ISSN (Print) 0473-453X
Discussion Paper No. 1117
ISSN (Online) 2435-0982

STRATEGIC UNCERTAINTY AND PROBABILISTIC SOPHISTICATION

Masaki Aoyagi
Takehito Masuda
Naoko Nishimura

February 2021

The Institute of Social and Economic Research
Osaka University
6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan

Strategic Uncertainty and Probabilistic Sophistication*

Masaki Aoyagi† Takehito Masuda‡ Naoko Nishimura§

March 22, 2021

Abstract

This paper uses laboratory experiments to study subjects' assessment of uncertainty resulting from strategic and non-strategic decisions of other players. Non-strategic events are defined by the colors of balls drawn from urns, whereas strategic events are defined by the action choice in Stag Hunt (SH) and Prisoners' Dilemma (PD) games. We elicit subjects' *matching probabilities* and examine if they satisfy the law of probability including monotonicity and additivity. Violations from the law are observed for both uncertainty sources, but are more substantial for strategic uncertainty. In particular, we observe a *coordination fallacy,* a violation of monotonicity whereby the probability weight placed on a symmetric coordination profile of the games exceeds that placed on the corresponding action choice. The violation is found to be severer for an efficient coordination profile.

Key words: matching probability, ambiguity, uncertainty, coordination, conjunction
JEL codes: C91, D01, D81, D91

*We are particularly grateful to Aurélien Baillon, Chen Li, Peter Wakker, and Songfa Zhong for providing very detailed comments. We also thank Yoichiro Fujii, Nobuyuki Hanaki, Takashi Hayashi, PJ Healy, Keigo Inukai, Asen Kochov, Yukio Koriyama, Hiroaki Masuhara, Yusuke Osaki, Yoshi Saijo, Tridib Sharma, Eli Spiegelman, and the participants of SAET 2018 Taiwan and 2019 North American and European ESA meetings as well as seminars at Osaka, Dijon, Erasmus, Waseda, Kyoto, Kochi Tech, Shinshu, and Hitotsubashi. Shuya Abe, Yuki Hamada and Yoshitaka Okano provided excellent research assistance. Financial support from the JSPS (grant numbers: 22330061, 23530216, 24330064, 24653048, 15K13006, 15H03328, 16H03597, 17K13701, 15H05728, 20H05631) and the Joint Usage/Research Center at ISER, Osaka University, and the International Joint Research Promotion Program of Osaka University is gratefully acknowledged.

†ISER, Osaka University, 6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan. E-mail: aoyagi@iser.osaka-u.ac.jp.

‡Faculty of Economics and Law, Shinshu University, 3-1-1 Asahi, Matsumoto, Nagano 390-8621, Japan. E-mail: tmasuda@shinshu-u.ac.jp.

§Department of Gastronomy Management, Ritsumeikan University, 1-1-1 Noji-higashi, Kusatsu, Shiga 525-8577, Japan. E-mail: nnaoko@fc.ritsumei.ac.jp.

1 Introduction

While there is by now an extensive literature on the theory of ambiguity preferences, the experimental work on the subject has focused almost exclusively on artificial and non-strategic uncertainty based on the colors of balls drawn from urns.[1] In contrast, uncertainty in the real world is often associated with the strategic decision making of human beings. For example, uncertainty in the exchange rate results from governments' strategic trade policies or central banks' market intervention. Uncertainty associated with climate change is not purely in the domain of natural sciences, but a large part of it is associated with how we respond to the change, and the responses are based on strategic interactions among different parties involved.

A natural question then concerns how humans perceive uncertainty when it results from strategic interactions of others. Would individuals facing strategic uncertainty behave in much the same way as they would in the presence of artificial uncertainty such as the colors of balls drawn from urns?

This paper uses a laboratory experiment to investigate if and how individuals' perception of uncertainty changes when uncertainty is of strategic nature compared with when it is not. Specifically, we have a group of subjects create strategic and non-strategic uncertainty in the first set of experiments, and then have a new group of subjects predict the outcome of the first experiments to measure their perception of uncertainty. Subjects in the first set of the experiments are called *players,* and they fill balls into urns and then play one-shot games against each other. Non-strategic events are defined by the (combinations of) colors of balls drawn from urns filled by these subjects, whereas strategic events are defined by individual action choices and joint action profiles in the games played by them. In the second set of the experiments, a new group of subjects called *observers* face *uncertain bets* whose outcome is determined by the events from the first set of experiments. Specifically, for each event E, the observers make a choice between an uncertain bet which yields a positive reward if and only if the event E occurs, and another *risky bet* which yields the same positive reward with an objectively known probability. For each observer i and for each event E, we elicit their *matching probability* $m_i(E)$ of E which equals the probability of the risky bet that they state is indifferent to the uncertain bet based on E.

For the observers' choices to be consistent with the theory of subjective expected utility, the matching probabilities thus elicited must satisfy the law of probability. Specifically, they must satisfy monotonicity: $E \subset F \Rightarrow m_i(E) \leq m_i(F)$, and additivity: $E \cap F = \emptyset \Rightarrow m_i(E) + m_i(F) = m_i(E \cup F)$. We follow the convention in the literature and call observer i *probabilistically sophisticated* (PS) if his matching probability m_i satisfies the law of probability including these conditions. Our empirical strategy is to

[1] See Trautmann & van de Kuilen (2015*a*) for a comprehensive survey of the literature.

examine if our observer subjects are PS by constructing various sophistication indices and computing the values of those indices from the stated matching probabilities.

The player subjects in our experiments play Stag-Hunt (SH) and Prisoners' Dilemma (PD) games. SH represents a coordination game with two pure Nash equilibria (NE). One of the NE yields each player a strictly higher payoff than the other NE, and hence is *payoff dominant.* On the other hand, which of the two NE is *risk dominant* depends on the payoff parameter.[2] We have the observer subjects predict the outcome of two SH games. The payoff dominant NE and the risk dominant NE coincide in one SH game, and they correspond to different action profiles in the other SH game. The interaction between the two notions of dominance in these games creates non-trivial uncertainty for the observers. In PD, on the other hand, there exists a unique NE that corresponds to the strictly dominant action of each player. Despite this, however, the choice of the dominated action is commonly observed in the laboratory, and is known to be more frequent when the temptation payoff is smaller. We have the observers predict the outcome of two PD games with different temptation parameters. The expected difference in the deviation from the theoretical prediction in these PD games also creates substantial uncertainty for the observers. Importantly, the player subjects in this first set of experiments play these games in perfect stranger format with no interim feedback so that they have no chance to coordinate their action choices.

The observers' assessment of uncertainty consistently violates the law of probability for both strategic and non-strategic events, but the patterns of violation are substantially different between the two uncertainty sources. First, regardless of the uncertainty sources, the *binary complementarity* (BC) index, which equals one minus the sum of the matching probabilities of complementary events, is negative on average. This implies that the observers are on average ambiguity seeking in line with some recent findings in the literature. However, this is almost the only similarity between the two sources, and substantial difference exists in the distributions of matching probabilities themselves: While a majority of observers place almost the same probability weights on two ball colors for the urns, they have little agreement on the likelihood of each action choice for the games. Interestingly, we confirm a version of the familiarity hypothesis whereby the observers who state in the post-experimental questionnaire that the games are easier to predict than the urns tend to be more ambiguity seeking in their prediction of the game events.

The most striking difference between the uncertainty sources arises in the matching probabilities of *conjunctional* events, which correspond to the color combination of balls from two urns, or the action profile of two players in games. Specifically, for the strate-

[2] A NE of a symmetric 2×2 coordination game is risk dominant when the action corresponding to the equilibrium yields a higher expected payoff when the other player chooses each action with probability one-half.

gic uncertainty, the matching probabilities of the coordination profiles of the games are extremely large, and in many cases violate *monotonicity* (MN) in the sense that they exceed the matching probabilities of the corresponding individual action choices. We call this phenomenon a *coordination fallacy.* Overall, nearly half of the observers display the coordination fallacy for at least one of the two symmetric profiles, and more frequently for the Pareto efficient coordination profile in both SH and PD. Such violation of monotonicity is not observed for the conjunctional urn events.[3]

To further evaluate the probability weight placed on the symmetric coordination profiles, we introduce *quasi-complementarity* (QC) and *coordination premium* (CP) indeices. QC equals one minus the sum of the matching probabilities of two conjunctional events, and is non-negative under PS. We however find that the index takes negative values for half of observations for the games. CP equals the matching probability of the coordination profile minus the product of the matching probabilities of the corresponding action choices of the two players, and equals zero if the observers believe that the two players make their action choices independently as in our experiments. We observe that CP is indistinguishable from zero for many of the conjunctional urn events, but is significantly positive for all conjunctional game events. The positive CP implies that the observers believe in the presence of some mechanism that helps the players coordinate their actions.

In order to account for the possibility that the observed violation of PS is caused by trembling hand rather than some fundamental deviation from probabilistic thinking, we examine k-PS whereby the values of the indices are allowed to deviate by a small amount $k > 0$ from what is implied by PS. Through this exercise, we again find substantial difference between the urns and games. In the case of the urns, BC is "binding" in the sense that observer i is k-PS whenever he satisfies BC by the margin k (i.e., $|BC_i| \leq k$), whereas in the case of the games, the proportion of k-PS observers is substantially lower for each value of k, and there is no single index that is binding in such a sense.

To the best of our knowledge, our experiments are the first to confirm the violation of monotonicity for strategic uncertainty using an incentivized elicitation of matching probabilities. While the literature usually takes monotonicity for granted in economics, the violation of monotonicity in various forms is discussed in the psychological literature.[4] The best known among them is the *conjunction fallacy,* which states that an individual, when informed of the occurrence of event X, often reports a higher value for

[3]Relatedly, Baillon et al. (2018) find violation of monotonicity to be infrequent even when the subjects had to make predictions on stock market indices under time pressure.

[4]Monotonicity is the bottom line for models of decision under uncertainty (Offerman et al. 2009). For example, the capacity function in Choquet expected utility is assumed monotone (Schmeidler 1989). Monotonicity is also assumed for the theories of maxmin preferences and their generalizations.

the probability $\Pr(A \cap B \mid X)$ than for the probability $\Pr(A \mid X)$.[5] The coordination fallacy identified in this paper is different from the conjunction fallacy in that it involves no conditioning event X.

This paper is organized as follows. We discuss the related literature in Section 2, and describe the experimental design in Section 3. Section 4 presents the sophistication indices used in our analysis, and Section 5 describes results on the violation of PS well as the difference between strategic and non-strategic uncertainty. Approximate PS is examined in Section 6. We offer one possible model which accommodates the violation of monotonicity in Section 7, and then conclude in Section 8.

2 Related Literature

We contribute to the growing experimental literature on descriptive and empirically relevant models of attitudes toward uncertainty. The experimental literature, especially since Halevy (2007), recognizes that decision making under uncertainty has much more diverse patterns than universal aversion typically assumed in theory: Attitudes toward uncertainty are found to depend on such factors as event likelihood, the sign of the payoff associated with the event, and familiarity with uncertainty sources (Trautmann & van de Kuilen 2015*a*, Kocher et al. 2018, Abdellaoui et al. 2011, Chew et al. 2012, Baillon & Bleichrodt 2015, Baillon & Emirmahmutoglu 2018).[6]

Uncertainty attitudes are studied most commonly in the context of ambiguity preferences. As mentioned in the Introduction, much of the experimental literature on ambiguity preferences (Halevy 2007, Chew et al. 2017, Kocher et al. 2018) follows the original implementation by Ellsberg and formulates uncertainty in terms of artificial and non-strategic events defined by the colors of balls drawn from urns.[7] Attempts to mea-

[5]The conjunction fallacy is usually explained by representativeness: The description of the conditioning event X is chosen to be representative of the properties of B, but not of A so that when evaluating $\Pr(A \cap B \mid X)$, attention is focused on B, but when evaluating $\Pr(A \mid X)$, A is discounted. In the famous Linda example in Tversky & Kahneman (1982), X is the statement "Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations," A is the statement "Linda is a bank teller," and B is the statement "Linda is active in a feminist movement." See also Gigerenzer (1996). Some authors use incentivized experiments to examine monotonicity in the context of the conjunction fallacy (*e.g.,* Zizzo et al. 2000, Charness et al. 2010), but do not quantify the degree of violation unlike in the present experiments.

[6]The proportion of subjects whose behavior conforms to a major decision model varies across studies. For example, the proportion of subjects whose behavior conforms to subjective expected utility is 15-20% in Halevy (2007) and about 60% in Ahn et al. (2014). Chew et al. (2017) and Yang & Yao (2017) report that a substantial proportion of subjects violate subjective expected utility and other decision models including maxmin and smooth ambiguity models.

[7]See Trautmann & van de Kuilen (2015*a*) for a comprehensive survey of the extensive literature.

sure ambiguity preferences when uncertainty comes from natural sources include Baillon & Bleichrodt (2015), and Baillon et al. (2018), who use uncertainty over stock market indices and observe that familiarity with the source of uncertainty and/or time pressure influence its perception, and Li et al. (2017), who use uncertainty over how other subjects rank three alternatives.[8]

On the perception of uncertainty involved in strategic decision making, Camerer & Karjalainen (1994) observe that subjects prefer objectively risky bets to the play of games against another subject, and conclude that strategic uncertainty entails ambiguity. Eichberger et al. (2008) change the characteristics of players that subjects play games against, and observe through their action choices that the opponents' characteristics influence their perception of strategic uncertainty. Heinemann et al. (2009) elicit from subjects playing threshold public good games the certainty equivalents of strategic uncertainty that the threshold is achieved.

Our experiments are characterized by the following design choices. First, we use multiple price lists and solicit matching probabilities as in Baillon & Bleichrodt (2015) and Dimmock et al. (2016).[9] This requires no assumption on risk or uncertainty attitudes and allows for a direct and quantifiable measurement of the perception of uncertainty. Second, the non-strategic uncertainty is created by subjects who fill balls into urns and receive a flat payment for the task. The process hence is in line with the creation of strategic uncertainty, which results from the same subjects' action choices in games. This also addresses the concern that the observers may think that an experimenter attempts to control the payment when they fill urns themselves.[10] Third, in contrast to the majority of belief elicitation experiments which elicit beliefs from the subjects who subsequently make decisions, observers in our experiments do not make decisions themselves. This not only removes restrictions on the type of events whose matching probabilities can be elicited, but also eliminates concerns over the possible effect of belief

[8]Empirical evidence of source preferences includes a home bias in stock holdings (French & Poterba 1991), and a trading volume bias based on the proximity of countries (Epstein & Schneider 2010). Ritov & Baron (1990) find public aversion to receiving vaccination which involves uncertainty over its intended as well as side effects.

[9]The concept of matching probability dates back to Raiffa (1968). See also Wakker (2010), Baillon et al. (2018) and Li et al. (2019) for the measurement of uncertainty attitudes using matching probabilities.

[10]See for example Al-Najjar & Weinstein (2009) and Dominiak & Duersch (2019).

elicitation on the decisions.[11,12]

3 Experimental Design

3.1 Outline

The experiments were conducted at Osaka University with subjects recruited from the student body through ORSEE (Greiner 2015) and the program coded using z-Tree (Fischbacher 2007).[13] The experiments consist of two types of sessions as follows: The first type is *player sessions* where subjects fill urns with colored balls and then play 2×2 games against each other. The second type is *observer sessions* where a new group of subjects make "predictions" about the outcome of the player sessions. There are two player sessions and sixteen observer sessions, and no subjects participated in more than one session. In both types of sessions, no communication is allowed among the subjects. The following is the description of the player and observer sessions in more detail.

3.2 Player Sessions

The player sessions took place in March 2018.[14] Each player session had 22 subjects and consisted of two parts: In Part 1, a subject was shown two urns on their screen, and asked to put one-hundred colored balls in each urn. Specifically, they can put red (R) and green (G) balls in Urn 1, and yellow (Y) and blue (B) balls in Urn 2. Any color composition was allowed in each urn as long as the total number equaled one hundred. A subject received a fixed payment of 500 JPY, or approximately 5USD, for this task. Non-strategic uncertainty is hence created by the subjects themselves just as strategic uncertainty is created by their action choices.

In Part 2, subjects formed pairs and played a sequence of 21 different 2×2 games. These are seven parameterizations each of Prisoner's Dilemma games (PD), Stag-Hunt games (SH), and Battle of the Sexes games (BoS). All payoffs are denominated in JPY. The 21 games appeared in random order, but the order was the same for all pairs in

[11]See Schlag et al. (2015) for a survey on the subject. On the effect of elicitation on the subsequent action choice, belief elicitation is found to increase cooperative behavior in public good games (Gächter & Renner 2010), and reduce the choice of dominated actions (Hoffmann 2014). Hedging behavior is discussed in Blanco et al. (2010).

[12]Among those experiments that have players and observers as two separate groups of subjects are Hyndman et al. (2009), Palfrey & Wang (2009), and Cason et al. (2020). See Section 5.5.

[13]The subject pool contains more than 3,000 students of various majors. Recruitment advertisement states that there is an opportunity to earn money in a research experiment.

[14]Upon arrival in the lab, the subjects received an instruction sheet and a record sheet which they can use to record their response. The instructions were read aloud by the same experimenter.

each session. After each round of play, random re-matching took place by the perfect stranger format. There was no feedback until after the session was over. At the end of each player session, one game was randomly chosen for payment, and the subjects were informed of their earnings and the choice of their matched partner in the chosen game. Payment was made in cash at the conclusion of each session, which lasted for about 90 minutes. Average earnings were 2617.4 JPY. We also elicited demographic information of the participants of these sessions. The instructions for the player sessions stated that the outcome of these experiments would be used for the decision making of other subjects in future experiments, but did not explicitly mention the nature of the observer sessions so that their choices would not be affected by the existence of observers.

	SHx					PDx			
P1 \ P2	A_2		B_2		P1 \ P2	A_2		B_2	
A_1	x	x	x	0	A_1	1	1	$-x$	$1+x$
B_1	0	x	1	1	B_1	$1+x$	$-x$	0	0

Table 1: Standardized SH and PD Games

Table 1 shows the SH and PD games with standardized payoffs $(g_1(\cdot), g_2(\cdot))$. SHx and PDx refer to the SH and PD games, respectively, with parameter x. For SH, x is the payoff associated with the safe action A_i, and the larger is its value, the lower is the relative attractiveness of the payoff dominant NE (B_1, B_2).[15] In fact, (B_1, B_2) is the risk dominant equilibrium if $x < \frac{1}{2}$, and (A_1, A_2) is the risk dominant equilibrium if $x > \frac{1}{2}$. For PD, x is the temptation parameter, and we expect the lower cooperation rate when x is larger as shown by Dal Bó & Fréchette (2018). The actual payoffs in the experiments are the affine transformations of these standardized payoffs: They are given by $1600g_i(\cdot) + 100$ for the SH games, and by $700g_i(\cdot) + 1000$ for the PD games. The player sessions had seven SH game with $x = 0.2, \ldots, 0.8$ and seven PD games with $x = 0.1, \ldots, 0.7$. Tables 2 and 3 depict four games SH0.2, SH0.8, PD0.1, and PD0.5 which were used in the observer sessions.

	SH0.2					SH0.8			
P1 \ P2	A_2		B_2		P1 \ P2	A_2		B_2	
A_1	420	420	420	100	A_1	1380	1380	1380	100
B_1	100	420	1700	1700	B_1	100	1380	1700	1700

Table 2: SH Games in Observer Sessions

[15] Our choice of SH games is based on Rankin et al. (2000).

P1 \ P2	PD0.1				P1 \ P2	PD0.5			
	A_2		B_2			A_2		B_2	
A_1	1700	1700	930	1770	A_1	1700	1700	650	2050
B_1	1770	930	1000	1000	B_1	2050	650	1000	1000

Table 3: PD Games in Observer Sessions

3.3 Observer Sessions

In the observer sessions which took place in April and November 2019, we elicit from a new group of subjects their predictions about the outcome of the player sessions. A participant in these sessions is referred to as an *observer* in what follows.

The observers' predictions are formally elicited in the form of *matching probabilities* (Baillon & Bleichrodt 2015, Baillon & Emirmahmutoglu 2018) defined as follows. Let y be a real number and E be an event corresponding to the set of outcomes in the player sessions. For any monetary reward y, denote by y_m0 the *risky bet* that pays out y with probability m and 0 with probability $1-m$, and by y_E0 the *uncertain bet* that pays out y when event E takes place and 0 otherwise. The *matching probability* $m = m_i(E)$ of event E for observer i is a probability that makes him indifferent between y_m0 and y_E0: $m = m_i(E) \Leftrightarrow y_m0 \sim y_E0$. We say that observer i is *probabilistically sophisticated* (PS) if his matching probability $m_i(E)$ satisfies the law of probability. Let Ω be the universal event so that $m_i(\emptyset) = 0$, and $m_i(\Omega) = 1$. For any events E, $F \subset \Omega$,

$$E \subset F \Rightarrow m_i(E) \leq m_i(F), \text{ and } E \cap F = \emptyset \Rightarrow m_i(E \cup F) = m_i(E) + m_i(F).$$

With some abuse of notation, we denote by R the event in which a Red ball is drawn from the first urn, and by G the event in which a Green ball is drawn from it. Events Y (Yellow) and B (Blue) for the second urn are similarly defined. Likewise, we denote by A_i and B_i the events in which player i chooses actions A_i and B_i, respectively. To simplify notation, we also use EE' to denote the intersection $E \cap E'$ of two events E and E' so that $RG = R \cap G$, $A_1A_2 = A_1 \cap A_2$, *etc.*

Each observer session consists of three parts. Questions in Parts 1 and 2 of the observer sessions elicit matching probabilities of events from the player sessions using multiple price lists as shown in Figure 1, and described in more detail below. Part 1 has eight questions on urn events $E \subset \{R, G\} \times \{Y, B\}$, and Part 2 has eight questions on game events $E \subset \{A_1, B_1\} \times \{A_2, B_2\}$.[16] Table 4 shows the correspondence between the events and the questions in Parts 1 and 2. Events are classified into three different types as follows. Events are *simple* if they correspond to the color of a ball drawn from

[16] The two urns used for Part 1 (i.e., one with R and G and the other with Y and B) are filled by two different player subjects, and this is stated in the instructions of the observer sessions.

Event Type	Part 1		Part 2	
	Question #	Event	Question #	Event
Simple	1	R	1	A_1
	2	G	2	B_1
	3	Y	3	A_2
	4	B	4	B_2
Conjunctional	5	RY	5	A_1A_2
	6	GB	6	B_1B_2
Diagonal	7	$RY \cup GB$	7	$A_1A_2 \cup B_1B_2$
	8	$RB \cup GY$	8	$A_1B_2 \cup B_1A_2$

Table 4: Correspondence between questions and events in observer sessions

Treatment	#Sessions (#Observers)	Action choices by players ($\%A, \%B$)	Average total earnings of observers in JPY (USD)
SH0.2	4 (24, 23, 23, 22)	(13.6, 86.4)	2096.3 (19.23)
SH0.8	4 (23, 24, 23, 20)	(56.8, 43.2)	1689.6 (15.50)
PD0.1	4 (21, 24, 22, 24)	(45.4, 54.6)	1641.2 (15.06)
PD0.5	4 (23, 23, 22, 24)	(20.4, 79.6)	1977.2 (18.14)

Notes. a) The male-to-female ratio among subjects (2.50) is close to that (2.58) of the undergraduate student body of Osaka University as of 1 May 2019. b) Conversion rate: 1 USD=109 JPY.

Table 5: Treatment summary for observer sessions.

a single urn or the action choice of a single player, *conjunctional* if they correspond to the combination of colors of balls drawn from the two urns, or the action profile of the two players, and *diagonal* if they correspond to the union of two (disjoint) conjunctional events.

For Parts 1 and 2, the reward y equals 1000 JPY.[17] Section 3 has five questions that elicit certainty equivalents of urn and game events and other measures of uncertainty attitudes.[18]

There are four sessions each for the four treatments that correspond to the four games SH0.2, SH0.8, PD0.1, and PD0.5 with a total of 365 observers participating. Table 5 presents the summary statistics of the observer sessions.

At the beginning of each observer session, an instruction sheet for each part was

[17] $y = 9.17$ USD according to the average exchange rate 1 USD=109 JPY in 2019.

[18] The questions in Part 3 are as follows. Q1 elicits the certainty equivalent (CE) of a bet $1000_{50}0$; Q2 elicits $n \in \{-350, -300, ..., +150\}$ which makes $1000_{50}0$ equivalent to a bet $(25, 1500; 25, 500 + n)$; Q3 elicits the CE of a bet $1000_G 0$; Q4 elicits the CE of a bet $1000_{B_1} 0$; Q5 elicits the CE of a bet $1000_{B_1 B_2} 0$.

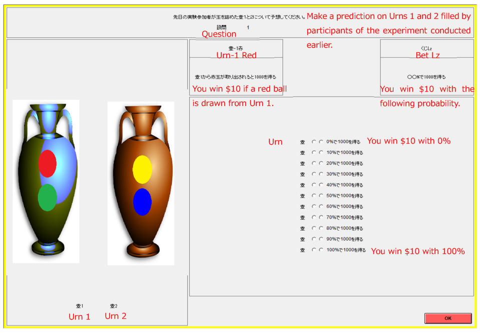


Figure 1: Urn part decision screen for observer sessions.

distributed along with a record sheet that the subjects used to write down their responses. These instructions included the reproduction of the instructions for the player sessions so that there would be clear understanding among our observer subjects on how the uncertainties had been generated. All instructions were read aloud by a text-to-speech software so as to control the time and tone of the articulation of the instructions across sessions. In addition, a snapshot of one of the player sessions was shown to inform that the questions asked would be about the behavior of other students from the same university. Payments ranged from 500 JPY to 4000 JPY, including the show-up fee of 500 JPY. A typical observer session lasted for approximately 100 minutes.

Every question of the observer session has a similar structure. For illustration, Figure 1 shows Question 1 in Part 1 of the observer sessions. It elicits a matching probability of event R by presenting a choice between "Bet R", which pays out 1000 JPY if R takes place and zero otherwise, and "Bet z", which pays out 1000 JPY with probability $z/100$, where $z = 0, 1, \ldots, 100$. The first screen shows a list of z's with 10% increments, and the second screen presents a list of z's with 1% increments based on the choice in the first screen. For example, if a switch from Bet R to Bet z takes place at $z = 40$ in the first screen, then the second screen zooms in to the list of z's with $z = 30, 31, \ldots, 39, 40$.[19] We call this switching value of z an observer's *choice* and identify it with his matching probability of the corresponding event.[20] After the choice is made, a confirmation screen

[19] In this case, the choices for $z = 30$ and $z = 40$ in the second screen are determined by the choices in the first screen. If the choice is either $z = 0$ or $z = 100$ in the first screen, then the second screen is skipped and the confirmation screen appears.

[20] More specifically, we follow Baillon & Emirmahmutoglu (2018) and convert the switching point to the matching probability as follows. For any uncertain event E, if observer i switches from $y_E 0$ to $y_z 0$

appears and prompts the subjects to either confirm their choice or return to the first screen to reenter their choice for the question. The experimenter makes sure that no one proceeds to the next question before everyone confirms his choice.

4 Sophistication Indices

We begin by the description of probabilistic sophistication (PS) measures used in our analysis. Application of the indices below to our data will be based on the consideration that there is in principle no reason for the observers to distinguish between the two urns or the two players. For this reason, we take the average of the matching probabilities of simple urn events R and Y which are the first color of each urn, and the average of the matching probabilities of G and B which are the second color of each urn. Likewise, we use the average of the matching probabilities of simple game events A_1 and A_2, and the average of the matching probabilities of B_1 and B_2.[21]

$$\begin{aligned} m_i^{R+Y} &= \tfrac{1}{2}\left(m_i(R) + m_i(Y)\right), & m_i^{G+B} &= \tfrac{1}{2}\left(m_i(G) + m_i(B)\right), \\ m_i^{A_1+A_2} &= \tfrac{1}{2}\left(m_i(A_1) + m_i(A_2)\right), & m_i^{B_1+B_2} &= \tfrac{1}{2}\left(m_i(B_1) + m_i(B_2)\right). \end{aligned} \tag{1}$$

4.1 Monotonicity

Monotonicity (MN) holds if the matching probability of an event is larger than that of its sub-event. Specifically, for any events E and E' such that $E' \subset E$, define

$$MN_i(E, E') = m_i(E) - m_i(E'). \tag{2}$$

Clearly, PS implies $MN_i(E, E') \geq 0$. Our special focus is on the case when E is simple and E' is conjunctional. Based on the combined matching probabilities introduced in (1), we define the indices MNU_i^{RY} and MNU_i^{GB} for the urns and MNG_i^{A} and MNG_i^{B} as in Table 6. We also define MNU_i and MNG_i to be the combined measures as in Table 6.

at $z = 1, \ldots, 100$, then his matching probability is set equal to the mid-point between $z-1$ and z. For example, when i switches at $z = 68$, then we set $m_i(E) = 0.675$. On the other hand, if subject i prefers $y_0 0$ to $y_E 0$ for $z = 0$, or prefers $y_E 0$ to $y_{100} 0$ for $z = 100$, we censor it from data as an inconsistent choice. Out of 5,840 (=365 subjects × 16 questions) elicited switching points, 5,397 responses (=92.4% of the initial) survived this process. The proportion of censored data (7.6%) is comparable to that in Baillon & Emirmahmutoglu (2018) (6.2%) who also use the matching probability elicitation method. Our analysis in what follows computes each index for a subsample of observations in which the matching probabilities required for the computation of the index are available. See Table 15 in Appendix A for the size of the subsample for each index.

[21] See Section 5.1 for the examination of this point.

	Index	Definition	PS implies
MN	MNU_i^{RY}	$m_i^{R+Y} - m_i(RY)$	≥ 0
	MNU_i^{GB}	$m_i^{G+B} - m_i(GB)$	
	MNG_i^{A}	$m_i^{A_1+A_2} - m_i(A_1A_2)$	
	MNG_i^{B}	$m_i^{B_1+B_2} - m_i(B_1B_2)$	
	MNU_i	$\min\{MNU_i^{RY}, MNU_i^{GB}\}$	
	MNG_i	$\min\{MNG_i^{A}, MNG_i^{B}\}$	
BC	BCU_i	$1 - m_i^{R+Y} - m_i^{G+B}$	$= 0$
	BCG_i	$1 - m_i^{A_1+A_2} - m_i^{B_1+B_2}$	
	BCU_i^d	$1 - m_i(RY \cup GB) - m_i(RB \cup GY)$	
	BCG_i^d	$1 - m_i(A_1A_2 \cup B_1B_2) - m_i(A_1B_2 \cup B_1A_2)$	
QC	QCU_i	$1 - m_i(RY) - m_i(GB)$	≥ 0
	QCG_i	$1 - m_i(A_1A_2) - m_i(B_1B_2)$	
CP	CPU_i^{RY}	$m_i(RY) - m_i(R)m_i(Y)$	$= 0$
	CPU_i^{GB}	$m_i(GB) - m_i(G)m_i(B)$	
	CPG_i^{A}	$m_i(A_1A_2) - m_i(A_1)m_i(A_2)$	
	CPG_i^{B}	$m_i(B_1B_2) - m_i(B_1)m_i(B_2)$	

Notes. m_i^{R+Y}, m_i^{G+B}, $m_i^{A_1+A_2}$, and $m_i^{B_1+B_2}$ are defined in (1).

Table 6: Definitions of indices

4.2 Binary Complementarity

Binary complementarity (BC) as proposed by Baillon & Bleichrodt (2015) measures the deviation from unity of the sum of the matching probabilities of an event and its complement: Let $E^c = \Omega \setminus E$ denote the complement of event E, and define the BC of event E by

$$BC_i(E) = 1 - m_i(E) - m_i(E^c). \tag{3}$$

Clearly, PS implies $BC_i(E) = 0$ for any event E. On the other hand, $BC_i(E) > 0$ corresponds to ambiguity aversion, and $BC_i(E) < 0$ corresponds to ambiguity seeking.[22] Define BCU_i and BCG_i for the combined simple events and BCU_i^d and BCG_i^d for the diagonal events as in Table 6.

4.3 Quasi-Complementarity

Quasi-complementarity (QC) is a weaker version of BC in the sense that it measures the difference between unity and the sum of matching probabilities of two disjoint events (whose union may not be universal):

$$QC_i(E, E') = 1 - m_i(E) - m_i(E') \quad \text{for } E,\ E' \text{ with } E \cap E' = \emptyset.$$

[22]See for example Baillon & Bleichrodt (2015).

Clearly, PS implies $QC_i(E, E') \geq 0$. We will use QC to evaluate the weights placed on the two conjunctional events and define QCU_i and QCG_i as in Table 6.

4.4 Coordination Premium

In the player sessions, the paired subjects make their choices simultaneously with no communication between them. As such, their choices are independent.[23] We test if the matching probabilities observe the independence of choices. Specifically, for non-empty subsets E and E', we consider the *coordination premium* (CP) index defined by

$$CP_i(E, E') = m_i(EE') - m_i(E)m_i(E'). \tag{4}$$

$CP_i = 0$ when E and E' are independent according to m_i ($m_i(EE') = m_i(E)m_i(E')$), and is positive when E and E' are positively correlated according to m_i. We will test this when E and E' are both simple and EE' is conjunctional, and define CPU_i^{RY} and CPU_i^{GB} for the urns and CPG_i^A and CPG_i^B for the games as in Table 6.

For each index introduced above, our aggregate analysis is based on the sample average expressed without the superscript i. That is, when we denote by N the number of observers, we let $MNU^{RY} = \frac{1}{N}\sum_i MNU_i^{RY}$, $BCG = \frac{1}{N}\sum_i BCG_i$ and so on.

5 Results

5.1 Distribution of Matching Probabilities

Table 7 shows the frequency of each action profile realized in the player sessions for each treatment along with a few statistics. As seen, the action choice is concentrated on (B_1, B_2) in SH0.2 and PD0.5, whereas it is evenly spread over four profiles in SH0.8 and PD0.1.

Table 8 lists the mean matching probabilities of each event along with the standard deviations. Table 16 in Appendix A shows that no statistical difference is found except in a few cases between the matching probabilities of two events which differ only in the identities of the urns or the players.[24]

[23] However, even when the choices in each pair are independent, the choices may appear correlated when they are aggregated over all pairs. This point is discussed in the next section.

[24] Based on the two-sided t-test, the difference is significant in one case at $p < 0.05$, and in two cases at $p < 0.10$.

Treatment	SH0.2		SH0.8		PD0.1		PD0.5	
Action	A_2	B_2	A_2	B_2	A_2	B_2	A_2	B_2
A_1	1	1	8	6	4	8	1	5
B_1	3	17	3	5	4	6	2	14
Coord. rate								
$p(A_1A_2)$	0.045		0.364		0.182		0.045	
$p(B_1B_2)$	0.773		0.227		0.273		0.636	
$p(A_1A_2)+p(B_1B_2)$	0.818		0.591		0.455		0.682	
Entropy								
H	1.085		1.917		1.936		1.418	
Correlation								
ρ	0.261		0.189		-0.069		0.054	

Notes. For each action profile x, $p(x) = \frac{\#\text{obs. of } x}{\#\text{player pairs}}$. Entropy $H = -\sum_x p(x)\log_2 p(x)$. A higher entropy implies more dispersion. Correlation $\rho = \frac{p(A_1A_2)-p(A_1)p(A_2)}{\sqrt{\{p(A_1)-p(A_1)^2\}\{p(A_2)-p(A_2)^2\}}}$.

Table 7: Realized action profile distributions by treatment

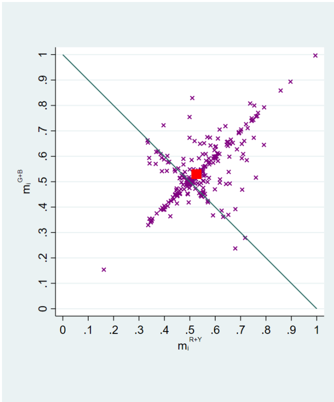
Treatment	Urn events							
# Observers	$m(R)$	$m(G)$	$m(Y)$	$m(B)$	$m(RY)$	$m(GB)$	$m(RY \cup GB)$	$m(RB \cup GY)$
SH0.2	0.54	0.53	0.52	0.55	0.31	0.31	0.5	0.5
92	(0.12)	(0.12)	(0.12)	(0.11)	(0.13)	(0.13)	(0.14)	(0.14)
SH0.8	0.54	0.51	0.52	0.53	0.3	0.32	0.5	0.5
90	(0.1)	(0.09)	(0.11)	(0.12)	(0.12)	(0.12)	(0.13)	(0.14)
PD0.1	0.53	0.53	0.52	0.52	0.3	0.31	0.47	0.47
91	(0.1)	(0.1)	(0.11)	(0.11)	(0.11)	(0.12)	(0.14)	(0.14)
PD0.5	0.53	0.52	0.51	0.53	0.32	0.32	0.48	0.48
92	(0.13)	(0.14)	(0.14)	(0.14)	(0.13)	(0.13)	(0.13)	(0.13)
Treatment	Game events							
# Observers	$m(A_1)$	$m(B_1)$	$m(A_2)$	$m(B_2)$	$m(A_1A_2)$	$m(B_1B_2)$	$m(A_1A_2 \cup B_1B_2)$	$m(A_1B_2 \cup B_1A_2)$
SH0.2	0.27	0.75	0.33	0.73	0.27	0.71	0.78	0.32
92	(0.26)	(0.24)	(0.3)	(0.25)	(0.26)	(0.27)	(0.21)	(0.24)
SH0.8	0.5	0.47	0.5	0.54	0.45	0.49	0.79	0.31
90	(0.36)	(0.33)	(0.34)	(0.32)	(0.32)	(0.33)	(0.17)	(0.21)
PD0.1	0.45	0.55	0.46	0.53	0.45	0.44	0.7	0.38
91	(0.33)	(0.31)	(0.32)	(0.31)	(0.31)	(0.29)	(0.22)	(0.19)
PD0.5	0.43	0.66	0.41	0.65	0.38	0.59	0.69	0.4
92	(0.29)	(0.25)	(0.26)	(0.24)	(0.28)	(0.27)	(0.2)	(0.22)

Notes. Standard deviation in parentheses.

Table 8: Average matching probabilities by treatment and by event

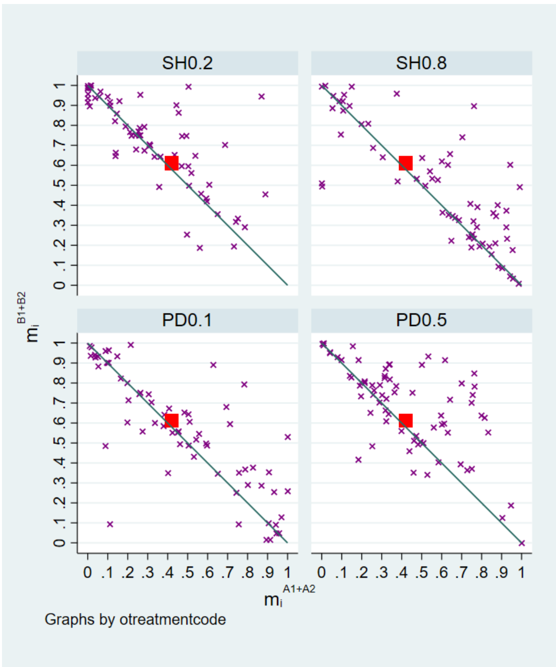
Figures 2, 3, and 4, and 5 display the distributions of matching probabilities of simple and conjunctional events.[25] As immediately seen, there is a clear difference between the distributions for the urns and those for the games. In the case of the urns, the distributions have spread along the 45-degree line, suggesting that the observers agree that each color or color combination is equally likely. In the case of the games, on the other hand, the distributions have large two-dimensional spread in each treatment.

[25]Figures 12 and 13 in Appendix A show the distributions of $(m_i(RY \cup GB), m_i(RB \cup GY))$ and $(m_i(A_1A_2 \cup B_1B_2), m_i(A_1B_2 \cup B_1A_2))$, respectively.



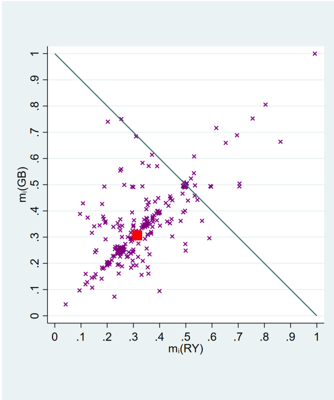
Notes. The red square represents the average point.

Figure 2: (m_i^{R+Y}, m_i^{G+B}): all treatments combined



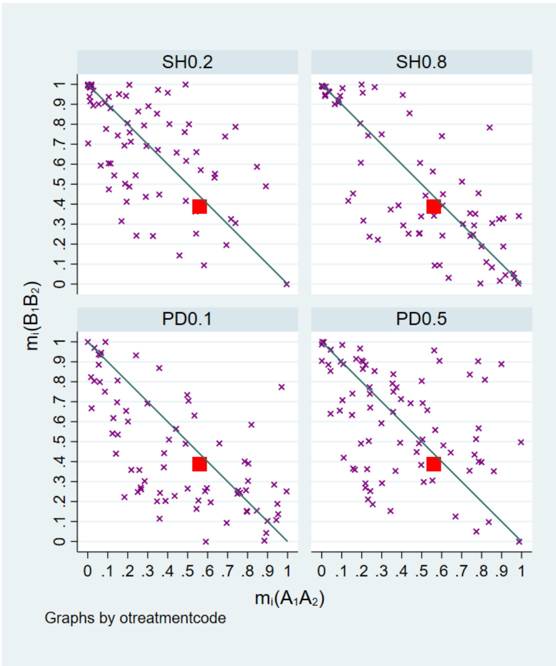
Notes. The red square in each panel represents the average point.

Figure 3: $(m_i^{A_1+A_2}, m_i^{B_1+B_2})$ by treatment



The red square represents the average point.

Figure 4: $(m_i(RY), m_i(GB))$: all treatments combined



The red square in each panel represents the average point.

Figure 5: $(m_i(A_1A_2), m_i(B_1B_2))$ by treatment

5.2 Binary Complementarity

Observations on BC are summarized as follows.

Result 1 *i) (Ambiguity seeking attitudes)* $BCU < 0$ *in every treatment, whereas*

$BCG < 0$ *in three out of four treatments.*

ii) (Familiarity effect) BCG_i *tends to be negative when observer* i *perceives the game to be easier to predict than the urn.*

Ambiguity seeking attitudes The negative 45-degree line in Figures 2 and 3 corresponds to PS $BCU_i = 0$ and $BCG_i = 0$, respectively. As seen, PS is violated for the majority of the observations. Furthermore, the patterns of violation are surprisingly different between the two uncertainty sources: While strong positive correlation exists between the two matching probabilities for the urns, negative correlation exists between them for the games.[26] More specifically, the spread along the 45-degree line for the urns suggests that the observers consider each color equally likely but do not necessarily think that their probability equals 0.5. In contrast, the large spread orthogonal to the 45-degree line for the games suggests that the observers differ substantially in their assessment of the likelihood of the two actions. On the other hand, Table 17 in Appendix A checks whether or not the observers who are nearly PS with $|BCU_i| \leq 0.1$ for the urns are more likely to have $|BCG_i| \leq 0.1$ for the games, and find correlation for SH0.8 and PD0.5 ($p < 0.05$, χ^2 test). On average, the observers are *ambiguity seeking* for the both uncertainty sources with $BCU < 0$ in every treatment and $BCG < 0$ except in PD0.1 ($p < 0.05$) as seen in Table 9. This observation corroborates findings in the recent experimental studies using matching probability elicitation (e.g., Baillon & Bleichrodt 2015, Baillon et al. 2018, Trautmann & van de Kuilen 2015*b*, Li et al. 2019).[27]

Kocher et al. (2018) survey the literature to point out that ambiguity seeking is particularly prevalent for low likelihood events. The diverse matching probabilities of the simple game events reported by our observers allow us to test this hypothesis for strategic uncertainty. Specifically, we take a small $\delta > 0$ and identify observer i with $\min\{m_i^{A_1+A_2}, m_i^{B_1+B_2}\} < \delta$ as having low likelihood assessment of either simple game event. We then test if their average BCG is different from that of the observers with $\min\{m_i^{A_1+A_2}, m_i^{B_1+B_2}\} \geq \delta$. The result is at odds with Kocher et al. (2018), and indicates that the observers with $\min\{m_i^{A_1+A_2}, m_i^{B_1+B_2}\} < \delta$ are less ambiguity seeking than the observers with $\min\{m_i^{A_1+A_2}, m_i^{B_1+B_2}\} \geq \delta$ for different values of $\delta \leq 0.25$.

Familiarity effect Heath & Tversky (1991), Fox & Tversky (1995) and subsequent studies show that decision makers are relatively more ambiguity-seeking when they face

[26] The Pearson correlation equals 0.640 for the urns and −0.784 for the games.

[27] Trautmann & van de Kuilen (2015*b*) use various elicitation methods to conclude that the sum of elicited probabilities of complementary events exceeds 1 on average. Table 1 of Oechssler & Roomets (2015) displays heterogeneity in ambiguity attitudes reported in the literature.

	Treatment				ANOVA for difference in
	SH0.2	SH0.8	PD0.1	PD0.5	BC among treatments
BCU	-0.080*** (0.188)	-0.044*** (0.136)	-0.053*** (0.173)	-0.047** (0.206)	$p = 0.508$
BCG	-0.044** (0.167)	-0.068*** (0.190)	-0.038 (0.194)	-0.104*** (0.198)	$p = 0.156$
$BCU - BCG$	-0.054*	0.026	-0.021	0.061**	
#Obs.	65	62	64	70	

Notes. a) * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$. b) $BC = 0$ is tested using two-sided t-test. c) Standard deviation in parentheses. d) The difference is computed for the subsample in which both terms are available, is tested using two-sided t-test.

Table 9: BC by treatment and by uncertainty source

uncertainty from sources they feel more familiar or competent with.[28] In order to determine which uncertainty source each observer finds more familiar, we use their response to the post-experimental questionnaire and suppose that observer i is more familiar with the games if he states that the games are easier to predict, and vice versa. Our hypothesis is that $BCG_i < BCU_i$ if observer i finds the game more familiar, and $BCG_i > BCU_i$ otherwise.[29] Formally, let $F_i \in \{u, g\}$ denote observer i's response to the post-experimental questionnaire: $F_i = u$ if observer i states that the urn was easier to predict, and $F_i = g$ otherwise. Let BCU^u and BCU^g be the mean BCU's for the urns of the observers with responses $F_i = u$ and g, respectively, in the questionnaire. Define BCG^u and BCG^g for the game similarly. Our hypotheses are

$$H_0^{BCU} : BCU^u = BCU^g \quad \text{and} \quad H_1^{BCU} : BCU^u < BCU^g$$

for the urns and

$$H_0^{BCG} : BCG^g = BCG^u \quad \text{and} \quad H_1^{BCG} : BCG^g < BCG^u$$

for the games.[30]

[28]The literature uses words familiar, knowledgeable, and competent interchangeably. A classic field example is the so-called home bias in stock markets (French & Poterba 1991): Investors' portfolio has too much weight on a domestic market even when a higher return is expected in overseas markets.

[29]The specification of familiarity resembles that by Heath & Tversky (1991), who classify subjects based on their stated familiarity with football and politics. Subsequent literature uses exogenous specification of familiarity according to the subjects' background. For example, Chew et al. (2012) and Baillon & Bleichrodt (2015) use geographical proximity as a proxy of familiarity. Fox & Weber (2002) show that when a pair of uncertain bets are presented simultaneously, subjects' attitudes toward the same bet change depending on whether or not they are more familiar with the uncertainty source of the other bet.

[30]The numbers of available observations are as follows: 106 (BCU^u), 235 (BCU^g), 92 (BCG^u), and 170 (BCG^g).

The results corroborate the finding in the literature. More precisely, H_0^{BCG} is rejected (by two-sided t-test) when data from the four treatments are combined ($t = -2.751$, $p = 0.006$), and also for SH0.2 ($t = -1.695$, $p = 0.0949$) and SH0.8 ($t = -3.3438$, $p = 0.001$) when data for each treatment are tested separately. In contrast, H_0^{BCU} is not rejected ($t = -1.028$, $p = 0.305$) when all urn data are combined.[31] In Appendix B, we present further examination on the effect of familiarity in conjunction with the entropy of play.

5.3 Monotonicity

Our primary findings on MN can be summarized as follows:

Result 2 *i) (Coordination fallacy)* MNU^{RY}, $MNU^{GB} > 0$ *in every treatment, while* MNG^A *and* MNG^B *are often indistinguishable from zero, and nearly one-half of the observers have* $MNG_i = \min\{MNG_i^A, MNG_i^B\} < 0$.

ii) (Association between urns and games) There is no *correlation between* MNU_i *and* MNG_i *in every treatment.*

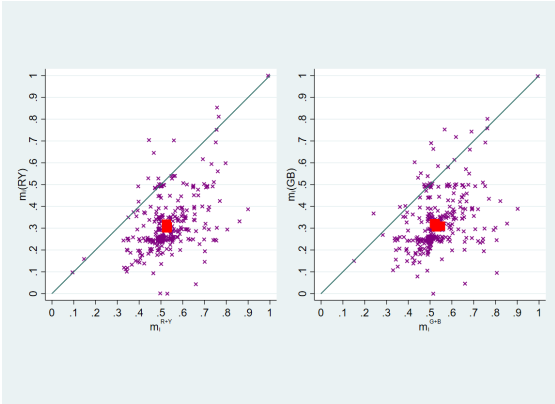
iii) (Efficient versus inefficient profile) MNG *is indistinguishable from zero more often for the Pareto efficient profile than for the inefficient profile.*

Coordination fallacy Figures 6 and 7 present scatter plots of matching probabilities of simple and conjunctional events for the urns and games, respectively. The region below the 45-degree line corresponds to PS: $MN_i \geq 0$. The difference between the urns and games is clear: While a predominant majority of observers lie below the 45-degree line for the urns, a substantial proportion lies above the line for the games. In fact, $MNG_i < 0$ in 48.5% (= 126/260) of the cases in the subsample in which both MNG_i^A and MNG_i^B can be computed.[32] (Figure 7). Table 10 shows that the average MNG for the games is indistinguishable from zero in many cases.[33]

[31] On the other hand, perceived familiarity creates no difference between MNU and MNG, or CPU and CPG.

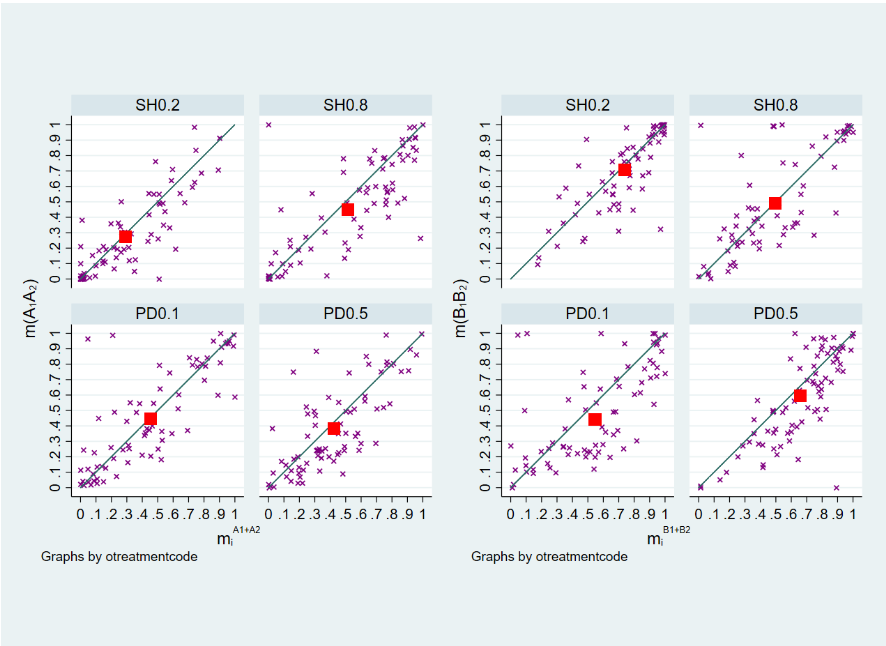
[32] See Footnote 20.

[33] Appendix C uses data from Part 3 to confirm that the violation of monotonicity is also substantial in terms of certainty equivalents for some pair of game events: For many observers, the certainty equivalent of the bet $1000_{B_1B_2}0$ is larger than that of the bet $1000_{B_1}0$.



Notes. The red square in each panel represents the average point.

Figure 6: $(m_i^{R+Y}, m_i(RY))$ (left) and $(m_i^{G+B}, m_i(GB))$ (right): all treatments combined



Notes. The red square in each panel represents the average point.

Figure 7: $(m_i^{A_1+A_2}, m_i(A_1A_2))$ (left) and $(m_i^{B_1+B_2}, m_i(B_1B_2))$ (right) by treatment

Association between MNU_i and MNG_i We may suspect that the violation of monotonicity is due to some hidden characteristics of individual observers, and that observers with those characteristics violate monotonicity whether the event is strategic or not. An observer's cognitive ability is one possible candidate for such characteristics.

	Treatment			
	SH0.2	SH0.8	PD0.1	PD0.5
MNU^{RY}	0.223***	0.222***	0.229***	0.206***
	(0.116)	(0.108)	(0.103)	(0.123)
MNU^{GB}	0.235***	0.198***	0.220***	0.205***
	(0.114)	(0.127)	(0.110)	(0.137)
$MNU^{RY} - MNU^{GB}$	-0.010	0.021**	0.012	0.008
#Obs.	90	87	85	88
MNG^{A}	0.026	0.068**	**0.005**	**0.038***
	(0.140)	(0.173)	(0.185)	(0.187)
MNG^{B}	**0.037****	**0.009**	0.102***	0.074***
	(0.151)	(0.189)	(0.182)	(0.145)
$MNG^{A} - MNG^{B}$	-0.013	0.078**	-0.101***	-0.034
#Obs.	66	62	61	71

Notes. a) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. b) $MN = 0$ is tested using two-sided t-test. c) Standard deviation in parentheses. d) The difference is computed for the subsample in which both terms are available (Footnote 20, Table 15), and is tested using two-sided t-test. d) Bold indicates the Pareto action profile (B_1B_2 in SH and A_1A_2 in PD).

Table 10: MN by treatment and by uncertainty source

For example, Chew et al. (2018) observe that subjects with better math skills are more ambiguity averse, suggesting correlation between the subjects' cognitive abilities and their BC values. Likewise, Oechssler et al. (2009) find that subjects with high CRT scores ($\geq$ 2) are less likely to exhibit the conjunction fallacy. In a similar vein, we may conjecture that an observer with a low cognitive ability may violate monotonicity both for the urns and games.[34] Table 11 tests this conjecture by classifying observers based on the signs of their MNU_i and MNG_i as defined in Table 6. While $MNU_i \geq 0$ for 93.4% (= 327/350) of observers, $MNG_i \geq 0$ for only 51.5% (= 134/260) of them. Furthermore, approximately one half of those observers with $MNU_i \geq 0$ have $MNG_i < 0$ in every treatment. As shown in the bottom line of Table 11, we cannot reject the null hypothesis that the likelihood of violation for one uncertainty source is independent of that of the other source (Fisher's exact test). In other words, this shows that the violation of monotonicity is strongly specific to strategic uncertainty and not specific to the observers.[35]

[34]The 742 subjects who were recruited from the same subject pool at Osaka University for an (unrelated) online experiment conducted in August 2020 scored on average 2.51 points for the three-item CRT. We should note that this score is higher than the average (2.05 points) of the high-score group of Oechssler et al. (2009).

[35]We find no clear evidence on the relationship between BCG_i and MNG_i either. For example, the average BCG value of the subjects with $MNG_i \geq 0$ is not statistically lower than the corresponding

		SH0.2		SH0.8		PD0.1		PD0.5	
		MNG_i							
		+	−	+	−	+	−	+	−
MNU_i	+	33	29	27	29	31	28	37	27
	−	1	2	1	3	0	1	1	4
#Obs.		65		60		60		69	
p		0.602		0.616		0.483		0.166	

Notes. a) Two-sided Fisher's exact test for independence of observations in a 2×2 contingent table.[36] b) MNU_i and MNG_i are computed for the subsamples in which every matching probability involved is available (Footnote 20, Table 15).

Table 11: Source dependence of monotonicity

Efficient versus inefficient profiles A closer look at Table 10 reveals that MNG for the Pareto efficient action profile (indicated by bold) is indistinguishable from zero more often: $MNG^A = 0$ cannot be rejected for PD0.1, and $MNG^B = 0$ cannot be rejected for SH0.8. To examine this point more closely, we classify observers i based on the signs of MNG_i^A and MNG_i^B in Table 12. In line with the above observation, it shows that the violation of monotonicity is more likely when it involves the Pareto action profile in both SH and PD. Formally, let ν_A be the number of subjects who violates monotonicity for A_1A_2 only, and ν_B be the number of subjects who violate monotonicity for B_1B_2 only:

$$\nu_A = \#\{i : MNG_i^A < 0 \text{ and } MNG_i^B \geq 0\},$$
$$\nu_B = \#\{i : MNG_i^A \geq 0 \text{ and } MNG_i^B < 0\}.$$

The relative degree of violation of monotonicity between the two profiles is captured by

$$\eta = \frac{\nu_B}{\nu_A}.$$

We see from Table 12 that $\eta < 1$ in both PD games whereas $\eta > 1$ in both SH games. In fact, the hypothesis $\nu_A = \nu_B$ is rejected for SH0.8 and PD0.5 ($p = 0.065$ for SH0.8 and $p = 0.089$ for PD0.5, one-sided McNemar test).[37]

value of the subjects with $MNG_i < 0$.

[36] Chi-square test is inappropriate given that the number of observations is less than five in some cell.

[37] The alternative hypothesis is $\nu_B > \nu_A$ for SH and $\nu_A > \nu_B$ for PD. The McNemar test is a version of chi-square test for binary matched pairs. See Moffatt (2015).

		SH0.2		SH0.8		PD0.1		PD0.5	
		MNG^B							
		+	−	+	−	+	−	+	−
MNG^A	+	34	12	29	18	32	10	40	10
	−	11	9	10	5	16	3	17	4
#Obs		66		62		61		71	
χ^2		0.04		2.29		1.38		1.81	
p		0.417		0.065		0.120		0.089	

Notes. p values are based on the one-sided McNemar test.

Table 12: MN values for symmetric profiles

5.4 Quasi-Complementarity

Result 3 $QCG_i < 0$ *for about a half of the observations for the games, whereas* $QCU_i < 0$ *is rare for the urns.*

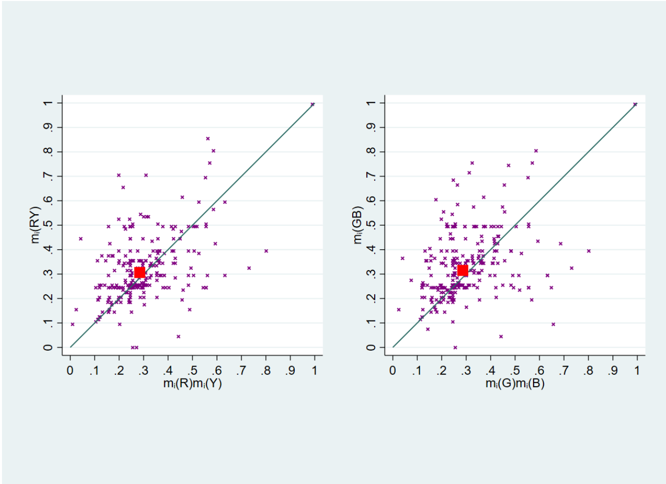
In Figures 4 and 5, the negative 45-degree line corresponds to $QCU_i = 0$ or $QCG_i = 0$ so that any point above this line presents a violation of QCU_i, $QCG_i \geq 0$. According to Figure 4, only 5.0% (=18/361) of the observations fall in the region $QCU_i < 0$ for the urns. We also see strong positive correlation between $m_i(RY)$ and $m_i(GB)$ (Pearson correlation 0.737, $p < 0.01$). We have a completely different picture when it comes to game prediction. In Figure 5, 44.9% (=132/294) of the observations fall into the region $QCG_i < 0$ for the games, and strong *negative* correlation exists between $m_i(A_1A_2)$ and $m_i(B_1B_2)$ (Pearson correlation: −0.594 (SH0.2), −0.782 (SH0.8), −0.666 (PD0.1), and −0.400 (PD0.5), $p < 0.01$ for each treatment.)

5.5 Coordination Premium and Correlation

Result 4 *i) (Positive CP for the games)* CPG^A, $CPG^B > 0$ *in every treatment. Moreover, CPG tends to be larger for the Pareto efficient profile.*

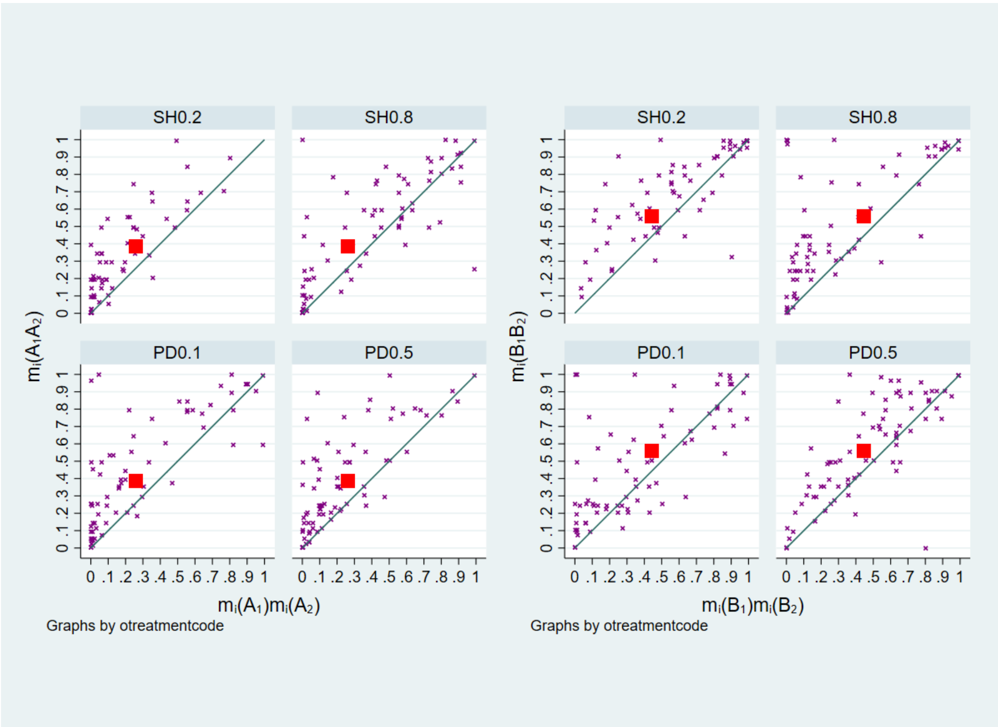
ii) (Ambiguous correlation) Based on comparison between the simple and diagonal events, approximately half of observers have assessment incompatible with independence of the two players' actions.

Positive CP for the games Figures 8 and 9 plot the matching probability of a symmetric profile against the product of the matching probabilities of the corresponding elements. The vertical distance between each point and the 45-degree line in Figure 9 equals the coordination premium CPG_i (CPG_i^A in the upper panels and CPG_i^B in the lower panels). As seen, the predominant majority of observations lie above the 45-degree line in most panels. Table 13 presents the mean CP values for both games and urns, and rejects the null hypotheses $CPG^A = 0$ and $CPG^B = 0$ in all four treatments (two-sided). Despite the fact that the observers were shown the instructions of the player sessions which stated that it adopted perfect stranger matching, hence, they believe in the presence of some mechanism that helps the players coordinate their actions. We also see that this tendency is stronger for the Pareto efficient profile in three out of the four treatments (SH0.8, PD0.1 and PD0.5). A closely related observation is made by Cason et al. (2020), who elicit observers' likelihood rankings over the four outcomes of 2×2 games played by others, and conclude that the result is consistent only with positive correlation of the action choices by the two players. In contrast, CPU^{RY} and CPU^{GB} for the urns are positive in only four out of eight cases, and are much smaller in magnitude than CPG^A and CPG^B for the games.



Notes. The red square in each panel represents the average point.

Figure 8: $(m_i(R)m_i(Y), m_i(RY))$ and $(m_i(G)m_i(B), m_i(GB))$: all treatments combined



Notes. The red square in each panel represents the average point.

Figure 9: $(m_i(A_1)m_i(A_2), m_i(A_1A_2))$ and $(m_i(B_1)m_i(B_2), m_i(B_1B_2))$ by treatment

	Treatment			
	SH0.2	SH0.8	PD0.1	PD0.5
CPU^{RY}	0.018	0.024**	0.0131	0.0346**
	(0.115)	(0.110)	(0.107)	(0.137)
CPU^{GB}	0.006	0.050***	0.023*	0.034**
	(0.122)	(0.130)	(0.117)	(0.148)
$CPU^{RY} - CPU^{GB}$	0.011	-0.023**	-0.014	-0.008
#Obs.	90	87	85	88
CPG^{A}	0.122***	0.074***	**0.152*****	**0.154*****
	(0.140)	(0.186)	**(0.193)**	**(0.185)**
CPG^{B}	**0.106*****	**0.171*****	0.066***	0.109***
	(0.179)	**(0.240)**	(0.183)	(0.151)
$CPG^{A} - CPG^{B}$	0.021	-0.112**	0.088**	0.048*
#Obs.	66	62	61	71

Notes. a) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard deviations in parentheses. b) $CP = 0$ tested using two-sided t-test. c) The difference is computed for the subsample in which both terms are available (Footnote 20, Table 15), and is tested using two-sided t-test. d) Bold indicates the Pareto action profile (B_1B_2 in SH and A_1A_2 in PD).

Table 13: CP by treatment and by uncertainty source

Ambiguous correlation Epstein & Halevy (2019) examine preferences over bets on the colors of balls drawn from two urns, and find violation from PS in terms of the assessment of the likelihood of simple and diagonal events. Define

$$\begin{aligned}
OG_{\max} &= \max\{m_i^{A_1+A_2}, m_i^{B_1+B_2}\}, \\
OG_{\min} &= \min\{m_i^{A_1+A_2}, m_i^{B_1+B_2}\}, \\
TG_{\max} &= \max\{m_i(A_1A_2 \cup B_1B_2), m_i(A_1B_2 \cup B_1A_2)\}, \\
TG_{\min} &= \min\{m_i(A_1A_2 \cup B_1B_2), m_i(A_1B_2 \cup B_1A_2)\}.
\end{aligned}$$

If an observer's assessment satisfies binary complementarity for both the simple and diagonal events, we should have $BC = BC^d$ $(= 0)$, or equivalently, $OG_{\max} + OG_{\min} = TG_{\max} + TG_{\min}$. This implies that under PS, if $OG_{\max} < TG_{\max}$, then $OG_{\min} > TG_{\min}$, and if $OG_{\max} > TG_{\max}$, then $OG_{\min} < TG_{\min}$. Furthermore, under PS, if an observer believes that the two players' action choices are independent, then $OG_{\max} \geq TG_{\max}$ and $OG_{\min} \leq TG_{\min}$ should hold although the reverse implication is not true. In our notation, the two non-PS preference patterns reported by Epstein & Halevy (2019) are: "two $\succ$ one" with $OG_{\max} < TG_{\max}$ and $OG_{\min} < TG_{\min}$, and "one $\succ$ two" with $OG_{\max} > TG_{\max}$ and $OG_{\min} > TG_{\min}$.[38] Similar definitions apply to the preference

[38] "two $\succ$ one" implies $BCG^d < BCG$ and "one $\succ$ two" implies $BCG < BCG^d$.

patterns for the urns.[39]

For the urns, "one $\succ$ two" holds for the average matching probabilities ($p < 0.05$, paired t-test), and for about 40% of observers as seen in Table 18 in Appendix A in line with the findings of Epstein & Halevy (2019). The proportion of observers with assessment compatible with independence is slightly less for the urns. For the games, on the other hand, neither "one $\succ$ two" nor "two $\succ$ one" holds for the average matching probabilities when data from all the treatments are combined. This is mostly consistent with the decomposition in Table 18. The proportion of observers with assessment compatible with independence is approximately one half, and the other one half shows various patterns with the proportion of "two $\succ$ one" the highest among them in three treatments.

6 Approximate Sophistication

Even when the matching probabilities of an observer exhibit violation of PS, the degree of violation may be relatively small. In such a case, we may interpret the violation as a result of trembling in the choice of a response in the multiple price list rather than fundamental deviations from probabilistic thinking. We investigate this possibility by relaxing the (in)equality defining sophistication for each index by a fixed constant k, and call observer i k-probabilistically sophisticated (k-PS) if his matching probabilities m_i satisfy the modified inequality. More specifically, define

$$\begin{aligned} k_i^{BCU} &= |BCU_i|, \quad k_i^{MNU} = \max\{0, -MNU_i\}, \quad k_i^{QCU} = \max\{0, -QCU_i\}, \\ k_i^{BCG} &= |BCG_i|, \quad k_i^{MNG} = \max\{0, -MNG_i\}, \quad k_i^{QCG} = \max\{0, -QCG_i\}. \end{aligned}$$

k_i^{BCU} and k_i^{BCG} measure the extents to which observer i deviates from $BCU_i = 0$ and $BCG_i = 0$, respectively, in either direction. Likewise, k_i^{MNU}, k_i^{MNG}, k_i^{QCU} and k_i^{QCG} equal zero if observer i is PS in terms of the respective indices, but take positive values otherwise and hence measure the extent of deviation from PS. For $k \geq 0$, we say that an observer i is

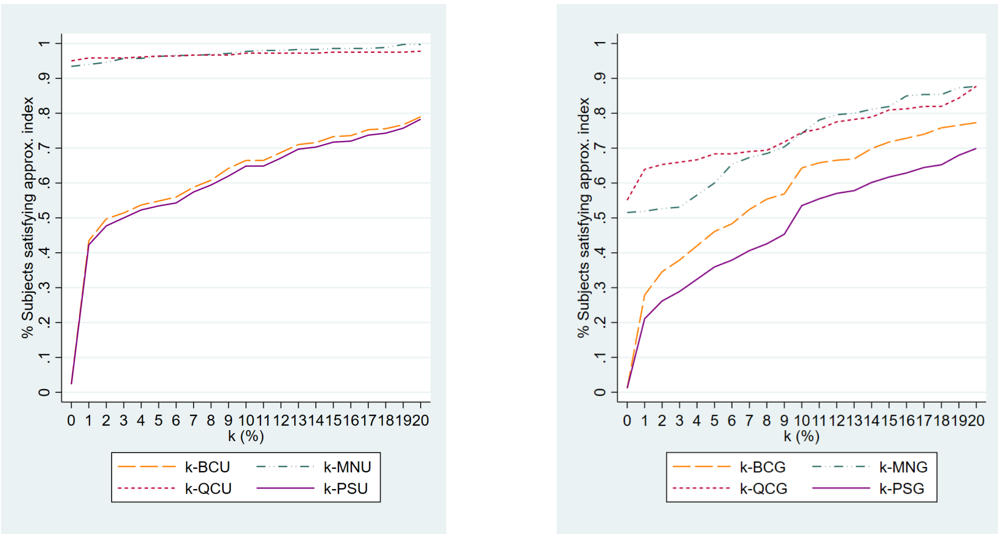
$$\begin{cases} k\text{-BC} \\ k\text{-MN} \\ k\text{-QC} \end{cases} \text{for the urns if} \quad \begin{cases} k_i^{BCU} \leq k \\ k_i^{MNU} \leq k \\ k_i^{QCU} \leq k \end{cases}.$$

An observer is k-PS for the urns if he is k-BC, k-MN, and k-QC for the urns. k-PS for the games is similarly defined. When $k = 0$, 0-PS is equivalent to PS elsewhere in the analysis. Figure 10 shows the proportions of observers who are k-PS for the urns and for the games as we vary k. As seen, the patterns are strikingly distinct between the

[39] The definitions are based on $OU_{\max} = \max\{m_i^{R+Y}, m_i^{G+B}\}$, $OU_{\min} = \min\{m_i^{R+Y}, m_i^{G+B}\}$, $TU_{\max} = \max\{m_i(RY \cup GB), m_i(RB \cup GY)\}$, and $TU_{\min} = \min\{m_i(RY \cup GB), m_i(RB \cup GY)\}$.

two uncertainty sources. In particular, the k-PS curve coincides with the k-BC curve for the urns, implying that k-BC is the defining condition for k-PS. For the games, on the other hand, the k-PS curve is below the other three curves by a significant margin for $k = 0.01, \ldots, 0.2$ ($p < 0.10$ against each of k-BC, k-MN and k-QC, t-test), implying that the three indices bind for different observers. Among them, k-BC is least likely satisfied. Between the two uncertainty sources, the proportion of k-PS observers is significantly lower for the games than for the urns for each value of k ($k = 0.01, 0.02, \ldots, 0.2$, $p < 0.05$, t-test).

Figure 11 shows the proportion of k-PS for the games by index and by treatment. While the difference across treatments is rarely significant, we see that the treatments are ordered remarkably consistently across indices with PD0.5 at the bottom and either SH0.2 or PD0.1 at the top for the most part. We also notice that for most values of k, PD0.1 lies above PD0.5 for every index, and SH0.2 lies above SH0.8 except QCG.



(a) Urns

(b) Games

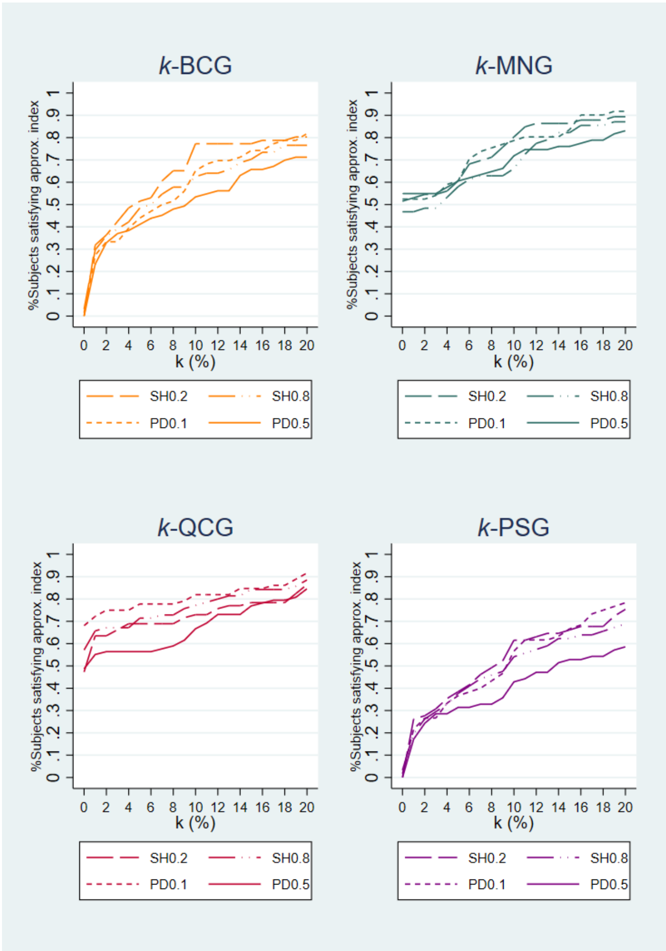
Notes. The horizontal axis is marked by $100k$.

Figure 10: k-PS by source

How does the degree of PS for the urns of a given observer correlate with the degree of his PS for the games? Table 19 in Appendix A presents OLS regressions of the degrees of deviation from PS for the games on those for the urns.[40] As seen, correlation tends to be positive although significance is weak (if any) except for MN.[41]

[40] The three models on the right include the dummy $1_{\{F_i=g\}}$ which equals one if an observer states that the game was easier to predict.

[41] In contrast with the finding in Section 5.3 that the tendency of an observer to violate MN for the



Notes. The horizontal axis is marked by $100k$.

Figure 11: k-PS by treatment

7 A Model of the Observers' Matching Probabilities

It is difficult to give a definitive account for the difference in the patterns of violation of PS, and monotonicity in particular, between the urns and games. As one attempt, we hypothesize here that an observer's matching probability $m_i(E)$ of any event E is computed in reference to some event which varies with E. In the terminology of the theory of ambiguity preferences, we suppose that the *predictive priors* have moving support as event E varies. Formally, for any event E, the *reference event of* E for i, denoted $\Lambda_i(E)$, is the event which observer i uses as a reference when computing the matching probability of E. We assume that $E \subset \Lambda_i(E)$ so that a reference event always contains the event in question.

Given event E and its reference event $\Lambda_i(E)$, we assume that i's matching probability $m_i(E)$ is given by

$$m_i(E) = \frac{\sigma_i(E)}{\sigma_i(\Lambda_i(E))},$$

where σ_i is the *support function* defined over the collection of events: $\sigma_i(E) > 0$ is the significance of the episodes that observer i associates with event E, and is similar to the "support value" of event E as assumed by Tversky & Koehler (1994). We assume that σ_i is monotone if $\sigma_i(E) \geq \sigma_i(E')$ if $E \supset E'$, but do not necessarily assume that it is additive. It follows that we may have $m_i(E) + m_i(E') \neq 1$ even when E and E' are complementary to each other.[42]

We hypothesize that the reference event of each event is given as follows:

a) (simple) $\Lambda_i(E) = \Omega$ if E is a simple urn or game event.

b) (conjunctional) $\Lambda_i(RY) = \Lambda_i(GB) = \Omega$, $\Lambda_i(A_1A_2) = \Lambda_i(B_1B_2) = A_1A_2 \cup B_1B_2$.

c) (diagonal) $\Lambda_i(E) = \Omega$ if E is a diagonal urn or game event.

The only difference between the games and urns is hence the reference events of the conjunctional events (b): While the observers take into consideration all four possible combinations of ball colors when contemplating on a conjunctional urn event, they take into consideration only the two coordination profiles when contemplating on a conjunctional game event. In this formulation, the violation of monotonicity for the games is not

games is independent of whether or not he violates MN for the urns, the positive correlation between k_i^{MNU} and k_i^{MNG} in Table 19 shows that an observer who violates MN by a large margin for the urns tends to do so for the games as well. The positive correlation between k_i^{BCU} and k_i^{BCG} in Model (1) of Table 19 corresponds to the finding in Table 17.

[42]For example, even if the reference set of both A_1 and B_1 equals Ω ($\Lambda_i(A_1) = \Lambda_i(B_1) = \Omega$), we have $BC_i(A_1) = 1 - m_i(A_1) - m_i(B_1) < 0$ if σ_i is sub-additive ($\sigma_i(A_1) + \sigma_i(B_1) < \sigma_i(\Omega)$), and $BC_i(A_1) > 0$ if σ_i is super-additive ($\sigma_i(A_1) + \sigma_i(B_1) > \sigma_i(\Omega)$).

inconsistent with the monotonicity of σ_i. Furthermore, it is consistent with the negative correlation between $m_i(A_1A_2)$ and $m_i(B_1B_2)$ found in Figure 5 in Section 5.4.[43]

Under the above hypothesis, quasi-complementarity $QCG_i = 1 - m_i(A_1A_2) - m_i(B_1B_2)$ is similar to binary complementarity BCG_i in the sense that A_1A_2 and B_1B_2 together constitute the reference event of each conjunctional event just like A_1 and B_1 together constitute the reference event of each simple event. In this sense, we would obtain indirect and partial support for our hypothesis if QCG_i and BCG_i correlate with each other at the individual level. Table 14 confirms this by classifying observers based on their values of QCG_i and BCG_i as well as BCG_i^d for the diagonal events. The top panel shows that correlation is positive between the signs of BCG_i and QCG_i in every treatment ($p < 0.10$, Fisher's exact test). Under the above hypothesis, this implies that an observer's support function σ_i tends to be sub-additive for simple events ($\sigma_i(\Omega) < \sigma_i(A_1) + \sigma_i(B_1)$) if and only if it is so for conjunctional events ($\sigma_i(A_1A_2 \cup B_1B_2) < \sigma_i(A_1A_2) + \sigma_i(B_1B_2)$).[44] The middle panel in Table 14 uses BCG_i^d for diagonal events and finds the same tendency at $p = 0.05$ except in PD0.5. The bottom panel of Table 14 checks if the observers tend to have $QCG_i < 0$ if they found the outcome of the games easier to predict (i.e., $F_i = g$) as in the case of BCG_i in Result 1(ii) on binary complementarity in Section 5.2. It shows that this is true in SH0.2 ($p < 0.05$) and marginally so in SH0.8 ($p = 0.126$).[45] This is in sharp contrast with Table 20 in Appendix A, where *no* correlation is found between the signs of BCU_i and QCU_i.

8 Conclusion

The elicited matching probabilities violate the law of probability including monotonicity and additivity for both strategic and non-strategic events, but the patterns of deviations are substantially different between the two uncertainty sources. The most stark difference between the non-strategic and strategic events is the coordination fallacy in terms of the monotonicity (MN) index. The violation of monotonicity (i.e., strictly negative MN

[43]Note that $m_i(A_1A_2) > m_i(A_1) \Leftrightarrow \frac{\sigma_i(A_1A_2)}{\sigma_i(A_1A_2 \cup B_1B_2)} > \frac{\sigma_i(A_1)}{\sigma_i(\Omega)}$ can hold even when $\sigma_i(\Omega) > \sigma_i(A_1) > \sigma_i(A_1A_2)$ and $\sigma_i(\Omega) > \sigma_i(A_1A_2 \cup B_1B_2) > \sigma_i(A_1A_2)$.

[44]By (b), we have

$$m_i(A_1A_2) = \frac{\sigma_i(A_1A_2)}{\sigma_i(A_1A_2 \cup B_1B_2)} \quad \text{and} \quad m_i(B_1B_2) = \frac{\sigma_i(B_1B_2)}{\sigma_i(A_1A_2 \cup B_1B_2)}. \tag{5}$$

Hence, sub-additivity $\sigma_i(A_1A_2 \cup B_1B_2) < \sigma_i(A_1A_2) + \sigma_i(B_1B_2)$ implies $QC_i = 1 - m_i(A_1A_2) - m_i(B_1B_2) < 0$. Likewise, sub-additivity $\sigma_i(\Omega) < \sigma_i(A_i) + \sigma_i(B_i)$ for $i = 1, 2$ implies $BC_i(A_1) < 0$.

[45]The positive impact of the dummy variable $1_{\{F_i=g\}}$ on k_i^{QCG} in Table 19 also supports the conjecture that those observers who find the games easier to predict are more likely to use the symmetric profiles A_1A_2 and B_1B_2 as their reference events when asked about the matching probability of one of those events.

	SH0.2		SH0.8		PD0.1		PD0.5	
	QCG_i							
BCG_i	+	−	+	−	+	−	+	−
+	20	6	19	6	19	5	15	6
−	12	28	16	21	20	17	19	31
p	0.000		0.018		0.059		0.018	
BCG_i^d	+	−	+	−	+	−	+	−
+	13	5	19	2	18	3	17	11
−	21	32	19	28	29	19	21	28
p	0.028		0.000		0.050		0.159	
F_i	+	−	+	−	+	−	+	−
g	21	34	22	22	25	14	21	27
u	14	5	17	7	21	8	16	10
p	0.015		0.126		0.602		0.223	

Table 14: Association between QCG_i and BCG_i, BCG_i^d, and F_i

values) is severe for the strategic events, and is caused by the excessively large matching probabilities placed on the symmetric coordination profiles of the games: Approximately one half of observers exhibit the coordination fallacy for at least one of the two symmetric profiles. The average MN values tend to be smaller for the Pareto efficient profile. In contrast, the violation of monotonicity is rare for the non-strategic events.

The weights placed on the two symmetric profiles of the games are so large that their sum exceeds one in half of the cases. This, along with the observation on the negative correlation between the matching probabilities of the two symmetric profiles, leads us to the conjecture that the subjects focus only on the two symmetric profiles when contemplating on the matching probability of such a profile. Under this hypothesis, the quasi-complementarity (QC) index, defined to be one minus the sum of the two matching probabilities of the two symmetric profiles, is similar in its functional form to the BC index of individual action choices. We find data to be consistent with this implication.

While the difference between the two types of uncertainty sources is substantial, the difference across the four game treatments is less clear. While Pareto dominance is found to induce the violation of monotonicity more often, risk dominance in the case of the SH games seems to have little impact on the predictions. There is also no obvious relationship between the predictions and the empirical statistics of how the games were played in the players sessions. One possibility is that the across-treatment variation is hidden behind the extremely large within-treatment variations in the predictions.

This is a first exploration into the perception of strategic uncertainty, and there are a number of possible extensions. We used stag-hunt and prisoners' dilemma games for the creation of strategic uncertainty. Given that both these games have symmetric profiles as focal outcomes, they may induce violation of monotonicity more strongly than some other games. The perception of strategic uncertainty in other games is the most obvious extension of the present analysis. In the sense that many of the strategic uncertainties in the reality involve more than two decision makers, it will be of practical importance to use games with a larger number of players.

Although we have focused attention on bets with a positive reward, the literature observes that uncertainty attitudes are often different in the positive and negative domains. Whether violation of PS is severer or not in the negative domain is an open question. In our experiments, the observers make predictions based on the information that the player subjects are students of the same university. An interesting extension that corresponds to the testing of the familiarity hypothesis would involve players and observers from totally different social backgrounds.

References

Abdellaoui, M., Baillon, A., Placido, L. & Wakker, P. P. (2011), 'The rich domain of uncertainty: Source functions and their experimental implementation', *American Economic Review* **101**(2), 695–723. 2

Ahn, D., Choi, S., Gale, D. & Kariv, S. (2014), 'Estimating ambiguity aversion in a portfolio choice experiment', *Quantitative Economics* **5**, 195–223. 6

Al-Najjar, N. I. & Weinstein, J. (2009), 'The ambiguity aversion literature: A critical assessment', *Economics and Philosophy* **25**(3), 249–284. 10

Baillon, A. & Bleichrodt, H. (2015), 'Testing ambiguity models through the measurement of probabilities for gains and losses', *American Economic Journal: Microeconomics* **7**(2), 77–100. 2, 3.3, 4.2, 22, 5.2, 29

Baillon, A. & Emirmahmutoglu, A. (2018), 'Zooming in on ambiguity attitudes', *International Economic Review* **59**(4), 2107–2131. 2, 3.3, 20

Baillon, A., Huang, Z., Selim, A. & Wakker, P. P. (2018), 'Measuring ambiguity attitudes for all (natural) events', *Econometrica* **86**(5), 1839–1858. 3, 2, 9, 5.2

Blanco, M., Engelmann, D., Koch, A. K. & Normann, H.-T. (2010), 'Belief elicitation in experiments: is there a hedging problem?', *Experimental Economics* **13**(4), 412–438. 11

Camerer, C. F. & Karjalainen, R. (1994), *Ambiguity-aversion and non-additive beliefs in non-cooperative games: Experimental evidence*, Springer Netherlands, Dordrecht, pp. 325–358. 2

Cason, T. N., Sharma, T. & Vadovič, R. (2020), 'Correlated beliefs: Predicting outcomes in 2×2 games', *Games and Economic Behavior* **122**, 256–276. 12, 5.5

Charness, G., Karni, E. & Levin, D. (2010), 'On the conjunction fallacy in probability judgment: New experimental evidence regarding Linda', *Games and Economic Behavior* **68**(2), 551–556. 5

Chew, S. H., Ebstein, R. P. & Zhong, S. (2012), 'Ambiguity aversion and familiarity bias: Evidence from behavioral and gene association studies', *Journal of Risk and Uncertainty* **44**(1), 1–18. 2, 29

Chew, S. H., Miao, B. & Zhong, S. (2017), 'Partial ambiguity', *Econometrica* **85**(4), 1239–1260. 2, 6

Chew, S. H., Ratchford, M. & Sagi, J. S. (2018), 'You need to recognise ambiguity to avoid it', *Economic Journal* **128**(614), 2480–2506. 5.3

Dal Bó, P. & Fréchette, G. R. (2018), 'On the determinants of cooperation in infinitely repeated games: A survey', *Journal of Economic Literature* **56**(1), 60–114. 3.2

Dimmock, S. G., Kouwenberg, R. & Wakker, P. P. (2016), 'Ambiguity attitudes in a large representative sample', *Management Science* **62**(5), 1363–1380. 2

Dominiak, A. & Duersch, P. (2019), 'Interactive Ellsberg tasks: An experiment', *Journal of Economic Behavior & Organization* **161**, 145–157. 10

Eichberger, J., Kelsey, D. & Schipper, B. C. (2008), 'Granny versus game theorist: Ambiguity in experimental games', *Theory and Decision* **64**(2-3), 333–362. 2

Epstein, L. G. & Halevy, Y. (2019), 'Ambiguous correlation', *Review of Economic Studies* **86**(2), 668–693. 5.5

Epstein, L. G. & Schneider, M. (2010), 'Ambiguity and asset markets', *Annual Review of Financial Economics* **2**(1), 315–346. 8

Fischbacher, U. (2007), 'z-Tree: Zurich toolbox for ready-made economic experiments', *Experimental Economics* **10**(2), 171–178. 3.1

Fox, C. R. & Tversky, A. (1995), 'Ambiguity aversion and comparative ignorance', *Quarterly Journal of Economics* **110**(3), 585–603. 5.2

Fox, C. R. & Weber, M. (2002), 'Ambiguity aversion, comparative ignorance, and decision context', *Organizational Behavior and Human Decision Processes* **88**(1), 476–498. 29

French, K. R. & Poterba, J. M. (1991), Investor diversification and international equity markets, Working Paper 3609, National Bureau of Economic Research. 8, 28

Gächter, S. & Renner, E. (2010), 'The effects of (incentivized) belief elicitation in public goods experiments', *Experimental Economics* **13**(3), 364–377. 11

Gigerenzer, G. (1996), 'On narrow norms and vague heuristics: A reply to Kahneman and Tversky.'. 5

Greiner, B. (2015), 'Subject pool recruitment procedures: Organizing experiments with ORSEE', *Journal of the Economic Science Association* **1**(1), 114–125. 3.1

Halevy, Y. (2007), 'Ellsberg revisited: An experimental study', *Econometrica* **75**(2), 503–536. 2, 6

Heath, C. & Tversky, A. (1991), 'Preference and belief: Ambiguity and competence in choice under uncertainty', *Journal of Risk and Uncertainty* **4**(1), 5–28. 5.2, 29

Heinemann, F., Nagel, R. & Ockenfels, P. (2009), 'Measuring strategic uncertainty in coordination games', *Review of Economic Studies* **76**(1), 181–221. 2

Hoffmann, T. (2014), The effect of belief elicitation game play, Annual Conference 2014 (Hamburg): Evidence-based Economic Policy 100483, Verein für Socialpolitik/German Economic Association. 11

Hyndman, K., Terracol, A. & Vaksmann, J. (2009), 'Learning and sophistication in coordination games', *Experimental Economics* **12**(4), 450–472. 12

Kocher, M. G., Lahno, A. M. & Trautmann, S. T. (2018), 'Ambiguity aversion is not universal', *European Economic Review* **101**, 268–283. 2, 5.2

Li, C., Turmunkh, U. & Wakker, P. P. (2019), 'Trust as a decision under ambiguity', *Experimental Economics* **22**(1), 51–75. 9, 5.2

Li, Z., Loomes, G. & Pogrebna, G. (2017), 'Attitudes to uncertainty in a strategic setting', *Economic Journal* **127**(601), 809–826. 2

Moffatt, P. G. (2015), *Experimetrics: Econometrics for Experimental Economics*, Macmillan International Higher Education. 37

Oechssler, J., Roider, A. & Schmitz, P. W. (2009), 'Cognitive abilities and behavioral biases', *Journal of Economic Behavior & Organization* **72**(1), 147–152. 5.3, 34

Oechssler, J. & Roomets, A. (2015), 'A test of mechanical ambiguity', *Journal of Economic Behavior & Organization* **119**, 153–162. 27

Offerman, T., Sonnemans, J., Van de Kuilen, G. & Wakker, P. P. (2009), 'A truth serum for non-bayesians: Correcting proper scoring rules for risk attitudes', *Review of Economic Studies* **76**(4), 1461–1489. 4

Palfrey, T. R. & Wang, S. W. (2009), 'On eliciting beliefs in strategic games', *Journal of Economic Behavior & Organization* **71**(2), 98–109. 12

Raiffa, H. (1968), *Decision Analysis: Introductory Lectures on Choices under Uncertainty*, Addison-Wesley. 9

Rankin, F. W., Huyck, J. B. V. & Battalio, R. C. (2000), 'Strategic similarity and emergent conventions: Evidence from similar stag hunt games', *Games and Economic Behavior* **32**(2), 315–337. 15

Ritov, I. & Baron, J. (1990), 'Reluctance to vaccinate: Omission bias and ambiguity', *Journal of Behavioral Decision Making* **3**(4), 263–277. 8

Schlag, K. H., Tremewan, J. & Van der Weele, J. J. (2015), 'A penny for your thoughts: A survey of methods for eliciting beliefs', *Experimental Economics* **18**(3), 457–490. 11

Schmeidler, D. (1989), 'Subjective probability and expected utility without additivity', *Econometrica* **57**(3), 571–587. 4

Trautmann, S. T. & van de Kuilen, G. (2015*a*), *Ambiguity Attitudes*, John Wiley & Sons, Ltd, chapter 3, pp. 89–116. 1, 2, 7

Trautmann, S. T. & van de Kuilen, G. (2015*b*), 'Belief elicitation: A horse race among truth serums', *Economic Journal* **125**(589), 2116–2135. 5.2, 27

Tversky, A. & Kahneman, D. (1982), *Judgments of and by Representativeness*, Cambridge University Press, pp. 84–98. 5

Tversky, A. & Koehler, D. J. (1994), 'Support theory: A nonextensional representation of subjective probability', *Psychological Review* **101**(4), 547–567. 7

Wakker, P. P. (2010), *Prospect Theory: For Risk and Ambiguity*, Cambridge University Press. 9

Yang, C.-L. & Yao, L. (2017), 'Testing ambiguity theories with a mean-preserving design', *Quantitative Economics* **8**(1), 219–238. 6

Zizzo, D. J., Stolarz-Fantino, S., Wen, J. & Fantino, E. (2000), 'A violation of the monotonicity axiom: Experimental evidence on the conjunction fallacy', *Journal of Economic Behavior & Organization* **41**(3), 263–276. 5

Appendices

A Tables and Figures

	Index	#Observations after censoring				
		SH0.2	SH0.8	PD0.1	PD0.5	total
MN	MNU_i^{RY}	92	87	86	88	353
	MNU_i^{GB}	90	88	88	91	357
	$MNU_i^{RY} - MNU_i^{GB}$	90	87	85	88	350
	MNG_i^{A}	75	73	70	77	295
	MNG_i^{B}	69	66	69	76	280
	$MNG_i^{A} - MNG_i^{B}$	66	62	61	71	260
	MNU_i	90	87	85	88	350
	MNG_i	66	62	61	71	260
	$MNU_i - MNG_i$	65	60	60	69	254
BC	BCU_i	90	88	85	89	352
	BCG_i	66	64	66	73	269
	$BCU_i - BCG_i$	65	62	64	70	261
	BCU_i^d	92	90	89	90	361
	BCG_i^d	75	72	75	80	302
	$BCU_i^d - BCG_i^d$	75	72	74	78	269
QC	QCU_i	92	89	90	90	361
	QCG_i	74	70	72	78	294
	$QCU_i - QCG_i$	74	70	72	77	293
CP	CPU_i^{RY}	92	87	86	88	353
	CPU_i^{GB}	90	88	88	91	357
	$CPU_i^{RY} - CPU_i^{GB}$	90	87	85	88	350
	CPG_i^{A}	75	73	70	77	295
	CPG_i^{B}	69	66	69	76	280
	$CPG_i^{A} - CPG_i^{B}$	66	62	61	71	260

Notes. See Footnote 20 for the description of the censoring criterion.

Table 15: Size of subsamples after censoring

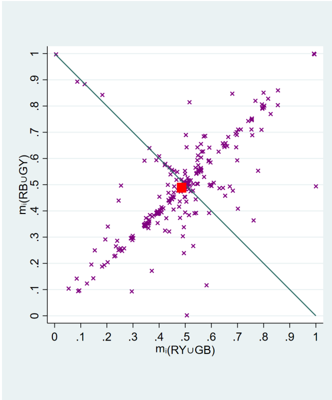


Figure 12: $(m_i(RY \cup GB), m_i(RB \cup GY))$: all treatments combined

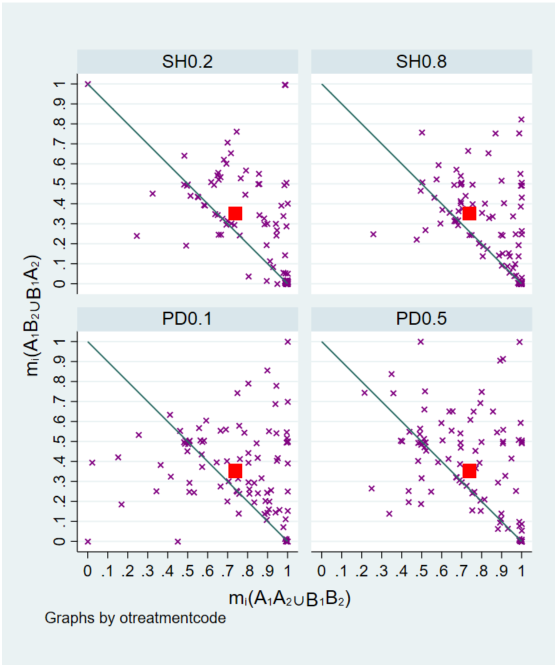


Figure 13: $(m_i(A_1A_2 \cup B_1B_2), m_i(A_1B_2 \cup B_1A_2))$ by treatment

Table 16 summarizes p-values for paired t-test for the difference in the average matching probabilities of pairs of events that are symmetric with respect to the urns or the players. It finds no statistical difference in the matching probabilities of these pairs of events except in one case ($p < 0.05$).

Hypothesis	Treatment			
	SH0.2	SH0.8	PD0.1	PD0.5
$m(R) = m(Y)$	0.1399 (92)	0.1485 (88)	0.9233 (86)	0.3135 (89)
$m(G) = m(B)$	0.1615 (90)	0.2910 (89)	0.5329 (88)	0.1909 (91)
$m(RY) = m(GB)$	0.5586 (92)	0.0601 (89)	0.4987 (90)	0.9230 (90)
$m(RY \cup GB) = m(RB \cup GY)$	0.7863 (92)	0.9208 (90)	0.5330 (89)	0.5548 (90)
$m(A_1) = m(A_2)$	0.0079 (75)	0.6801 (75)	0.7415 (72)	0.2906 (77)
$m(B_1) = m(B_2)$	0.7010 (70)	0.1409 (68)	0.7567 (73)	0.6629 (79)

Notes. The number in the parentheses is the size of the subsample in which both matching probabilities are available after data censoring described in Footnote 20.

Table 16: Paired t-test of event exchangeability.

	SH0.2			SH0.8			PD0.1			PD0.5		
	$\|BCG_i\| \leq 0.1$	$\|BCG_i\| > 0.1$	total	$\|BCG_i\| \leq 0.1$	$\|BCG_i\| > 0.1$	total	$\|BCG_i\| \leq 0.1$	$\|BCG_i\| > 0.1$	total	$\|BCG_i\| \leq 0.1$	$\|BCG_i\| > 0.1$	total
$\|BCU_i\| \leq 0.1$	33	8	41	33	12	45	30	11	41	28	15	43
$\|BCU_i\| > 0.1$	18	6	24	6	11	17	13	10	23	9	18	27
total	51	14	65	39	23	62	43	21	64	37	33	70
p-value	0.603			0.006			0.173			0.010		

Table 17: Association between BCU_i and BCG_i

	$O_{\max} < T_{\max}$ $O_{\min} > T_{\min}$	$O_{\max} \geq T_{\max}$ $O_{\min} \leq T_{\min}$	$O_{\max} > T_{\max}$ $O_{\min} > T_{\min}$	$O_{\max} < T_{\max}$ $O_{\min} < T_{\min}$	total
Urns	0.05	0.37 (0.06)	0.38	0.17	0.97
SH0.2	0.08	0.54 (0.31)	0.05	0.26	0.92
SH0.8	0.16	0.40 (0.25)	0.13	0.22	0.91
PD0.1	0.13	0.49 (0.36)	0.13	0.18	0.93
PD0.5	0.17	0.49 (0.32)	0.18	0.11	0.95

Notes. $O = OU$ and $T = TU$ for the urns (all treatments combined), and $O = OG$ and $T = TG$ for the games. The independence of the two players' action choices implies, but is not implied by, the second case with $O_{\max} \geq T_{\max}$ and $O_{\min} \leq T_{\min}$. The numbers in the parentheses represent the proportions of observations satisfying strict inequalities.

Table 18: Types based on the assessment of simple and diagonal events

B Entropy and Familiarity

We may conjecture that more dispersion in play makes prediction more difficult, and hence implies less confidence in prediction. This in turn leads to the conjecture that a higher entropy implies a higher degree of ambiguity aversion. We hence hypothesize that BCG_i is larger if the game has a higher entropy. Figure 14 shows the proportion

Models	(1)	(2)	(3)	(4)	(5)	(6)
Dept. var.	k_i^{BCG}	k_i^{MNG}	k_i^{QCG}	k_i^{BCG}	k_i^{MNG}	k_i^{QCG}
k_i^{BCU}	0.089*			0.224		
	(0.048)			(0.213)		
k_i^{MNU}		0.261**			1.907***	
		(0.120)			(0.185)	
k_i^{QCU}			0.123			-0.013
			(0.146)			(0.039)
$1_{\{F_i=g\}}$				2.924	-0.154	4.151***
				(2.307)	(1.924)	(1.354)
$k_i^{BCU} * 1_{\{F_i=g\}}$				-0.162		
				(0.240)		
$k_i^{MNU} * 1_{\{F_i=g\}}$					-1.783***	
					(0.287)	
$k_i^{QCU} * 1_{\{F_i=g\}}$						0.332
						(0.225)
Constant	11.14***	7.329***	7.381***	8.952***	7.311***	4.412***
	(0.668)	(0.736)	(0.943)	(1.449)	(1.800)	(1.278)
Observations	261	254	293	254	247	283
R-squared	0.007	0.002	0.006	0.017	0.011	0.041

Notes. a) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. b) Robust standard errors clustered by session in parentheses. c) $1_{\{F_i=g\}} = 1$ if observer i states that the game was easier to predict than the urns.

Table 19: Association of deviations from PS between the urns and games: OLS regressions

	SH0.2		SH0.8		PD0.1		PD0.5	
	QCU_i							
BCU_i	+	−	+	−	+	−	+	−
+	38	1	45	1	48	2	43	2
−	50	3	41	2	35	5	43	2
p	0.635		0.608		0.235		1.000	
BCU_i^d	+	−	+	−	+	−	+	−
+	55	0	49	0	55	2	52	1
−	33	4	37	3	27	5	33	3
p	0.024		0.087		0.093		0.299	
F_i	+	−	+	−	+	−	+	−
g	66	3	59	1	47	4	56	4
u	22	1	24	2	33	2	26	0
p	1.000		0.216		1.000		0.310	

Table 20: Association between QCU_i and BCU_i, BCU_i^d, and F_i

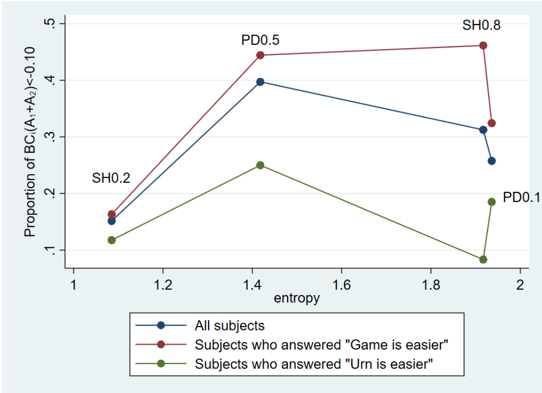


Figure 14: Entropy and the proportion of observers with $BC_i < -0.10$

of observers with $BC_i < -0.10$ plotted against the realized entropy of each game. The red line corresponds to the observers with $F_i = g$, the green line to the observers with $F_i = u$, and the blue line to both types. While we find no trend when both types are combined, there exist a positive trend for $F_i = g$, and a negative trend for $F_i = u$. In other words, among those observers who perceive game prediction to be easier, a higher entropy implies a higher proportion of $BCG_i < -0.10$ contrary to our conjecture.

C Matching Probabilities and Certainty Equivalents

One important question concerns whether violation of PS is observed even when we use measures other than matching probabilities. To answer this question, we use data from Part 3 of the observer sessions to check if monotonicity is violated also in terms of certainty equivalents. For event E, let $c_i(E)$ denote observer i's certainty equivalent of the bet 1000_E0. Part 3 of the observer experiments elicits $c_i(E)$ for the urn events $E = G$ as well as the game events $E = B_1$ and B_1B_2.[46] Given this limited set of questions, violation of MN is defined by $MNG_i(B_1, B_1B_2) \equiv m_i(B_1) - m_i(B_1B_2) < 0$, and violation of monotonicity in terms of certainty equivalents is defined by $MNGC_i(B_1, B_1B_2) \equiv c_i(B_1) - c_i(B_1B_2) < 0$. As seen in Table 21, 28.5% (= 81/284) of observers exhibit violation of monotonicity in terms of certainty equivalents of these events, and the proportion is higher than that in terms of matching probabilities (18.4% = 52/284).[47]

The scatter plots in Figure 15 show positive correlation between the difference in

[46]See Footnote 18.

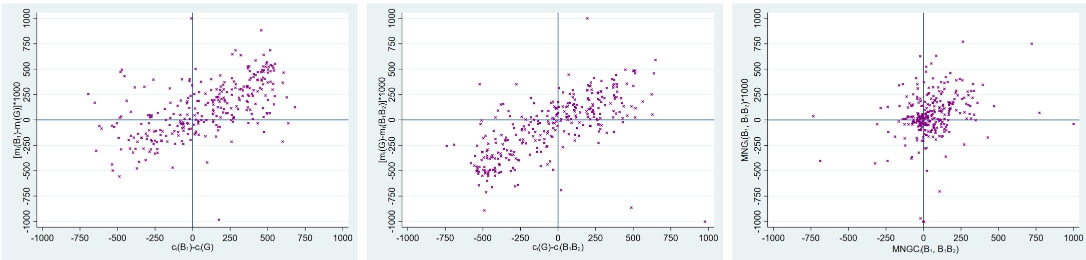
[47]We should however note that there is little correlation between those who violate MN in terms of matching probabilities and those who violate MN in terms of certainty equivalents. On the other hand, the predominant majority are PS among those who show consistent responses between the two measures of MN.

$MNGC_i(B_1, B_1B_2)$	$MNG_i(B_1, B_1B_2)$ < 0	≥ 0	total
< 0	18	63	81
≥ 0	34	169	203
total	52	232	284
p	0.282		

Notes. p-value for χ^2 test.

Table 21: Association between $MNG_i(B_1, B_1B_2)$ and $MNGC_i(B_1, B_1B_2)$

certainty equivalents against the difference in matching probabilities. This implies the first-order stochastic dominance (FOSD) property of observers' preferences, which justifies the use of multiple price list in the elicitation of their matching probabilities.[48]



Notes. Left panel: $(c_i(B_1) - c_i(G),\, m_i(B_1) - m_i(G))$, middle panel: $(c_i(G) - c_i(B_1B_2),\, m_i(G) - m_i(B_1B_2))$, right panel: $(MNGC_i(B_1, B_1B_2),\, MNG_i(B_1, B_1B_2))$. Pearson correlation = 0.59, 0.67, and 0.27, respectively, with $p = 0.000$ for all three cases.

Figure 15: Matching probabilities and certainty equivalents: All treatments combined

[48]Observer i's preferences $\succeq_i$ over bets satisfy satisfy first-order stochastic dominance (FOSD) if for any two events E and E', $m_i(E') \geq m_i(E)$ implies $1000_{E'}0 \succeq_i 1000_E 0$. Since i is indifferent between receiving $c_i(E)$ for sure and $1000_E 0$, FOSD holds if for any two events E and E', $m_i(E') \geq m_i(E)$ implies $c_i(E') \geq c_i(E)$. It follows that for any E and E', positive correlation between $m_i(E') - m_i(E)$ and $c_i(E') - c_i(E)$ implies consistency with FOSD.