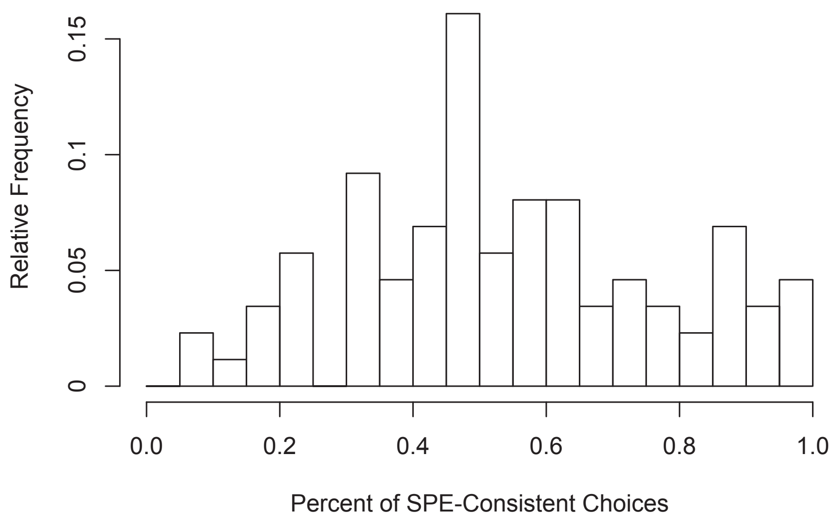

Figure F1: **Histogram of the Individual Rates of SPE-consistent Choices.** The figure excludes all periods in which the player had a dominant strategy and in which choice under the rule of thumb corresponded to the SPE.

Figure F1 displays substantial across-subject heterogeneity in the likelihood of choosing the *SPE* action when doing so was non-trivial.

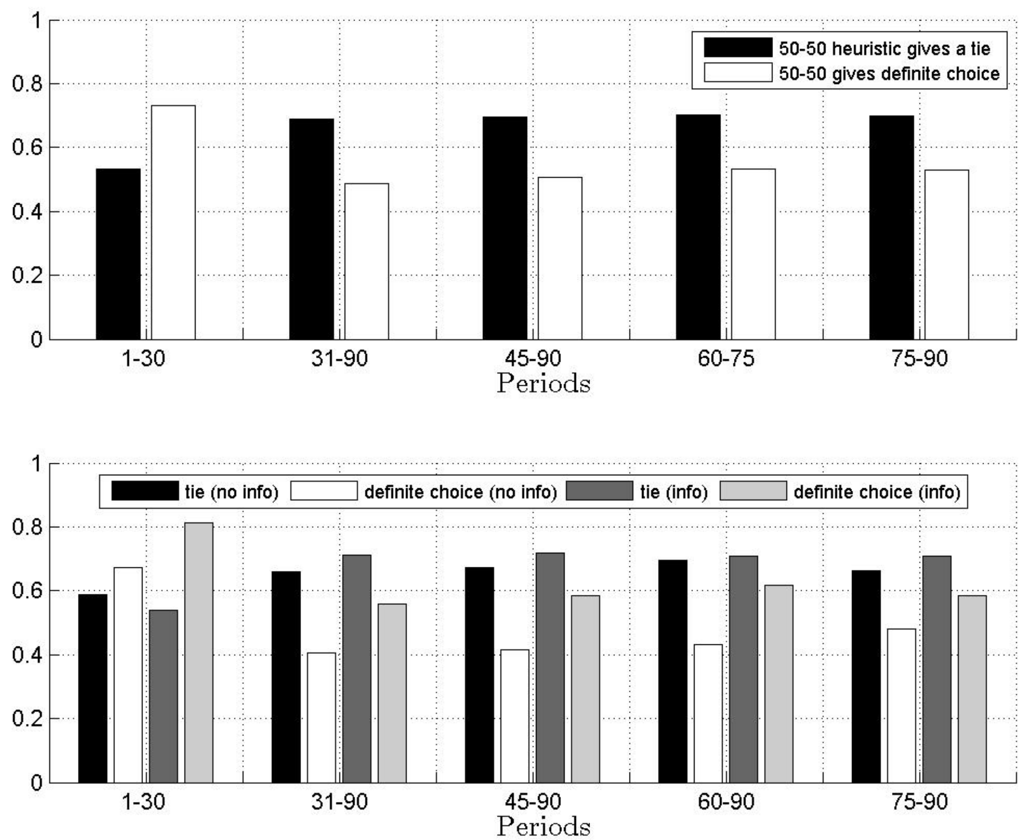

FIGURE F2: **The 50-50 Heuristic and the Fraction of Subjects Choosing the *SPE*.** The top figure gives the fraction of subjects choosing the *SPE* by period intervals (before Treatment NoHeuristic [Periods 1-30], and during NoHeuristic]) in the case where the 50-50 heuristic gives a definite choice, and then when mean payoffs from both actions of Player 1 are equal. The bottom figure considers separately instances where (1) the player 1 faced a new game, and at the same time a player with *ToP* could not infer the *SPE* choice given the public history (no info), and (2) either the game is familiar, or a *ToP* player could infer the *SPE* choice given the public history (info).

Table F2 displays the regression estimates that were used to construct Table 2 above. In particular, linear combinations of coefficients from this regression provide estimates of how the ability to learn from the past play of one's counterparts varies across the AQ distribution reported above. Note that we observe large, positive and significant coefficients on K_{jt} and S^*_{ijt}, while the interactions between those variables and AQ score are negative and significant, explaining the declining marginal effect of those variables across the AQ distribution reported in Table 2.

P1 Chose SPE	(1)
Period	0.002***
	(0.001)
AQ Score	-0.000
	(0.005)
Has Dominant Strategy	0.491***
	(0.123)
Has Dominant Strategy × AQ Score	-0.000
	(0.006)
"Highest Mean" Heuristic = SPE	0.329**
	(0.116)
"Highest Mean" Heuristic = SPE × AQ Score	0.002
	(0.006)
Fraction i's Counterparts Chose Dominant$_{t-1}$ (D_{ijt})	0.384**
	(0.136)
Game Played Previously (K_{jt})	0.237**
	(0.092)
Game Played Previously (K_{jt}) × AQ Score	-0.009**
	(0.004)
SPE Directly Implied (by i's private history) (S_{ijt})	0.139
	(0.081)
SPE Directly Implied (by i's private history) (S_{ijt}) × AQ Score	-0.002
	(0.004)
SPE Indirectly Implied (by i's private history) (S^*_{ijt})	0.340**
	(0.139)
SPE Indirectly Implied (by i's private history) (S^*_{ijt}) × AQ Score	-0.016**
	(0.006)
Constant	0.060
	(0.170)
Observations	7743
R-Sq.	0.109

Clustered standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table F2: **OLS Regression Analysis of SPE Choices Across the AQ Distribution.**

F.1. Alternative Analysis of Theory of Mind

Here we report an alternative way of showing the relationship between our subjects' capacity to learn the *SPE* action and theory of mind as it is understood by psychologists. We estimate learning rates separately for each Player 1 with OLS regressions where the dependent variable takes the value of 1 when the player makes the *SPE* choice and 0 otherwise, and the relevant independent variables are our measures of the information available to Player 1 subject i about the *SPE* from previous direct and indirect revelations of Player 2 preferences (S_{ijt} and S^*_{ijt} as described above). The β coefficients on these variables provide estimates of each individual's ability to use direct and indirect revelation of preferences in order to make the *SPE* choice. In these regressions, we also include a constant term, dummy variables that take a value of 1 when Player 1 had a dominant action or when the "highest mean" heuristic yields the *SPE* choice, as well as the additional controls K_{jt} and D_{ijt}. As a further robustness check, we also estimate in a separate specification the β coefficient on the variable $\max(K_{jt}, S_{ijt}, S^*_{ijt})$, again including the other controls; this coefficient gives an estimate of a more general learning rate from past experience.

We then compute simple correlation coefficients between estimated learning rates and measures of theory of mind from the AQ and BAP questionnaires. On both survey instruments, a higher score indicates increased presence of autism spectrum behaviors. Thus, negative correlations will indicate that our concept of theory of preferences is analogous to the information in the AQ and BAP surveys, while the absence of correlation or positive correlations will indicate otherwise.

	S_{ijt}	S^*_{ijt}	$\max(K_t, S_{ijt}, S^*_{ijt})$
BAP	-0.08	-0.08	-0.22**
BAP_Rigid	-0.14	0	-0.19*
BAP_Aloof	-0.01	0.02	-0.14
BAP_Prag	-0.05	-0.24**	-0.17
AQ	-0.05	-0.21*	-0.29***
AQ_Social	-0.04	-0.24**	-0.2*
AQ_Switch	-0.06	0	-0.22**
AQ_Detail	-0.05	-0.08	-0.13
AQ_Commun	-0.09	-0.21*	-0.24**
AQ_Imagin	0.12	-0.06	0.04

Table F3: **Correlations between Autism Spectrum Measures and Individual Learning from Direct and Indirect Preference Revelation (S_{jt}, S^*_{jt}).** BAP and AQ are overall scores from each instrument. Other variables are individual scores on subscales of each instrument. BAP_Rigid = Rigidity, BAP_Aloof = Aloofness, BAP_Prag = Pragmatic Language Deficit, AQ_Social = Social Skills, AQ_Switch = Attention Switching, AQ_Detail = Attention to Detail, AQ_Commun = Communication Skills, and AQ_Imagin = Imagination. Reported tests are one-tailed.

Table F3 reports these simple correlations between measures from our experiment and survey measures of autism spectrum intensity.[29] From the table, we can see that learning rates from direct revelations are only weakly negatively correlated with the BAP and AQ

[29] Figures F4, F5, and F6 display scatterplots of these correlations for the AQ and BAP scores.

scores and some of the subscales.[30] The correlations are stronger for the learning rates from indirect revelations, and stronger still for the variable $\max(K_{jt}, S_{ijt}, S^*_{ijt})$. This provides further evidence that our measure of theory of preferences corresponds to the notion of theory of mind, as it is used in psychology.

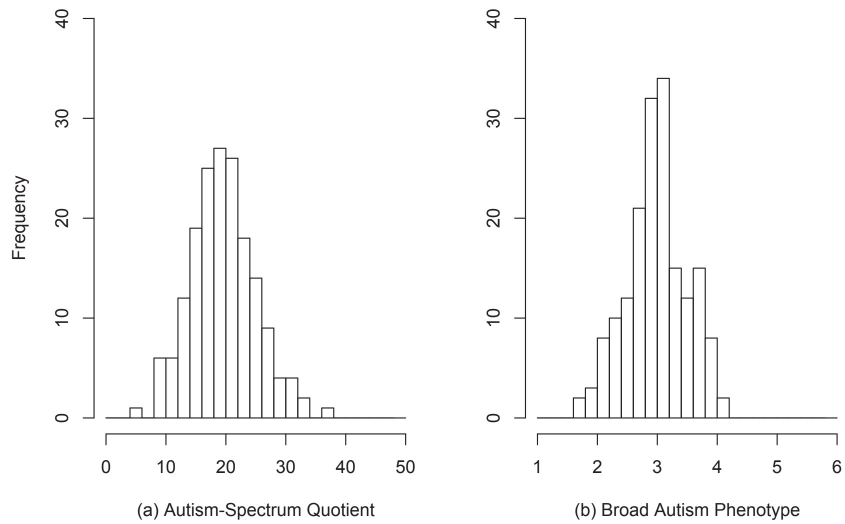

FIGURE F3: **Histograms of AQ and BAP Scores.** This includes the scores of player 2s, which do not feature in the analysis elsewhere. Each panel shows the range of possible scores.

[30] Following convention, the BAP score is the mean of the three BAP subscale scores, and the AQ score is the sum of the five AQ subscale scores.

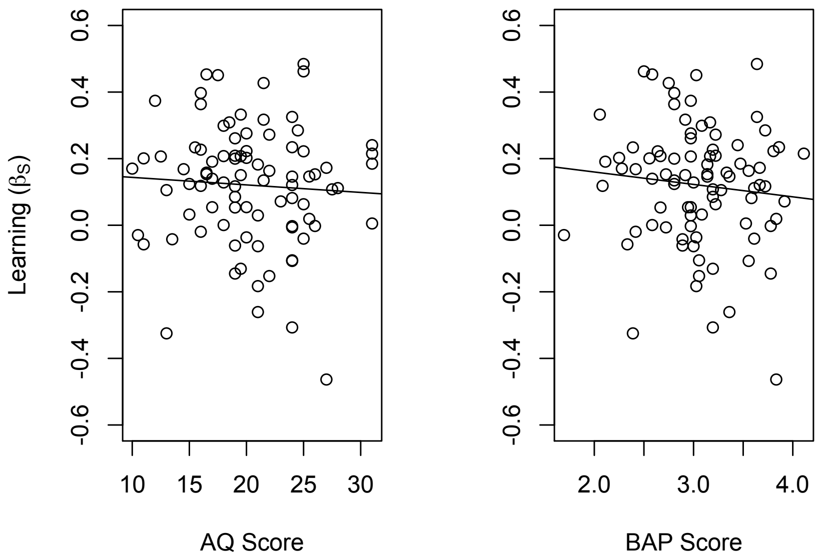

FIGURE F4: **Scatterplots Comparing Personal Direct Revelation (S_{ijt}) Learning Rates to Theory of Mind Measures from Psychology.** The solid lines plot OLS fits of the data.

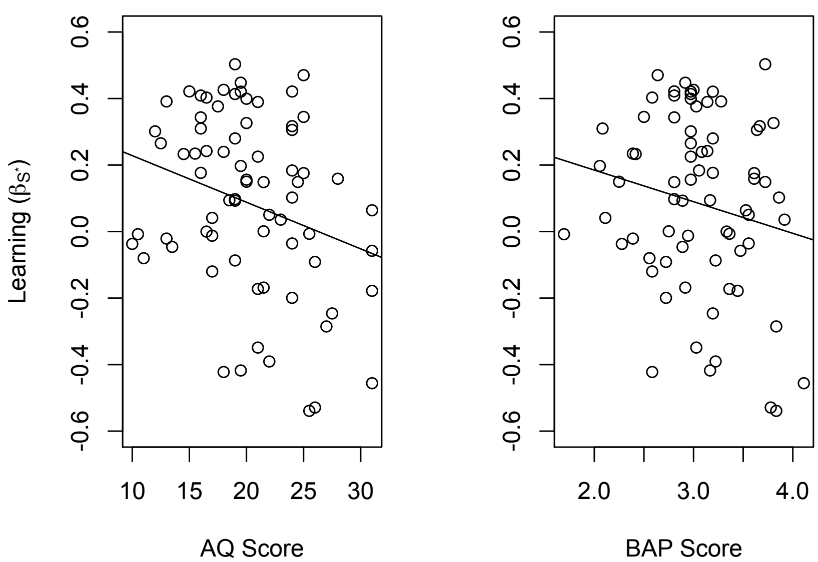

FIGURE F5: **Scatterplots Comparing Personal Indirect Revelation (S^*_{ijt}) Learning Rates to Theory of Mind Measures from Psychology.** The solid lines plot OLS fits of the data.

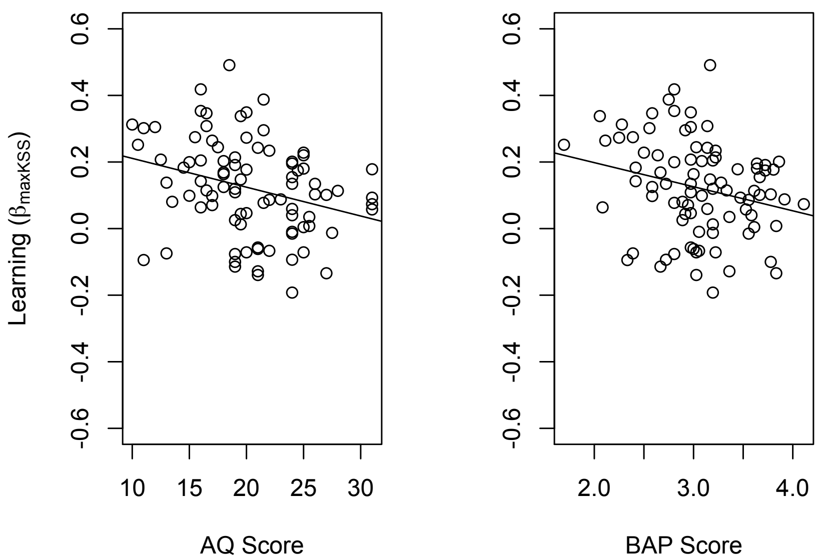

FIGURE F6: **Scatterplots Comparing** $(\max(K_{jt}, S_{ijt}, S^*_{ijt}))$ **Learning Rates to Theory of Mind Measures from Psychology.** The solid lines plot OLS fits of the data.