Purdue University Economics Department Working Paper No **1321**

Correlated beliefs: Predicting outcomes in 2×2 games

Timothy N. Cason, Tridib Sharma and Radovan Vadovič*

November 4, 2019

Abstract

Studies of strategic sophistication in experimental normal form games commonly assume that subjects' beliefs are consistent with independent choice. This paper examines whether beliefs are consistent with correlated choice. Players play a sequence of 2×2 normal form games with distinct opponents and no feedback. Another set of players, called predictors, report a likelihood ranking over possible outcomes. A substantial proportion of the reported rankings are consistent with the predictors believing that the choice of actions in the 2×2 game are correlated. Predictions seem to be correlated around focal outcomes and the extent of correlation over action profiles varies systematically between the type of games (i.e., prisoner's dilemma, stag hunt, coordination, and strictly competitive).

1 Introduction

A 2×2 game is played in a laboratory by randomly matched anonymous subjects. Should one expect their actions to be correlated or independent? To

*TC: Purdue University; cason@purdue.edu. TS: Instituto Tecnológico Autónomo de México (ITAM) - Centro de Investigación Económica (CIE); sharma@itam.mx. RV: Carleton University; radovan.vadovic@carleton.ca. We thank Martin Dufwenberg, Piotr Evdokimov, Andrei Gomberg, Yoram Halevy, P.J. Healy, Guillaume Frechette, Dan Friedman, Vai-Lam Mui, Roman Pancs, Yaroslav Rosokha, Levent Ülkü, John Wooders, an anonymous Advisory Editor, referee, and audiences at several university seminars and conferences for valuable comments. TS thanks Asociación Méxicana de Cultura for financial support.

iteratively delete dominated strategies, subjects need to have mutual knowledge of rationality and the game. For Nash equilibrium, players additionally need mutual knowledge of conjectures (Aumann and Brandenburger, 1995). Anonymous subjects who do not communicate may not satisfy these requirements (see for example Cason and Sharma, 2007). Even correlated equilibria may require some form of interaction for the formation of belief hierarchies (Brandenburger and Friedenberg, 2008), or for the acceptance of and adherence to a common coordination device (Aumann, 1974). On the other hand, even naive agents may learn to coordinate their actions over time (e.g., Foster and Vohra, 1997; Hart and Mas-Colell, 2000). Such learning could lead to coordinated actions in the laboratory.

This paper reports on an experiment in which subjects predict outcomes in 2×2 games played by others. We ask: Do predictors believe that players are able to coordinate their actions? In other words, are the predicted distributions over outcomes consistent with independent or correlated choice of actions?

Predictions consistent with independent choice of actions are given the best chance of succeeding. Our subjects interact anonymously and independently. We develop a novel procedure to allow subjects to verify that they are matched independently. Other subjects, aware of this matching procedure, predict outcomes. Surprisingly, we find overwhelming support for predictions that are consistent with correlated actions.

To get at beliefs (behind the predictions) in a clean and direct manner, we choose a framework where the predictors' actions do not affect the payoffs of other players. This eliminates incentives to hedge in reporting beliefs. Even then, eliciting probability distributions is a complex task (Schlag, Tremewan and Van de Weele, 2015; Hossain and Okui, 2013). Hence, we elicit beliefs in the form of likelihood rankings. Likelihood rankings are easier to elicit using a simple incentive scheme. All we require is that the predictor's preferences respect first order stochastic dominance (see Section 2.1). Although likelihood rankings provide a coarse measure of underlying probability beliefs, there exists a class of rankings that can only result from the predictor believing that players' actions are correlated (see Fact 2 in Section 2.2).

To our knowledge, our data provide the cleanest evidence to date on beliefs about correlated actions.[1] In our experiment, 24 subjects (players)

[1] In an experiment conducted online, Rubinstein and Salant (2014) document correlation in beliefs over outcomes of a hawk-dove game by asking directly for estimates of the

were paired up and interacted with a new opponent (i.e., perfect stranger matching) in 11 different 2×2 games. They received no feedback between games, and outcomes were revealed at the end of a session through a credible "public" procedure (see Section 3). We included prisoner's dilemma games, stag-hunt games, coordination games and strictly competitive games. We call this session the Behavior session. A set of 53 different subjects (predictors) were then asked to rank the likelihood of outcomes in the 11 games. Predictors went through the same set of instructions and practice rounds as the actual players in the initial (Behavior) session. The instructions included photographs documenting the procedures, to highlight the independence of the row and column player choices. Nevertheless, in most games at least 60% of predictors stated rankings that were consistent with only correlated actions. Interestingly, such rankings were rarely stated in a matching pennies game.[2] We also find that more predictors expected correlated behavior in games with multiple Nash equilibria. However, in the Behavior session, players' actions were not correlated in the manner predicted.

As a control, we also had another set of 48 subjects first play the games themselves. They then predicted outcomes of other players in the original session. Their predictions were also overwhelmingly consistent with correlated actions. We also ran additional control sessions with another 46 new predictors, but where the outcomes of most games were not determined by human players but rather by random draws from two bingo cages (one representing the row and the other the column player). In these control sessions predicting bingo cage outcomes, correlated rankings accounted for only about 4-11% of total reports. This indicates that predictors associate correlated actions with humans but not with innate bingo cages. The result also indicates that subjects are perfectly capable of reporting independent rankings.

frequency distribution over outcomes. Their procedure of eliciting beliefs did not include salient rewards as in a controlled lab experiment, but their intriguing data set provides important suggestive evidence. Our experiment includes a novel procedure for eliciting beliefs that is properly incentivized and cognitively simpler than asking for probabilities or outcome frequencies. We also examine correlation across eleven different games chosen to investigate the relationship between belief correlation and payoff asymmetry, off-equilibrium payoff magnitudes, and multiple Nash equilibria in pure strategies.

[2]The unique correlated equilibrium outcome in the matching pennies game is the same as that of the unique Nash equilibrium.

1.1 Related Literature

There is growing empirical support for the fact that subjects act as if choices of others are correlated. Ho, Camerer and Weigelt (1998) study learning dynamics in p-beauty contest games using a Stahl and Wilson (1995)-type model. They observe that allowing beliefs of players to be correlated, substantially improves overall fit relative to the restriction of independent beliefs. Costa-Gomes, Crawford and Iriberri (2009) use data from the Van Huyck, Battalio and Biel (1990, 1991) coordination games to evaluate the performance of some leading behavioral models, including the quantal response equilibrium, the level-k/cognitive hierarchy model and the noisy introspection model. Similar to Ho *et al.*, they allow players to hold correlated beliefs. For all of their model estimates, beliefs consistent with correlated actions fit better than those with independent actions.[3]

Beliefs consistent with correlated actions may lead to interesting outcomes. In finitely repeated anonymous games, players often cooperate in experimental settings even when they should not. Healy (2007) shows that this behavior is consistent with *stereotyping*–letting beliefs over types of actions be correlated. Rubinstein and Salant (2014) show that decisions in an entry game can change dramatically with correlated beliefs. Actions, correlated through "sunspots," could possibly lead to bank runs (Duffy and Fisher, 2005; Ennis and Keister, 2010).

In the experimental literature, Moreno and Wooders (1998) induce correlated equilibria via pre-play communication. Cason and Sharma (2007) show that correlated equilibria can be induced through mediated private signals, as long as there is mutual knowledge of rationality. Duffy and Feltovich (2010) show that for a mediator's suggestions to be followed, they have to be consistent with correlated equilibrium. Conversely, Guillen and Hing (2014) find that subjects respond to mediated messages even when they should not. Palfrey and Pogorelskiy (2019) find support for correlated equilibrium in a voter turnout game with communication within parties.

Somewhat related is the recent work on the consensus effect, where individuals believe (at times falsely) that others make choices similar to their own. Offerman et al. (1996) and Iriberri and Rey Biel (2013) find support for

[3]In a decision-theoretic setting, Epstein and Halevy (2019) distinguish between source ambiguity (e.g., an ambiguous urn) and ambiguity regarding the relationship between sources (i.e., a degree of correlation between multiple ambiguous urns). They show that uncertainly about the relationship between sources is a determinant of ambiguity aversion.

the false consensus effect. Engelmann and Strobel (2000) find that the effect is not necessarily false. Consistent with the false consensus effect, Rubinstein and Salant (2016) observe that subjects are more likely to play hawk in the hawk-dove game when believing the opponent plays hawk and vice versa. Vanberg (2008) posits that the effect can be interpreted as an outcome of some correlated equilibria.

2 The Framework

In period 1, players R and C simultaneously choose from action sets $\{u, d\}$ and $\{l, r\}$ respectively, to receive payoffs as shown below:

Payoffs

	l	r
u	x_{11}, y_{11}	x_{12}, y_{12}
d	x_{21}, y_{21}	x_{22}, y_{22}

Belief distribution

	l	r
u	b_1	b_2
d	b_3	b_4

Let $O = \{(u, l), (u, r), (d, l), (d, r)\} \equiv \{o_1, o_2, o_3, o_4\}$ denote the set of outcomes and let $\hat{o} \in O$ be the realized outcome in period 1. The Predictor (P) does not observe $\hat{o}$. Let $b(o_i) = b_i$, for $o_i \in O$, be probabilities that represent P's beliefs over the outcome. We denote a belief distribution by $b = (b_1, b_2, b_3, b_4)$. Let B be the set of possible beliefs.

In period 2, P reports a likelihood ordering over O. For our purpose, a ranking (ordering) is a vector $k = (k_1, k_2, k_3, k_4)$, where $k_j \in \{1, 2, 3, 4\}$ denotes the likelihood rank of outcome o_j. For all k, there exists an element k_i which takes the value of 1. Furthermore, if $k_i \neq 1$ is an element of k, then there exists k_j in k which takes the value $k_i - 1$. That is, non consecutive ranks are not allowed. For example, $k = (3, 1, 2, 4)$ is a likelihood ranking which indicates that the outcome $(u, r) \equiv o_2$ is most likely to occur, followed by $(d, l) \equiv o_3$, $(u, l) \equiv o_1$ and $(d, r) \equiv o_4$. We allow for ties in reported rankings, i.e. the ranking $(2, 1, 1, 3)$ is allowed. Let K be the set of all rankings. For a given k let $t(k) = (t_1, t_2, t_3, t_4)$ be the collection of distinct outcomes such that t_1 is the outcome with the highest rank in k; t_2 is the outcome with the second highest rank, and so on.

Given a ranking vector selected by P, her payoff is determined as:

Ranked outcome/actual outcome:	Payoff:
$t_1 = \hat{o}$	π_1
$t_2 = \hat{o}$	π_2
$t_3 = \hat{o}$	π_3
$t_4 = \hat{o}$	π_4

where $\pi_i > \pi_{i+1}$. If the selected ranking includes a tie, and this tied ranking matches the actual outcome, then the payoff equals one of the corresponding rewards, each chosen with equal probability. For example let $k = (2, 1, 1, 3)$ be the chosen ranking, and suppose $\hat{o} = (d, l)$. As the chosen outcome matched one of the two outcomes ranked highest by P, her payoff would be either π_1 or π_2, each chosen with equal probability. This completes the description of the game.

2.1 Predictions

P forms beliefs over O, on the basis of some behavioral model. For example P may believe that R and C coordinate on some focal point, but P does not know which one. Or it could be that P believes that the players cannot coordinate and hence choose their actions randomly. We will remain agnostic about P's model, and let it belong to a very general class. A minimum restriction on this class will allow us to distinguish between P believing that period 1 actions are correlated or independent.

P believes that there is a set of "states of the world" given by $\Omega = \{\omega_1, \omega_2\}$, with priors: $\Pr(\omega_i) = \theta_i$ for $i \in \{1, 2\}$.[4] Conditional on ω_i, P believes that player R receives signal $s_1^i \in \{u', d'\}$ and C receives signal $s_2^i \in \{l', r'\}$, for $i \in \{1, 2\}$. Players are believed to see the signal but not the state. P sees neither the state nor the signals.

Conditional on ω_i, signals s_1^i and s_2^i are believed to be independently drawn with $\Pr(s_1^i = u' \mid \omega_i) = p_u^i$ and $\Pr(s_2^i = l' \mid \omega_i) = p_l^i$, for $i \in \{1, 2\}$. After receiving the signal, P believes that R and C play according to the

[4] Aumann (1987, p. 16) notes that a predictor's perceived "correlation may ... [have] no connection with any overt or even covert collusion" between players. The states in our model could refer to some shared history among the players that is unknown to our predictor, and this could give rise to correlation among their actions.

signal. This then generates a belief b over outcomes O, as follows.

	l	r
u	$b_1 = \theta_1 p_u^1 p_l^1 + \theta_2 p_u^2 p_l^2$	$b_2 = \theta_1 p_u^1 p_r^1 + \theta_2 p_u^2 p_r^2$
d	$b_3 = \theta_1 p_d^1 p_l^1 + \theta_2 p_d^2 p_l^2$	$b_4 = \theta_1 p_d^1 p_r^1 + \theta_2 p_d^2 p_r^2$

This class of models can generate any belief of P's, as we have not taken a stand on R and C's payoffs or rationality. However, specific values of θ_i, p_u^i and p_l^i, $i \in \{1, 2\}$, will be consistent with beliefs that regard players actions as correlated or independent. This is stated as Fact 1 below. The proof is simple and hence omitted.

Fact 1: *The belief distribution b is consistent with independent choice of actions if and only if at least one of the following conditions is satisfied.*

$$(1)\ \theta_1 = 0 \text{ or } \theta_2 = 0$$
$$(2)\ p_l^1 = p_l^2 \text{ or } p_u^1 = p_u^2$$

Fact 1 tells us that if only one state is relevant for choice or states of the world are not relevant for choice (for at least one player), then P believes that players actions are independent (the converse also holds). This, for example, is consistent with P believing that players are able to coordinate on one particular Nash equilibrium or focal point. If Fact 1 does not hold then we shall say that P has multiple, state dependent, conjectures about players' choices. For example, let $\theta_1 = \theta_2 = \frac{1}{2}$, and $p_l^1 = 1$; $p_l^2 = 0$; $p_u^1 = 1$; and $p_u^2 = 0$. Fact 1, then, states that P believes that the actions of R and C are correlated. In particular, b is as follows:

	l	r
u	$b_1 = \frac{1}{2}$	$b_2 = 0$
d	$b_3 = 0$	$b_4 = \frac{1}{2}$

The discussion above justifies the following assumption.

Assumption 1: *P's belief is a distribution over outcomes induced by some behavioral theory of the Period 1 stage game.*

Slightly abusing notation, let the function k,

$$k : B \to K$$

denote a likelihood representation of P's beliefs. Recall that B is the set of all beliefs. We assume that P has *consistent* rankings.[5]

Definition 1: *Ranking k is consistent with belief b if: $[b_i > b_j] \Longrightarrow [k_i < k_j]$ and $[b_i = b_j] \Longrightarrow [k_i = k_j]$.*

We now show that, given the payoffs of P defined previously, our elicitation process makes truthful revelation of the likelihood ranking a (weakly) dominant strategy.

Assume that P has preferences that respect (first order stochastic dominance) FOSD ordering. Let w.l.g., P's belief be such that $b_1 \geq b_2 \geq b_3 \geq b_4$. Hence P's true ranking is $k = (k_1, k_2, k_3, k_4)$ where $k_1 \geq k_2 \geq k_3 \geq k_4$. Reporting k induces a distribution over $\{\pi_1, \pi_2, \pi_3, \pi_4\}$ such that $\Pr(\pi_i) = b_i$. Name this distribution F. Now suppose P reports k' instead, where $k' \neq k$. Then, k' induces a distribution such that $\Pr(\pi_i) = b'_i$ where (b'_1, b'_2, b'_3, b'_4) is a permutation of (b_1, b_2, b_3, b_4). Name this distribution F'. As $b_l = \min\{b_i \mid i \leq l\}$, we have $\sum\limits_{i=l}^{4} b'_i \geq \sum\limits_{i=l}^{4} b_i$ for all $l \in \{1, 2, 3, 4\}$. Hence, due to FOSD, F' cannot be strictly preferred to F. Assuming P reports the truth when she is indifferent, we have that reporting k is a weakly dominant strategy.

2.2 Correlated Rankings and Hypotheses.

Rankings are a coarse representation of beliefs in the following sense: two distributions, one consistent with independent actions and the other correlated actions, may generate the same ranking. As an example consider $b' = (\frac{4}{9}, \frac{2}{9}, \frac{2}{9}, \frac{1}{9})$ and $b'' = (\frac{3}{9}, \frac{25}{90}, \frac{25}{90}, \frac{1}{9})$. Both are represented by the ranking $k = (1, 2, 2, 3)$, yet b' is consistent with independent choice while b'' is consistent with correlated choice.[6] However, certain rankings are generated by correlated and only correlated distributions.

[5] Chapter 3 of Fishburn (1970) has details on the representation of likelihood orderings by probability distributions.

[6] Of the 75 possible rankings that we elicit in our experiment, 49 (65%) would be classified as independent.

Fact 2. *If k is consistent with belief b and: (i) $k_1 < k_2$* and $k_4 < k_3$, *or; (ii) $k_2 < k_1$* and $k_3 < k_4$, *then b is consistent with correlated choice of actions.*

Proof: We prove (i), and the proof of (ii) is similar. Suppose $k_1 < k_2$ and $k_4 < k_3$. Since k is consistent with b we have that $b_1 > b_2$ and $b_4 > b_3$. Suppose b is independent, then by definition of independence,

$$\begin{aligned}&b_1 > b_2\\&\quad \Longleftrightarrow (b_1 + b_2)(b_1 + b_3) > (b_1 + b_2)(b_2 + b_4)\\&\quad \Longleftrightarrow (b_1 + b_3) > (b_2 + b_4) \text{ as } (b_1 + b_2) > 0.\end{aligned}$$

Similarly,

$$\begin{aligned}&b_4 > b_3\\&\quad \Longleftrightarrow (b_2 + b_4)(b_3 + b_4) > (b_1 + b_3)(b_3 + b_4)\\&\quad \Longleftrightarrow (b_2 + b_4) > (b_1 + b_3) \text{ as } (b_3 + b_4) > 0.\end{aligned}$$

But we cannot have $(b_1 + b_3) > (b_2 + b_4)$ and $(b_2 + b_4) > (b_1 + b_3)$. *QED*

All other distinct rankings are consistent with at least one independent belief b. This is easy to show. We provide just one example. Consider k such that $k_1 > k_2 > k_3 > k_4$. This ranking is consistent with the independent distribution $b = (\frac{20}{49}, \frac{15}{49}, \frac{8}{49}, \frac{6}{49})$.

Fact 2 motivates the following definitions.

Definition 2. *Ranking $k = (k_1, k_2, k_3, k_4)$ is said to be correlated if: either (i) $k_1 < k_2$* and $k_4 < k_3$, *or; (ii) $k_2 < k_1$* and $k_3 < k_4$

A subset of correlated rankings is intuitively appealing. In these rankings, the highest ranks are provided to either (u, l) and (d, r) or (u, r) and (d, l) (i.e. the diagonal outcomes).

Definition 3. *$k = (k_1, k_2, k_3, k_4)$ is said to be diagonally correlated if: either (i) $k_1 < \min\{k_2, k_3\}$* and $k_4 < \min\{k_2, k_3\}$, *or; (ii) $k_2 < \min\{k_1, k_4\}$* and $k_3 < \min\{k_1, k_4\}$.

Definition 4. *$k = (k_1, k_2, k_3, k_4)$ is said to be circularly correlated if: either (i) $k_1 < k_2 < k_4 < k_3$, or; (ii) $k_2 < k_1 < k_3 < k_4$.*

Rankings that satisfy Definitions 3 or 4 exhaust the set of correlated rankings. In passing we note that k, such that $k_1 = k_2 = k_3 = k_4 = 1$, is consistent with only one distribution $b = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$, which is independent.

Definition 5. *$k = (k_1, k_2, k_3, k_4)$ is said to an independent ranking if k is not diagonally or circularly correlated.*

Using Fact 1 and Definitions 1 through 5 we put forth two hypotheses. *Single conjecture hypothesis (H-s):* P's conjecture is contingent on only one (or no) state of nature, and so she reports an independent ranking. *Multiple conjectures hypothesis (H-m):* P's conjectures are contingent on both the states in a manner that she reports a diagonally correlated ranking.

3 Experimental Design and Procedures

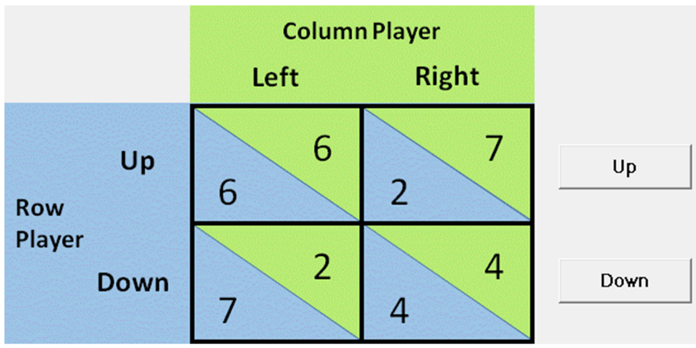
In the Behavior part subjects played a sequence of eleven 2×2 games.[7] Games were presented in normal form. Each subject was assigned a role of either the Row or the Column player and kept the same role in all games. To ease comprehension, we presented the games with color-coded roles and strategies as shown in Figure 1.

For a clean test of our hypotheses it was important to make clear that there were no repeated interactions and that matching could not possibly depend on subjects' previous choices.[8] To address the first concern we used perfect stranger matching. The Behavior session had 24 subjects play the 12 games with different opponents, with no feedback provided between games. To address the second concern we developed an experimental procedure to make the matching explicit and ex-post verifiable. In particular, each subject was identified by an ID number, and a record sheet at the top of their computer screen showed throughout the experiment the predetermined matched partners' IDs as well as the subject's own actions for all 12 games. At the end of the experiment, the experimenter evaluated the games publicly by listing the matches and corresponding actions on the main projector screen. Subjects were encouraged to verify that the information on the projector screen (i.e., their actions and IDs of their opponents) indeed matched the

[7]Another 3×3 game was added at the end of the sequence of 2×2 games as a matter of curiosity. Following this game subjects also responded to one framed investment game. The results from these last two games are not analyzed in this paper.

[8]If this were not the case, then actions could be correlated via the matching procedure.

Figure 1: Game presentation



information displayed in their own record sheets. This procedure ensured that the experimenter could not secretly manipulate the matching.

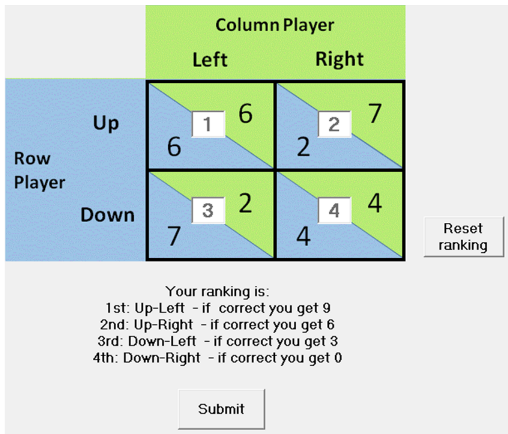
In the Prediction part, the primary task was to predict the outcome of play for a randomly selected pair of players from each game of the Behavior session. Predictors were asked to rank outcomes in terms of likelihood of occurrence. They were to assign numbers from 1 (most likely outcome) to 4 (least likely outcome) in the fields provided inside each cell of the decision matrix (see Figure 2). As mentioned in Section 2.1, ties were allowed.

For each game we randomly selected a pair of subjects from the Behavior session. Their choices represented the *actual outcome* ($\hat{o}$ in Section 2.1). The predicting subject then received a reward according to the following schedule:

Guess/Outcome:	Earning:
The most likely guess = actual outcome	9
Second most likely guess = actual outcome	6
Third most likely guess = actual outcome	3
Least likely guess = actual outcome	0

We allowed for indifference by letting subjects assign the same rank to multiple outcomes. If the actual outcome matched one of the tied outcomes,

Figure 2: Ranking task



then the payoff was randomly chosen from the corresponding rewards.[9] Assuming identical play across games, it follows from Section 2.1 that our payment structure induces truthful revelation of rankings.[10]

The 11 2×2 games are listed below. The games were not presented in this logical order and row/column orderings were also varied so that the equilibria did not always fall on the main (*ul*-*dr*) diagonal.

PD-S

6 , 6	2 , 7
7 , 2	4 , 4

PD-P

5 , 5	1 , 9
9 , 1	2 , 2

PD-A

5 , 6	2 , 9
7 , 1	3 , 2

SH-S

6 , 6	2 , 5
5 , 2	4 , 4

SH-P

7 , 7	0 , 3
3 , 0	4 , 4

SH-A

7 , 6	1 , 5
6 , 2	3 , 4

[9]The following example illustrates the procedure: if the ranking was as $k = (1, 2, 2, 3)$ and if the actual outcome happened to be (d, l), then the earning was either 6 or 3, each with probability 1/2.

[10]To aid subjects in ranking outcomes, the software guided them through the process using a step-wise procedure. Subjects assigned ranks incrementally from the most likely to the least likely (allowing for possible indifference).

CO-S

6 , 6	1 , 1
1 , 1	6 , 6

CO-A1

7 , 3	1 , 1
1 , 1	4 , 4

CO-A2

3 , 7	3 , 3
2 , 2	6 , 4

SC-S

6 , 2	2 , 6
2 , 6	6 , 2

SC-D

8 , 4	1 , 5
1 , 5	8 , 4

We included four types of games: three prisoner's dilemma (PD) games, three stag-hunt (SH) games, three coordination (CO) games and two strictly competitive games (SC). The PD and SH games included a symmetric version (S), a version with high off-diagonal ("temptation") payoffs (P), and an asymmetric version (A). We included one symmetric and two asymmetric versions of the coordination games, and a symmetric (S) and asymmetric version with higher off-diagonal payments (D) for the SC games.

The games were chosen to generate variation to test the hypotheses outlined in the preceding section. We conjectured that beliefs that are consistent with correlated choice of actions are best identified in games that allow for two distinct and mutually exclusive conjectures about how the game would be played. For example, in the stag-hunt game the conjectures could correspond to the two PNE's that are on the diagonal. When both equilibria (focal points) are equally compelling, they could be assigned the highest rank by P. We varied the monetary payoffs across games in ways that at times magnified and at times minimized the plausibility of multiple conjectures.

In addition to eliciting rankings, we also sought to gauge subjects' level of confidence in their stated rankings. We asked predictors to guess how many of the 12 pairs from the Behavior session played the predictor's top ranked outcome. To see why this information is useful, notice that a prediction that 10 to 12 pairs played the most likely outcome indicates that the predictor is largely indifferent between the remaining three outcomes. In each game, this guessing task immediately followed the ranking task. We incentivized guessing by paying 5 experimental currency units for a correct response.[11]

The Behavior part began with an experimenter reading the instructions aloud (see Appendix B) and subjects following along on their own hardcopy. Subjects then completed four unpaid practice rounds to became familiar with

[11]To avoid incentives for hedging we only allowed subjects to make a guess for the outcome(s) that they ranked as the most likely.

the computer interface and understand the experimental procedures. Each practice round involved control questions that had to be completed correctly before the experiment was allowed to continue.

Our main prediction sessions were of two different types:

(i) **Only-predict:** Several groups of subjects (53 in total) participated in the Prediction part only. They predicted the outcomes of play from the Behavior session.

(ii) **Play-and-predict:** Two groups of 24 subjects (48 in total) participated in both parts of the experiment in a single session. In the Prediction part they did not predict the outcomes of their own group's play, but rather the outcomes of subjects in the original Behavior session.

It was important that everyone properly understood the decision problem in the Behavior session. So, we began each session with the same instructions and the same practice rounds as in the Behavior part.[12] In the Only-predict sessions this was followed by the instructions on ranking outcomes and three practice rounds with control questions involving the ranking task. In the Play-and-predict sessions, subjects first played the games and only then were instructed on how to make predictions.[13]

We also ran control sessions to rule out the possibility that correlated rankings are driven by misconceptions regarding the matching procedure. Similar to the main experiment, the control part also had a single Behavior session and four Prediction sessions. The Behavior session had 9 games (instead of 12) where draws from two bingo cages determined the games' outcomes.[14] One bingo cage made the draws for the row player and the other for the column player. Each bingo cage had 12 balls of one of two colors, with each color representing an action. A ball drawn from each of

[12]For authenticity purposes the instructions for all Prediction sessions included photographs taken during the Behavior session, which included photos of the public ex-post verification of the choice implementation and matching procedure.

[13]In each of the four Only-predict sessions, subjects encountered a different game presentation ordering shown in Appendix A.1 Table 5. This turned out to have no impact on the predictions and so in the two Play-and-Predict sessions we kept the same game ordering.

[14]Two human participants were recruited to observe the bingo cage drawings. They were paid the sum of games' payoffs as determined by one randomly chosen pair of bingo balls for each game. This procedure provided a natural justification for the inclusion of the games' payoffs in their presentation in some of the prediction sessions.

the two bingo cages then resulted in an action pair that determined the outcome. Drawing were made with replacement. The number of “Up” and “Down” balls in the row bingo cage and the number of “Left” and “Right” balls in the column bingo cage corresponded to the marginal frequencies of play obtained for different games in the original (human) Behavior session.[15]

Forty six subjects participated in the bingo cage Prediction sessions. They ranked outcomes for 12 games. Nine games corresponded to the 9 games “played” through bingo cages and 3 games were selected from the original human Behavior session.[16] Two sessions had the block of 9 bingo cage games first, followed by the block of 3 human player games. Another two sessions were run with the reverse ordering. Two of the four control sessions (one in each order) displayed the game payoffs in the bingo cage prediction games even though this information is irrelevant for the ranking task. In the other two control sessions the payoffs were not visible to the predicting subjects. This is to address a possible concern that the size and relative standing of payoffs in individual cells could be nudging subjects toward some simple ranking heuristic, such those based on equity and/or efficiency considerations. It turns out, however, that the control session predictions were not sensitive to either the presentation ordering or the display of the game payoffs.

All 173 participants were undergraduate students with a variety of academic backgrounds at Purdue University, recruited from a database of about 3000 subjects using ORSEE (Greiner, 2015).[17] The software was programmed in Visual Basic. The Behavior session lasted about 30 minutes, and the Prediction sessions lasted about 50 (for Only-predict) to 60 minutes for (Play-and-predict). At the end of each session subjects were paid privately in cash, one USD for every 4 experimental currency units, plus a 5 USD show-up payment. Subjects earned on average 21 USD in the Behavior session, 17 USD in the Only-predict sessions and 25 USD in the Play-and-predict sessions.

[15] Two pairs of games produced identical frequencies. Instead of doubling up, we only used 9 instead of 11 games.

[16] We wanted each game chosen from the human behavior session to represent a different strategic setting (PD, SH, SC). From each category we also chose a game with the highest frequency of correlated rankings: PD-S, SH-P and SC-D.

[17] Only 10 out of 147 (6.8%) predicting subjects reported economics as their major. This means a possible background in game theory could not be a major factor affecting our results.

4 Results

We start by comparing data from Only-predict and Play-and-predict sessions. Table 6 in Appendix A shows very similar prediction patterns in the two types of sessions. For the SH-S game, the frequency of correlated predictions is lower in the Play-and-predict sessions (54%) than in the Only-predict sessions (74%). For all other games, we cannot reject the hypothesis that predictions were generated by the same process. Therefore, in the rest of the analysis we pool the prediction data across these session types.

The rest of this section is divided into five parts. First, we summarize the data on rankings. Second, we construct an empirical model and formally test whether the underlying beliefs are more likely to be consistent with correlated or independent actions. Third, we describe behavior from the Behavior sessions. Fourth, we compare rankings over outcomes generated by the actions of human subjects to those generated by draws from bingo cages. Last, we provide observations on the systematic variation in correlated rankings across games.

4.1 Correlated rankings: main sessions

The left panel of Figure 3 shows the proportion of correlated rankings for each of the eleven games. Each bar is divided into three parts. The dark-colored part refers to the proportion of *diagonally correlated rankings* along the *main* (*ul-dr*) diagonal. The lighter color refers to the proportion of *diagonally correlated rankings* along the *off* (*ur-dl*) diagonal. The uncolored part refers to *circularly correlated rankings*. The right side of the figure displays correlation rates for the bingo cage control sessions, discussed below in Section 4.2.

The bar chart shows a substantial number of correlated rankings, and most of these are *diagonally correlated rankings*. However, the proportions vary considerably between games. Predictors report diagonally correlated rankings most frequently in the stag hunt (SH) and pure coordination (CO) games, which have two pure strategy Nash equilibria in monetary payoffs. Interestingly, correlated rankings are less frequent in the prisoner's dilemma games and least frequent in the symmetric version of the strictly competitive game. This suggests that behavior, in these games is more consistent with the single conjecture hypothesis. The Mann-Whitney tests shown in Table 1 indicate that proportions of diagonally correlated rankings vary significantly

Figure 3: Frequency of correlated rankings by game

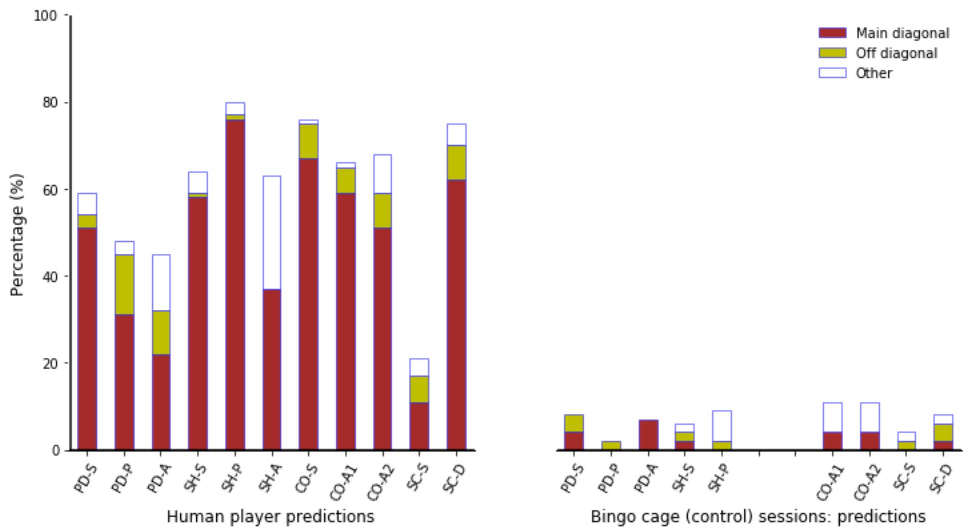


Table 1: Classification of rankings (101 predictors)

	PD			SH			CO			SC	
	S	P	A	S	P	A	S	A1	A2	S	D
Correlated											
Main dg.	51	32	22	59	77	37	68	60	51	11	63
Off dg.	3	14	10	1	1	0	8	6	8	6	8
Circular	5	3	13	5	3	26	1	1	10	4	5
Corr. total	59	49	45	65	80	63	77	67	69	21	76
Indep. total	42	52	56	36	21	38	24	34	32	80	25
Mann-Whitney Test p-values within game types (main diagonal correlation):											
	$p_{\text{S-P}} = 0.001$			$p_{\text{S-P}} = 0.002$			$p_{\text{S-A1}} = 0.106$			$p_{\text{S-D}} =$ 0.000	
	$p_{\text{S-A}} = 0.000$			$p_{\text{S-A}} = 0.001$			$p_{\text{S-A2}} = 0.002$				
	$p_{\text{P-A}} = 0.035$			$p_{\text{P-A}} = 0.000$			$p_{\text{A1-A2}} = 0.052$				

within all classes of games. This is most pronounced in cases of prisoner's dilemma games and strictly competitive games.

4.1.1 Testing for correlation

In this section we formulate an error model to determine rankings that best fit the data. In line with the left panel of Figure 3, we find a strong support for correlated rankings.

Suppose the ranking that the subject intends to report is $\rho = (\rho_1, \rho_2, \rho_3, \rho_4) \in K$. Given ρ, for every $k \in K$, define the Euclidean distance between ρ and k as: $d(k;\rho) = \sqrt{\sum_{i=1}^{4} (\rho_i - k_i)^2}$. Let $d^* = \max_{\rho,k \in K} d(k;\rho)$. For a given ρ, we assume that ranking k is chosen with probability

$$\Pr(k \mid \rho, \mu) = \frac{\exp\left(\frac{d^* - d(k;\rho)}{\mu}\right)}{\sum_{k \in K}\left[\exp\left(\frac{d^* - d(k;\rho)}{\mu}\right)\right]},$$

where $\mu > 0$ is a precision parameter. The idea here is that P, after due introspection, chooses to report her true ranking ρ. In the process, however, her hands "tremble" and she mistakenly picks a ranking k. Rankings, closer to ρ (in Euclidean distance) are chosen with higher probability.

Let D be the set of rankings reported in the experiment. The likelihood function for a given ρ and μ is

$$L(\rho, \mu) = \prod_{k \in D} \Pr(k \mid \rho, \mu)$$

We look for a pair ρ and μ that maximizes $L(\rho, \mu)$ in the domain $\rho \in K$ and $\mu \in [0, \mu']$, where μ' is some large number.[18,19]

[18]Much of the literature uses a logistic error model which perturbs actions in proportion to their implied relative payoffs, e.g., see Dal Bó and Fréchette (2011) for a recent approach in a strategic setting. Goeree and Holt (2004) use the logistic error structure to perturb players' beliefs. For our case, the expected payoff from individual rankings depends on the underlying belief distribution over the game's outcomes. A subject's payoff follows from how accurately her belief distribution predicts the realized outcome, and this depends on the behavior of two other players.

[19]For related approaches see, e.g., Costa-Gomes, Crawford Broseta (2001) for a simple error structure in strategic setting, or, e.g., Ivanov, Levin and Peck (2009) for sequential

As K is finite and $L(\rho, \mu)$ is continuous in μ, where μ belongs to a compact set, a maximum exists. We test for correlation on the main diagonal using a likelihood ratio test. This involves estimating the most likely ranking via maximum likelihood on the unrestricted domain (including all rankings) and then again on a restricted domain that includes only independent rankings.

One additional aspect of the data must be taken into consideration. Our method of eliciting rankings gives only a coarse description of the underlying beliefs. When beliefs are relatively extreme, such that most probability is assigned to a single outcome, then ordering the remaining outcomes may be noisy due to the predictor's near indifference. To measure this near indifference, recall that we rewarded subjects for accurate guesses about the number of pairs (0 to 12) who played the outcome that they designated as most likely. This gives us some idea about the shape of the underlying distribution, including the belief weight placed on lower-likelihood outcomes. In the following table we exclude rankings with the most extreme beliefs (guesses of 12 pairs or implausible beliefs of 2 or fewer pairs on the most likely outcome) from the analysis.[20]

Table 2 shows that in 9 of 11 games a correlated ranking maximizes the likelihood function. In 8 of those games the correlation is on the main diagonal. In the PD-A game the most likely ranking has circular correlation. For 9 of 11 games the best fitting ranking corresponds to the modal (most common) ranking (cf. Table 7 in Appendix A). When correlation is on the main diagonal the LR test is always highly significant.

Correlation on the main diagonal is mostly observed in games with two pure strategy Nash equilibria. In these games the two equilibria are likely to be focal, giving P two natural conjectures about how the game is played. The SH and CO games have two PNE on the main diagonal, the PD games we have one PNE (dr) and the SC games have no PNE. We therefore observe fewer rankings that are correlated on the main diagonal in PD and SC games

error assignment. In the sequential assignment model mistakes are made in assigning ranks outcome wise. Furthermore, the mistakes are not correlated across outcomes. This kind of independence assumption is not appropriate for our model. In our setup, the set K is restricted so ranks such as $(2, 2, 2, 2)$ are not feasible. If the subject assigns ranks $1, 2$ and 3 to the first three outcomes, she must assign rank of 3 or 4 to the fourth outcome.

[20]This excludes a total of 211 out of 1010 predictions (20.9%), about two-thirds of which represent highly confident predictions indicating all 12 pairs selecting the most likely outcome. The analysis was also conducted with the full sample and the results remain qualitatively unchanged, as shown in Appendix A.6.

Table 2: ML estimation of best-fitting ranking

	PD games			SH games		
	S	P	A	S	P	A
Rnk	$\begin{bmatrix}1&3\\3&2\end{bmatrix}$	$\begin{bmatrix}1&2\\2&3\end{bmatrix}$	$\begin{bmatrix}1&2\\4&3\end{bmatrix}$	$\begin{bmatrix}1&3\\3&2\end{bmatrix}$	$\begin{bmatrix}1&3\\3&2\end{bmatrix}$	$\begin{bmatrix}1&4\\3&2\end{bmatrix}$
μ	0.782	0.942	1.764	0.729	0.566	0.969
Obs.	89	94	92	68	56	74
$-\log L$	327.9	373	383.3	243.4	171.4	275.4
$-\log \tilde{L}$	365.5	373	383.7	272.2	219.4	294.4
p_{LR}	0.000	1	0.371	0.000	0.000	0.000
	Cor	*Ind*	*Cir*	*Cor*	*Cor*	*Cor*

	CO games			SC games	
	S	A1	A2	S	D
Rnk	$\begin{bmatrix}1&2\\2&1\end{bmatrix}$	$\begin{bmatrix}1&3\\3&2\end{bmatrix}$	$\begin{bmatrix}1&4\\3&2\end{bmatrix}$	$\begin{bmatrix}1&1\\1&1\end{bmatrix}$	$\begin{bmatrix}1&2\\2&1\end{bmatrix}$
μ	0.44	0.805	1.01	0.694	0.515
Obs.	80	94	99	92	93
$-\log L$	211.9	349.8	374	304.7	286
$-\log \tilde{L}$	267.2	371.6	410.7	304.7	340.4
p_{LR}	0.000	0.000	0.000	1	0.000
	Cor	*Cor*	*Cor*	*Ind*	*Cor*

Note: $\tilde{L}$ is the log-likelihood of the estimation restricted to independent rankings; p_{LR} refers to the likelihood ratio test p-value. *Cor/Cir/Ind* indicates whether the estimated ranking is diagonally correlated, circularly correlated or independent.

than in SH and CO games.

4.1.2 Behavior and predictions

We now look at the performance of predictions in light of the data generated in the Behavior sessions.

Table 3 shows the distribution of outcomes in the Behavior sessions, the rankings denote the rank of frequencies. A quick comparison of rankings in

Table 3: Behavior part: outcomes frequencies and corresponding rankings

	PD games						SH games					
	S		P		A		S		P		A	
Outcomes	10	14	4	12	4	7	16	9	22	4	16	9
	8	4	7	13	7	18	9	2	8	2	8	3
Ranking	2	1	4	2	3	2	1	2	1	3	1	2
	3	4	3	1	2	1	2	3	2	4	3	4
	Cir		*Ind*		*Ind*		*Ind*		*Ind*		*Ind*	

	CO games						SC games			
	S		A1		A2		S		D	
Outcomes	13	13	8	11	15	12	6	15	12	11
	7	3	5	12	6	3	3	12	11	2
Ranking	1	1	3	2	1	2	3	1	1	2
	2	3	4	1	3	4	4	2	2	3
	Cir		*Cir*		*Ind*		*Ind*		*Ind*	

Note: Outcomes refer to the frequency distribution of play among the 36 pairs in our sample (Behavior session and two Play-and-predict sessions). Ranking states the likelihood ranking for the corresponding outcome distribution in the rows immediately above.

Table 3 with the estimated rankings in Table 2 reveals that there is not a single match among the 11 games. Predictors correctly guess that the payoff dominant equilibrium is most frequent in the SH games, but they incorrectly predict that the cooperative outcome is most frequent in the PD games. Importantly, in no case is the behavior consistent with a correlated ranking on the main diagonal. A clear discrepancy exists between the behavior and prediction data.

4.2 Predictions: bingo cage control

The analysis of the previous subsection raises an important question. Are we sure that our subjects properly understood the implications of independence of play on outcomes? Correlated rankings could possibly be reported by predictors who could not perceive what independent matching meant; or, it could be that the presentation of payoffs nudged subjects in the direction of some ranking heuristic. It is also possible that some subjects misunderstood the ranking procedure.

In light of our stark results these concerns gain significance. To address them we ran several control sessions described in the experimental design section. The key difference between the main experiment and the control sessions was the nine games that subjects predicted with outcomes determined by random draws from two bingo cages. Subjects were informed about the chances of drawing each action. In this setting one would be hard pressed to hold multiple conjectures. Our theory would then rule out correlated rankings. On the other hand, reasons mentioned in the previous paragraph could lead to correlated rankings.

The control sessions provide a between-subjects comparison of the frequency of correlated rankings in the games that were played by human players and the same games that were played by bingo cages. Furthermore, each control session included predictions for three games from the main experiment that were played by human players. This allows for a within-subject comparison for these three games. Table 4 and the right panel of Figure 3 summarize the results.[21]

The results paint a clear picture. For all 9 games, correlated rankings are at least five times more frequent in the human players data than in the bingo cage data. The p-values shown in Table 4 are (Fisher's) exact tests and differences are all highly significant. No more than 11 percent of the bingo cage control rankings indicate beliefs consistent with correlated outcomes. Furthermore, in 72-85% of instances subjects report rankings that match the belief implied by the induced marginals (see Table 11 in Appendix A.2); and, of course, this belief is consistent with independent actions for all games.

Predictions for the three games in the control sessions that were played by human participants are sharply different. Here we cannot reject the hypothesis that any of the three games is the same as its counterpart in the main experiment in terms of the frequencies of correlated rankings (p-values of 0.73, 0.82 and 0.61 in Table 4). Thus, the same treatment difference is observed for within-subjects variation as with between-subjects variation.

The control sessions provide clear evidence that subjects understand how

[21]Further tables with descriptive statistics can be found in the "Additional tables: control sessions" Section A.2 of the Appendix. No significant differences exist in the ranking distributions between individual sessions or between sessions with shown vs. hidden payoffs (see Table 10). Therefore, we pool the data across all four control sessions.

Table 4: Correlated rankings (in %): main vs. control sessions

	PD			SH		CO		SC	
	S	P	A	S	P	A1	A2	S	D
Corr.: all types									
Main	58	49	45	64	79	66	68	21	75
Control									
B-cage	9	9	7	9	7	11	11	4	9
(p-val.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Human	65				89				70
(p-val.)	(0.59)				(0.26)				(0.56)
Corr.: the main dg.									
Main	50	32	22	58	76	59	50	11	62
Control									
B-cage	4	0	7	2	4	4	4	0	2
Human	52				85				63

Note: All percentages were calculated using the full sample of observations. $n = 101$ for the main sessions and $n = 46$ for the control sessions. The p-values report Fisher's exact tests comparing the main and control sessions.

to rank outcomes consistent with independent occurrence, in the setting devoid of human interaction. Subjects report correlated rankings only when predicting outcomes of games played by human participants. This is consistent with the idea that humans are somehow able to coordinate their actions. However, the outcomes they coordinate on are unknown. In other words, predictors hold multiple conjectures about coordinated outcomes.

4.3 Additional observations

In this section we return to Tables 2 and 7 and make further observations on some patterns of systematic variation in correlated rankings between games. We have already documented that rankings and outcomes do not match up in terms of correlation. But how about other common rankings and the most frequent outcomes? Do they match up? Consider first the PD games. In all three games predictors rank the cooperative outcome (ul) as most likely. Actual players, however, were more likely to choose equilibrium actions. This is not anticipated by many predictors. In the other games (SH, CO and SC), the most likely predicted outcomes are generally in line with the most frequently occurring outcome. Guessing the most frequent outcome correctly is where the highest degree of consistency exists between the outcomes and predictions data, at least for a majority of games.

Another interesting observation concerns the estimated rankings in the strictly competitive games SC-S and SC-D. The absence of correlation and the high frequency of the (indifference) ranking (1,1,1,1) in SC-S is exactly what we would expect in that game (also see Table 7 in the Appendix). However, rankings in SC-D exhibit considerable correlation on the main diagonal. The estimated ranking matches that for the coordination game CO-S. In SC-D the outcomes on the main diagonal are efficient and the deviation gain for the column player is quite weak. It is plausible that some predicting subjects view this game as a coordination game and rank similarly to CO-S. Nevertheless, outcomes on the main diagonal are not very frequent

The overall picture seems to support the idea that monetary and psychological incentives act in tandem and provide subjects with multiple conjectures that correlate beliefs. Assuredly, correlated rankings are most common in games, such as SH-P or CO-S, where the incentives reinforce one another.[22]

Before closing this subsection we make two additional observations. The first concerns the difference between symmetric and asymmetric games. The symmetric games produce smaller variation in the number of unique rankings as well as a higher concentration on a few particular rankings than asymmetric games.[23] In addition, Table 1 shows that for the PD and SH games the correlation on the main diagonal is significantly higher in the symmetric than asymmetric games. This is consistent with an interpretation that in asymmetric games tracing the payoffs and evaluating the incentives requires more attention and cognitive effort.

Our final observation concerns the nature of the "dilemma" between the two outcomes on the main diagonal. In the -S and -P versions of the PD and SH games the two outcomes are attractive for different reasons – Pareto efficiency vs. incentives to deviate in the PD game and Pareto vs. risk dominance in the SH game. Moreover, the -P version reduces the dilemma

[22]In Appendix A.5 we explore further the interaction between the perception of payoffs and the belief correlation. We define several behavioral types and estimate their relative likelihood. The results are consistent with the claims just made.

[23]This is also consistent with a higher measure of noise, μ, in our maximum likelihood estimates in Table 2.

relative to the -S version.[24] This is in contrast with CO-S and SC-D where the two outcomes on the main diagonal are identical and hence equally attractive. Does the nature of the dilemma have any bearing on whether beliefs are consistent with correlated or independent actions? This does not seem to be the case. First, Table 1 shows fewer correlated rankings in PD-P relative to PD-S, but an opposite pattern in the SH games. Second, both CO-S and SC-D record some of the highest proportions of correlated rankings among all games. It seems there is no clear connection between the nature of the dilemma and the degree of correlation in a given game.

5 Summary and Discussion

This experiment provides direct evidence on whether beliefs over strategic behavior of others are consistent with independent or correlated actions. The possibility that beliefs are consistent with correlated actions has recently been suggested by Crawford and Iriberri (2009). To address this question directly we used simple 2×2 games presented in normal form. One group of subjects played the games and another group predicted the outcomes. To obtain a reliable measure of beliefs we elicited likelihood rankings rather than direct probabilities over outcomes. We incentivized truthful reporting of likelihood rankings with a simple payment scheme. In the majority of cases (7 of 11 games) we detect a high frequency (between 50-80%) of diagonally correlated rankings. With the exception of gender effect we do not see a significant impact of any other demographic variables that we collected in our post experimental questionnaire, e.g., type of major, number of semesters completed.[25] Our control sessions rule out possible misconceptions regarding the proper understanding of the outcome ranking procedure, the independence of play or the framing of the games.

The prevailing frequency of correlated rankings is inconsistent with a model of other players' behavior in which a single state generates independent signals (resulting in). It is, however, consistent with a model in which people are believed to be able to coordinate their actions on one of multiple possible outcomes. In games with multiple Nash equilibria, these outcomes seem to be the ones on which players are expected to coordinate. Interestingly, in one

[24] Note that under risk neutrality, in the SH-P game neither equilibrium is risk dominant.

[25] For regression results see Appendix A.4. They show that women are more likely than men to state correlated rankings.

prisoner's dilemma game and one strictly competitive game, ranking reports suggest predictors believe that players' actions are contingent on one state or are independent of any states. This lends some support to the idea that pure strategy Nash equilibria play an important role in correlating beliefs.

We find strong evidence that subjects tend to think of others as behaving in a correlated manner. Interestingly, an emerging literature has considered a seemingly opposite phenomenon called correlation neglect, i.e., Enke and Zimmermann (2019) or Eyster and Weizsacker (2016). These studies present individuals with correlated information and they are made aware of this correlation. Their decisions, however, are consistent with them treating the information as independent. While these correlation neglect studies suggest that information from others tends to be treated as independent, our paper identifies conditions where beliefs about others' behavior tend to be correlated. This contrast suggests an interesting avenue for further research.

Our experiment is only the first step in studying how subjects form beliefs of others' play. We subjected predictions to a stress test that strongly favored predictions consistent with independent play. Our players interacted anonymously and were matched independently. The predicting players, however, reported rankings consistent with correlated actions. We document that such reports occur in large proportions. However, will such reports persist as subjects gain more experience? Is the correlation likely to get stronger in larger or more complex games? Will reports agree with correlated equilibrium probabilities in hawk-dove games where some equilibria require private signals? What are the implications for economically relevant scenarios, such as, entry games, public goods games, or auctions? These are all open questions that we leave for future research.

References

[1] Aumann, R., 1974. Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics*, 67-96.

[2] Aumann, R., 1987. Equilibrium as an expression of Bayesian rationality. *Econometrica*, 1-18.

[3] Aumann R. and Brandenburger A. (1995) Epistemic conditions for Nash equilibrium. *Econometrica* 63, 1161–1180

[4] Brandenburger, A. and Friedenberg, A., 2008. Intrinsic correlation in games. *Journal of Economic Theory*, 28-67.

[5] Cason, T. and Sharma, T., 2007. Recommended play and correlated equilibria: An experimental study. *Economic Theory*, 11-27.

[6] Charness, G. and Rabin, M., 2002. Understanding social preferences with simple tests. *Quarterly Journal of Economics*, 817-869.

[7] Costa-Gomes, M., Crawford, V. P. and Broseta, B., 2001. Cognition and behavior in normal-form games: An experimental study. *Econometrica*, 1193-1235.

[8] Costa-Gomes, M. A., Crawford, V. P. and Iriberri, N., 2009. Comparing models of strategic thinking in Van Huyck, Battalio, and Beil's coordination games. *Journal of the European Economic Association*, 7(2-3), 365-376.

[9] Dal Bó, P. D. and Fréchette, G. R., 2011. The evolution of cooperation in infinitely repeated games: Experimental evidence. *The American Economic Review*, 101(1), 411-429.

[10] Duffy, J. and Feltovich, N., 2010. Correlated equilibria, good and bad: An experimental study. *International Economic Review*, 701-721.

[11] Duffy, J. and Fisher, E. O. N., 2005. Sunspots in the Laboratory. *American Economic Review*, 95(3), 510-529.

[12] Engelmann, D. and Strobel, M., 2000. The false consensus effect disappears if representative information and monetary incentives are given. *Experimental Economics*, 3(3), 241-260.

[13] Enke, B., and Zimmermann, F. 2019. Correlation neglect in belief formation. *Review of Economic Studies*, 86, 313-332.

[14] Ennis, H. M. and Keister, T., 2010. Banking panics and policy responses. *Journal of Monetary Economics*, 404-419.

[15] Epstein, L. G., and Halevy, Y. 2019. Ambiguous Correlation. *Review of Economic Studies*, 86, 668-693.

[16] Eyster, E., and Weizsacker, G. 2016. Correlation Neglect in Portfolio Choice: Lab Evidence. Available at SSRN: https://ssrn.com/abstract=2914526 or http://dx.doi.org/10.2139/ssrn.2914526

[17] Fishburn, P. C., 1970. Utility Theory for Decision Making. Wiley, New York. Krieger edition, 1979.

[18] Foster, D. and R. V. Vohra, 1997 : Calibrated Learning and Correlated Equilibrium, *Games and Economic Behavior*, 21, 40-55.

[19] Guillen, P. and Hing, A., 2014. Lying through their teeth: Third party advice and truth telling in a strategy proof mechanism. *European Economic Review*, 70, 178-185.

[20] Goeree, J. K. and Holt, C. A., 2004: A model of noisy introspection. *Games and Economic Behavior*, 46(2), 365-382.

[21] Greiner, B., 2015. Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1(1), 114-125.

[22] Hart, S. and A. Mas-Colell, 2000: A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68(5), 1127-1150.

[23] Healy, P. J., 2007. Group reputations, stereotypes, and cooperation in a repeated labor market. *American Economic Review*, 97(5), 1751-1773.

[24] Ho, T. H., Camerer, C., and Weigelt, K., 1998. Iterated dominance and iterated best response in experimental "p-beauty contests." *American Economic Review*, 947-969.

[25] Hossain, T. and Okui, R., 2013. The Binarized Scoring Rule. *Review of Economic Studies*, 80, 984-1001.

[26] Iriberri, N. and Rey-Biel, P., 2013. Elicited beliefs and social information in modified dictator games: What do dictators believe other dictators do? *Quantitative Economics*, 4(3), 515-547.

[27] Ivanov, A., Levin, D. and Peck, J., 2009. Hindsight, foresight, and insight: an experimental study of a small-market investment game with

common and private values. *The American Economic Review*, 1484-1507.

[28] Moreno, D. and Wooders, J., 1998. An experimental study of communication and coordination in noncooperative games. *Games and Economic Behavior*, 24(1), 47-76.

[29] Offerman, T., Sonnemans, J. and Schram, A., 1996. Value orientations, expectations and voluntary contributions in public goods. *The Economic Journal*, 817-845.

[30] Palfrey, T. R. and Pogorelskiy, B. K., 2019. Communication Among Voters Benefits the Majority Party. *The Economic Journal*, 129, 961-990.

[31] Rey-Biel, P., 2009. Equilibrium play and best response to (stated) beliefs in normal form games. *Games and Economic Behavior*, 65(2), 572-585.

[32] Rubinstein, A. and Salant, Y., 2014. They do what I do: Positive correlation in ex-post beliefs. Unpublished, Tel Aviv University.

[33] Rubinstein, A. and Salant, Y., 2016. Isn't Everyone Like Me. On the Presence of Self-Similarity in Strategic Interactions. *Judgment and Decision Making*, 11(2), 168-173.

[34] Schlag, K., Tremewan, J. and Van der Weele, J., 2015. A penny for your thoughts: A survey of methods for eliciting beliefs. *Experimental Economics*, 18, 457-490.

[35] Stahl, D. O. and Wilson, P. W., 1995. On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior*, 10(1), 218-254.

[36] Van Huyck, J. B., Battalio, R. C. and Beil, R. O., 1990. Tacit coordination games, strategic uncertainty, and coordination failure. *American Economic Review*, 80(1), 234-248.

[37] Van Huyck, J. B., Battalio, R. C. and Beil, R.O., 1991. Strategic uncertainty, equilibrium selection, and coordination failure in average opinion games. *Quarterly Journal of Economics*, 106(3), 885-910.

[38] Vanberg, C., 2008. A short note on the rationality of the false consensus effect. Unpublished note, University of Heidelberg.

Appendix A

A.1 Game sequencing

Table 5: Game ordering between sessions

	Beh. part	Pred. part			
		OP 1 PP 1&2	OP 2	OP 3	OP 4
1	PD-S	PD-S	SH-P	CO-A2	SC-S
2	SC-S	SC-S	PD-A	PD-P	CO-A1
3	SH-P	SH-P	CO-A2	SC-A	SH-S
4	CO-A2	CO-A2	SH-A	SH-P	PD-S
5	PD-P	PD-P	SC-A	CO-S	SH-A
6	CO-A1	CO-A1	CO-A1	PD-A	CO-A2
7	SH-S	SH-S	PD-S	SC-A	PD-P
8	PD-A	PD-A	SH-S	SH-A	CO-S
9	CO-S	CO-S	SC-S	PD-S	SH-P
10	SH-A	SH-A	PD-P	SH-S	SC-A
11	SC-A	SC-A	CO-S	CO-A1	PD-A

Note: OP and PP refers to Only-preddict and Play-and-predict session respectively.

A.2 Additional tables

Table 6: Correlated rankings (%): Only-predict vs. Play-and-predict sessions

	PD			SH			CO			SC	
	S	P	A	S	P	A	S	A1	A2	S	D
Only-Predict sessions (53 predictors)											
Main Diag.	51	30	21	64	77	47	66	58	57	13	60
Off Diag.	6	13	11	2	2	0	9	6	8	8	8
Circular	4	4	11	8	6	23	2	2	9	4	8
Total	61	47	43	74	85	70	77	66	74	25	76
Play-and-Predict sessions (48 predictors)											
Main Diag.	50	33	23	52	75	25	69	60	44	8	65
Off Diag.	0	15	8	0	0	0	6	6	8	8	8
Circular	6	2	15	2	0	29	0	0	1	4	2
Total	56	50	46	54	75	54	75	66	53	20	75
Fisher's p-value	.552	.692	.692	.038	.225	.101	1	1	.203	.317	1

Table 7: Descriptive statistics for main sessions

Games	Payoffs		Beh. pt.		Pred. pt.							
			no. of pairs		mean ranks		mode		2nd & 3rd most frq.			
PD-S	6,6	2,7	10	14	1.6	2.5	1	3	1	2	3	2
	7,2	4,4	8	4	2.6	2.0	3	2	2	3	2	1
							Cor (31)		Ind (16)		Ind (13)	
PD-P	5,5	1,9	4	12	1.6	2.2	1	2	1	3	2	1
	9,1	2,2	7	13	2.2	2.2	2	3	3	2	1	3
							Ind (27)		Cor (17)		Cor (11)	
PD-A	5,6	2,9	4	7	2.0	2.2	1	2	3	2	1	2
	7,1	3,2	7	18	2.8	2.3	3	4	2	1	3	3
							Ind (18)		Ind (7)		Cir (6)	
SH-S	6,6	2,5	16	9	1.3	2.7	1	3	1	2	3	2
	5,2	4,4	9	2	2.7	2.0	3	2	2	3	2	1
							Cor (40)		Ind (15)		Ind (7)	
SH-P	7,7	0,3	22	4	1.2	2.8	1	3	1	3	1	2
	3,0	4,4	8	2	2.8	2.0	3	2	4	2	2	3
							Cor (51)		Cor (11)		Ind (6)	
SH-A	7,6	1,5	16	9	1.4	3.0	1	4	1	4	1	2
	6,2	3,4	8	3	2.5	2.2	3	2	2	3	2	2
							Cor (19)		Cir (4)		Ind (9)	
CO-S	6,6	1,1	13	13	1.2	1.8	1	2	1	1	2	1
	1,1	6,6	7	3	1.8	1.2	2	1	1	1	1	2
							Cor (61)		Ind (19)		Cor (6)	
CO-A1	7,3	1,1	8	11	1.8	2.2	2	3	1	3	2	1
	1,1	4,4	5	12	2.9	1.6	3	1	3	2	3	2
							Cor (25)		Cor (16)		Ind (11)	
CO-A2	6,4	2,2	15	12	1.7	2.9	1	4	2	4	2	1
	3,3	3,7	6	3	2.6	2.0	3	2	3	1	3	2
							Cor (23)		Cor (15)		Ind (5)	
SC-S	6,2	2,6	6	15	1.3	1.4	1	1	1	1	1	2
	2,6	6,2	3	12	1.6	1.6	1	1	2	2	2	1
							Ind (47)		Ind (13)		Cor (8)	
SC-D	8,4	1,5	12	11	1.2	1.9	1	2	1	1	2	1
	1,5	8,4	11	2	1.9	1.3	2	1	1	1	1	2
							Cor (46)		Ind (18)		Cor (8)	

Note: below each of the three most frequent rankings we report whether the ranking can be classified as correlated (*Cor*), independent (*Ind*), or circular (*Cir*); The numbers of subjects reporting these exact rankings (out of 101) are shown in parentheses.

Table 8: Descr. stat. for control sessions (b-cage play prediction)

Games			Drawn pairs		Pr.: hid. payoffs (26 obs.)				Pr.: sh. payoffs (20 obs.)			
					means		mode		means		mode	
PD-S	7	5										
6	6,6	2,7	2	1	1	2.1	1	2	1.1	2.4	1	2
6	7,2	4,4	7	2	1	2	1	2	1.5	2.2	1	2
							(24)				(12)	
PD-P	5	7										
4	5,5	1,9	1	1	3.7	2.7	4	3	3.4	2.8	4	3
8	9,1	2,2	6	4	2	1.1	2	1	1.9	1.2	2	1
							(19)				(14)	
PD-A	5	7										
3	5,6	2,9	3	1	3.9	3	4	3	3.7	3	4	3
9	7,1	3,2	2	6	2	1	2	1	2.2	1.1	2	1
							(23)				(13)	
SH-S	8	4										
10	6,6	2,5	7	2	1.1	2	1	2	1.4	2.1	1	2
2	5,2	4,4	2	1	2.8	3.8	3	4	3	3.7	3	4
							(21)				(16)	
SH-P	11	1										
9	7,7	0,3	8	0	1.1	2.8	1	3	1	3.1	1	3
3	3,0	4,4	4	0	2.1	3.8	2	4	2.1	3.6	2	4
							(20)				(14)	
CO-A1	3	9										
4	7,3	1,1	0	5	3.8	2	4	2	3.7	2.1	4	2
8	1,1	4,4	1	6	2.8	1.1	3	1	2.9	1	3	1
							(19)				(16)	
CO-A2	6	6										
11	6,4	2,2	4	6	1	1	1	1	1.2	1.3	1	1
1	3,3	3,7	1	1	2	2.1	2	2	2.3	2.2	2	2
							(24)				(14)	
SC-S	2	10										
5	6,2	2,6	0	1	3.9	2	4	2	3.7	2	4	2
7	2,6	6,2	2	9	2.9	1.1	3	1	3	1	3	1
							(22)				(17)	
SC-D	6	6										
9	8,4	1,5	5	4	1	1	1	1	1.1	1.3	1	1
3	1,5	8,4	2	1	2.1	2.1	2	2	2.1	2.3	2	2
							(24)				(22)	

Table 9: Descr. stat. for control sessions (human play prediction)

Games			Pred. pt.: payoffs hid. (obs.: 26)				Pred. pt.: payoffs sh. (obs.: 20)			
			means		mode		means		mode	
PD-S										
	6,6	2,7	1.7	2.3	1	3	1.7	2.4	1	3
	7,2	4,4	2.4	2	3	2	2.5	1.8	3	2
					(9)				(7)	
SH-P										
	7,7	0,3	1.2	2.8	1	3	1.4	2.8	1	3
	3,0	4,4	2.8	1.8	3	2	1.9	1.8	3	2
					(18)				(14)	
SC-D										
	8,4	1,5	1.2	1.7	1	2	1.4	2	1	2
	1,5	8,4	1.8	1.2	2	1	1.9	1.3	2	1
					(15)				(10)	

Table 10: Freq. of correlated rankings (control session)

	B-cage prediction									Human pr.		
	PD			SH		CO		SC		PD	SH	SC
	S	P	A	S	P	A1	A2	S	D	S	P	D
Payoffs hidden ($n = 26$)												
ul-dg.	1	0	0	1	1	0	0	0	1	13	24	16
ur-dg.	1	0	0	0	0	0	0	0	0	4	1	0
Other	0	0	0	1	0	1	1	0	1	1	0	0
Payoffs shown ($n = 20$)												
ul-dg.	1	0	3	0	1	2	2	0	0	11	15	13
ur-dg.	1	1	0	1	1	0	0	1	2	1	0	2
Other	0	3	0	1	0	2	2	1	0	0	1	1

Table 11: Perc. of reported rankings matching the implied rankings

Gm. & impl. belief				Impl. rnk.		Hid. payoffs (obs.: 26)	Sh. payoffs (obs.: 20)	Comb.
PD-S		.58	.42					
	.5	.29	.21	1	2	92	60	78
	.5	.29	.21	1	2			
PD-P		.42	.58					
	.33	.14	.19	4	3	73	70	72
	.67	.28	.39	2	1			
PD-A		.58	.42					
	.25	.15	.10	3	4	88	65	78
	.75	.44	.31	1	2			
SH-S		.67	.33					
	.83	.56	.28	1	2	81	80	81
	.17	.11	.06	3	4			
SH-P		.92	.08					
	.75	.69	.06	1	3	77	70	74
	.25	.23	.02	2	4			
CO-A1		.25	.75					
	.33	.08	.25	4	2	73	80	76
	.67	.17	.50	3	1			
CO-A2		.5	.5					
	.92	.04	.04	2	2	92	70	82
	.08	.46	.46	1	1			
SC-S		.17	.83					
	.42	.07	.35	4	2	85	85	85
	.58	.10	.49	3	1			
SC-D		.5	.5					
	.75	.38	.38	1	1	92	75	85
	.25	.13	.13	2	2			

A.3 Most likely ranked outcomes

We ask whether monetary payoffs played a significant role in ranking outcomes. One way to approach this question is to check whether various types of incentives, such as the incentive to deviate or the outcome's equitable and efficiency properties, had any impact on which outcome was ranked as the most likely. Table 12 presents a Probit model in which the dependent variable takes a value of 1 if the outcome was ranked as the most likely (i.e., received the rank of 1) and 0 otherwise. This is regressed on several dummy variables: "No profitable deviation" takes a value of 1 if, in a given game, the outcome was a pure strategy Nash equilibrium; "Most equitable payoffs" takes a value of 1 if the game's outcome minimized the payoff difference between the two players; and "Efficient payoffs" takes a value of 1 if the outcome maximized the sum of players' payoffs.[26]

Table 12: Probit regression of the outcome ranked most likely on incentives

	Coeff	Std. err.	p-value
Constant	−1.723	0.148	0.000
No profitable deviation	0.556	0.070	0.000
Most equitable payoffs	0.939	0.117	0.000
Efficient payoffs	1.129	0.122	0.000
No. obs.	4443		
Log likelihood	−2171.6		

Note: Game fixed effects were included; errors were clustered by subject.

Results reveal that all three types of incentives had a significant impact on the outcome being ranked as the most likely with the psychological incentives playing a more important role than the monetary incentives. This suggests that concerns for equity and efficiency play a role in determining how subjects rank outcomes.

A.4 Demographics

In this section we examine whether there is a relationship between the propensity to report a correlated ranking and some of the demographic variables collected in the post-experiment questionnaire. We run a logistic regression in which the dependent variable on the left hand side is coded 1 if the ranking was correlated and 0 otherwise. Among the regressors are a gender dummy

[26]The correlations between regressors are: ρ(No dev., Equit.) = 0.337; ρ(No dev., Effic.) = 0.114; ρ(Equit., Effic.) = −0.126.

(female = 1, 42.8% of our sample), number of college semesters completed, the engineering dummy indicates whether the subject had an engineering major[27] (25.1% of our sample), and a non-US dummy takes on value of 1 if subject's country was outside of the US (30.6% of our sample). We have included game fixed effects and clustered the errors by subject.

Table 13: Probit regression of the type of ranking on demographics

Dep. variable = 1 if reported ranking is correlated	(1)	(2)	(3)
Constant	0.259 (0.205)	0.004 (0.199)	−0.121 (0.206)
Gender (F)	0.340** (0.138)	0.254* (0.136)	0.312** (0.142)
Semester	−0.013 (0.03)	−0.006 (0.029)	0.009 (0.028)
Engineering major	−0.132 (0.164)	−0.05 (0.17)	0.079 (0.169)
Non-US	−0.183 (0.157)	−0.081 (0.159)	−0.205 (0.161)
No. obs.	1078	1078	722
Log likelihood	−651.01	−656.92	−465.8

Notes: Standard errors clustered on subjects included in parentheses. * and ** denote significantly different from 0 at the five- and one-percent levels.

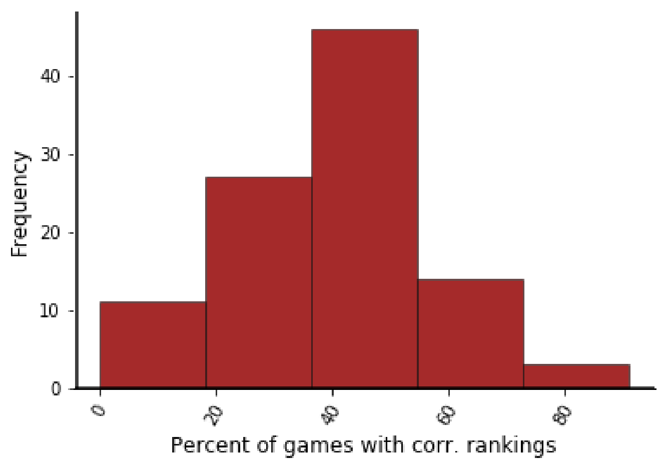
Table 13 reports the results. The first regression considers all types of correlated rankings. In column (2) the dependent variable took on value of 1 only if the rankings was correlated on the main diagonal. Finally, regression (3) was identical to (2) except that it excluded all rankings corresponding to extreme beliefs (with guesses of 12 or smaller than 4 on the number of pairs at the most likely outcome). The results reveals a significant gender effect, with women more likely to report correlated rankings. No other regressors had a significant impact on the ranking behavior.

[27]Engineering is highly selective at Purdue University and thus serves as a proxy for cognitive ability.

A.5 Behavioral models

Figure 4 shows the frequency of correlated rankings across all 101 subjects in the main human players prediction sessions. Most subjects report diagonally correlated rankings in about 20-60% of games. In this subsection we define four behavioral types and estimate their relative likelihood in the data.

Figure 4: Frequency of correlation within subjects



The predictor may perceive the stage one payoffs in terms of the monetary payments or may incorporate concerns for efficiency and equity in the payoffs (e.g., in the spirit of Charness and Rabin, 2002). Incorporating sufficiently strong efficiency and equity considerations into the payoffs, all of the games (except SC-S) could be considered coordination games with two pure strategy NE on the main diagonal. For brevity, we refer to such social preferences as psychological payoffs. Apart from this, we would like to distinguish between individuals whose beliefs are based on a single conjecture versus multiple conjectures. The former group believe in independent play. Our formulation, thus, gives us four types. Type t_{MI} considers only the monetary payments as payoffs and predicts according to the pure strategy Nash equilibrium induced by these payoffs. Type t_{MC} takes into account only monetary payoffs and predicts according to some diagonally correlated equilibrium. Similarly, for those who take psychological payoffs into consideration we have the types

t_{PI} and t_{PC}, respectively. Let $T = \{t_{MI}, t_{MC}, t_{PI}, t_{PC}\}$ and let t denote an element of T. Γ denotes the set of our eleven games and g denotes an element of Γ. Let N be the set of predictors.

Our objective is to estimate the proportion of the four types in our data. Let q_t be the proportion of type t and $q = (q_{MI}, q_{MC}, q_{PI}, q_{PC})$. For any game g in Γ, and type t in T, let P_{tg} be the set of consistent rankings, i.e., a ranking that t may provide.

Table 14 summarizes the relevant parts of elements in our predicted sets P_{tg}, for all four types.

Table 14: Predicted highest ranked outcome for behavioral types

	Monetary Correlated	Monetary Independent	Psychological Correlated	Psychological Independent
PD	(d, r)	(d, r)	(u, l) *and/or* (d, r)	(u, l) *or* (d, r)
SH & CO	(u, l) *and/or* (d, r)	(u, l) *or* (d, r)	(u, l) *and/or* (d, r)	(u, l) *or* (d, r)
SC-S	-	-	-	-
SC-D	-	-	(u, l) *and/or* (d, r)	(u, l) *or* (d, r)

Let $\Pr(k \mid \rho, \mu)$ be as defined earlier. Fix μ. For a given predictor $i \in N$, in game $g \in \Gamma$, let $\rho(i, t, g, \mu)$ in P_{tg} be the ranking she wishes to report if she were to be of type t. Given i's stated ranking k_{ig} in game g, let:

$$\rho(i, t, g, \mu) \in \max_{\rho \in P_{t,g}} \Pr(k_{ig} \mid \rho, \mu).$$

Now define the likelihood function:

$$L(\rho, q, \mu) = \prod_{i \in N} \prod_{t \in T} q_t \prod_{g \in \Gamma} \Pr(k_{i,g} \mid \rho(i, t, g, \mu), \mu).$$

We then estimate the proportions q_t's as a solution to the following optimization program:

$$\max_{\rho,q,\mu} L(\rho, q, \mu) \text{ s.t. } \sum_{t\in T} q_t = 1.$$

Table 15 summarizes the results. The SC-S game is excluded as the predictions for the four types coincide for this game. SC-D provides specific prediction only for P types but not for M types. Because of this, we exclude this game from estimates shown in columns (1) and (2).

Table 15: ML estimation of behavioral types

	(1)	(2)	(3)
q_{MC}	0.051 (0.000)	0.849 (0.000)	-
q_{MI}	0.000 (0.000)	0.151 (0.000)	-
q_{PC}	0.759 (0.006)	-	0.806 (0.006)
q_{PI}	0.190 (0.006)	-	0.194 (0.006)
μ	0.484 (0.000)	0.626 (0.000)	0.490 (0.001)
$-\log L$	2051.7	2600.6	2069.9

Note: (1) and (2) exclude data from the SC games (both are uninformative); (3) excludes data from game SC-S. Rankings with extreme frequency guesses (<4 or >11) are excluded. Number of observations used: 799. Standard errors are in the parenthesis (all coefficients are highly significant at 1% level). Corresponding table with the full sample estimates can be found in Appendix A.6.

The first column of Table 15 shows that the PC type whose payoffs incorporate a psychological transformations of monetary payoffs and who forms conjectures around PNE's that leads to correlated beliefs commands strong support in the data. A large majority, about 76% of the population, is estimated to be consistent with this behavioral type. The type (PI) who does not think in terms of such conjectures does receive a non negligible weight of 19%. Restricted estimates in columns (2) and (3) perform two robustness exercises where we only consider types with the same payoffs: monetary in (2) and psychological in (3). The results are qualitatively similar and if anything put even more weight on the type C that lead to correlated beliefs.

A.6 Full sample estimates

Table 16: ML estimation and LR tests - full sample

	PD games			SH games		
	S	P	A	S	P	A
Rnk	$\begin{array}{\|c\|c\|}\hline 1 & 3 \\ \hline 3 & 2 \\ \hline\end{array}$	$\begin{array}{\|c\|c\|}\hline 1 & 2 \\ \hline 2 & 3 \\ \hline\end{array}$	$\begin{array}{\|c\|c\|}\hline 1 & 2 \\ \hline 4 & 3 \\ \hline\end{array}$	$\begin{array}{\|c\|c\|}\hline 1 & 3 \\ \hline 3 & 2 \\ \hline\end{array}$	$\begin{array}{\|c\|c\|}\hline 1 & 3 \\ \hline 3 & 2 \\ \hline\end{array}$	$\begin{array}{\|c\|c\|}\hline 1 & 4 \\ \hline 3 & 2 \\ \hline\end{array}$
μ	0.745	0.916	1.825	0.572	0.472	0.884
Obs.	101	101	101	101	101	101
$-\log L$	264.9	398.7	421.9	311.8	256.2	362
$-\log \tilde{L}$	411.8	398.7	422.5	385.4	373.3	384.1
p_{LR}	0.000	1	0.273	0.000	0.000	0.000
	Cor	*Ind*	*Cir*	*Cor*	*Cor*	*Cor*

	CO games			SC games	
	S	A1	A2	S	D
Rnk	$\begin{array}{\|c\|c\|}\hline 1 & 2 \\ \hline 2 & 1 \\ \hline\end{array}$	$\begin{array}{\|c\|c\|}\hline 1 & 3 \\ \hline 3 & 2 \\ \hline\end{array}$	$\begin{array}{\|c\|c\|}\hline 1 & 4 \\ \hline 3 & 2 \\ \hline\end{array}$	$\begin{array}{\|c\|c\|}\hline 1 & 1 \\ \hline 1 & 1 \\ \hline\end{array}$	$\begin{array}{\|c\|c\|}\hline 1 & 2 \\ \hline 2 & 1 \\ \hline\end{array}$
μ	0.409	0.774	1.017	0.648	0.507
Obs.	101	101	101	101	101
$-\log L$	243.8	370.6	382.3	320.3	306.5
$-\log \tilde{L}$	339	397.5	419.2	320.3	367.7
p_{LR}	0.000	0.000	0.000	1	0.000
	Cor	*Cor*	*Cor*	*Ind*	*Cor*

Note: $\tilde{L}$ indicates the log-likelihood the restricted estimation; p_{LR} refers to the likelihood ratio test. *Cor/Cir/Ind* indicates whether the estimated ranking is diagonally correlated, circularly correlated or independent.

Table 17: ML estimation of behavioral types - full sample

	(1)	(2)	(3)
q_{MC}	0.065 (0.001)	0.852 (0.001)	-
q_{MI}	0.000 (0.000)	0.148 (0.001)	-
q_{PC}	0.780 (0.006)	-	0.826 (0.004)
q_{PI}	0.155 (0.007)	-	0.174 (0.004)
μ	0.470 (0.001)	0.597 (0.000)	0.473 (0.001)
$-\log L$	2509	3178.7	2527.5

Note: (1) and (2) exclude data from the SC games (both are uninformative); (3) excludes data from game SC-S. Number of observations used: 1010. Standard errors are in the parenthesis (all coefficients are highly significant at 1% level).