Consciously stochastic in preference reversals

Author(s):
Shi, Liu; Qiu, Jianying; Li, Jiangyan; Bohn, Frank

Publication date:
2024-06

Permanent link:
https://doi.org/https://doi.org/10.3929/ethz-b-000677299

Originally published in:
Journal of Risk and Uncertainty 68(3), https://doi.org/10.1007/s11166-024-09430-w

Journal of Risk and Uncertainty (2024) 68:255–297
https://doi.org/10.1007/s11166-024-09430-w

Consciously stochastic in preference reversals

Liu Shi[1] · **Jianying Qiu**[1] · **Jiangyan Li**[2] · **Frank Bohn**[1,3]

Accepted: 13 March 2024 / Published online: 31 May 2024

Abstract

Stochastic choice, the act of choosing differently in repeated decisions, can be a conscious decision made by individuals who are aware of their inability to make a definitive choice. To examine the prevalence and implications of conscious stochastic choice, we developed a novel method and implemented it in a preference reversal experiment: In each valuation choice between the bet and a varying reference option, subjects could either pay a small cost to select a specific option or opt for a free randomization choice where a computer randomly selects an option. Our findings revealed that the majority of subjects exhibited conscious stochastic choice, and further that their choices were significantly affected by the elicitation procedures.

Keywords Stochastic choice · procedure invariance · preference reversals · incomplete preferences · imprecise preferences · deliberately stochastic

JEL Classification B40 · C91 · D81

✉ Jianying Qiu
jianying.qiu@ru.nl

✉ Jiangyan Li
jiangyanli@dufe.edu.cn

Liu Shi
liu.shi@ru.nl

Frank Bohn
frank.bohn@ru.nl

[1] Department of Economics, IMR, Radboud University, Heyendaalseweg 141, Nijmegen 6525, AJ, Netherlands

[2] School of Finance, Laboratory of Experimental Economics, Dongbei University of Finance and Economics, Dalian, China

[3] KOF Swiss Economic Institute, ETH Zürich, Switzerland

1 Introduction

Individuals often make stochastic choices, selecting different options in repeated decisions (Mosteller & Nogee, 1951; Camerer, 1989; Starmer & Sugden, 1989), even when they are informed of the repetitions in advance (Agranov & Ortoleva, 2017). Stochastic choice can be either unconscious or conscious. Unconscious stochastic choice occurs when random preference shocks influence decision-making, as in random (expected) utility models (Luce & Suppes, 1965; Eliashberg & Hauser, 1985; Gul & Pesendorfer, 2006; Apesteguia & Ballester, 2018), where individuals have a clear preference in each decision and are not aware that they are choosing differently. In contrast, conscious stochastic choice can arise from preference incompleteness (Aumann, 1962), preference imprecision (MacCrimmon & Smith, 1986; Dubourg et al., 1994; Butler & Loomes, 2007), or hedging of preference uncertainty (Fudenberg et al., 2015), where individuals are aware that they do not have a clear preference and are intentionally randomizing their choices.[1]

Models of stochastic choice have been extensively studied and applied in various fields. However, previous research has primarily focused on models of unconscious stochastic choices (Harless & Camerer, 1994; Hey & Orme, 1994; Rieskamp et al., 2006), while the study of conscious stochastic choice has been limited. One of the main challenges in investigating conscious stochastic choice is the difficulty of distinguishing between unconscious stochastic choices that are made based on a clear preference (as in random utility models) and conscious stochastic choices that individuals find difficult to make and randomize intentionally.

This paper presents a novel method to address the challenge of distinguishing between conscious and unconscious stochastic choice, and demonstrates its usefulness in one of the most studied anomalies in economics and management: preference reversals (PR). A typical PR experiment involves a pair of lotteries with similar expected payoffs: a P-bet with a more attractive winning probability and a $-bet with a more attractive payoff. PR occurs when individuals choose one bet over the other in the direct choice between the two bets, but place a higher value on the other in the indirect valuation of each bet separately. PR poses a fundamental challenge to the theory of revealed preferences and has important policy implications because it questions whether preferences can be reliably revealed from choices. While alternative explanations for PR exist, such as regret aversion (Loomes & Sugden, 1983; Loomes et al., 1989, 1991), flaws in some particular experimental designs such as the Becker-DeGroot-Marschak (Becker et al., 1964) mechanism (Holt, 1986; Karni & Safra, 1987; Segal, 1988), or scale compatibility (Tversky et al., 1990), there has been a persistent interest in explaining PR by stochastic choice. The stochastic

[1] Stochastic choice can also be explained by deterministic non-expected utility theory (Machina, 1985), such as Rank-dependent expected utility theory (Quiggin, 1982), in which individuals choose randomly not because they are unsure about their preferences, but because they violate the independence axiom (more precisely, the betweenness axiom, Chew, 1989) and have a strict preference for randomization. It is important to recognize, however, that the violation of betweenness may be due to imprecise preferences (Butler & Loomes, 2011).

choice explanation is attractive because it can account for PR while retaining procedurally invariant preferences, a fundamental postulate of many theories.

Our novel method involves incentivizing individuals to provide an interval of values that they consider comparable to each bet. In each valuation choice between the bet and a varying reference option, subjects could either pay a small cost to select a specific option or opt for a free randomization choice, where the selection is made randomly by the computer. We demonstrate that this method, together with repeating the direct choice between the P-bet and the $-bet nine times (Loomes & Pogrebna, 2017), allows us to consider the three major models of conscious stochastic choice (incomplete preference, preference imprecision, and hedging of preference uncertainty). Our experimental results reveal substantial conscious stochasticity in choices, and further suggest that subjects' stochastic choice differs across the elicitation procedures (direct choice vs. indirect valuation).

Specifically, our experiment used the same P-bet and $-bet as in Butler and Loomes (2007), except that the currency was Chinese Yuan (¥) instead of Australian dollar. The experiment consisted of mainly two parts. Part I was a standard PR experiment. Subjects chose directly between the P-bet and the $-bet, which was later repeated eight more times (separated by other choices). They also evaluated the two bets in terms of certainty equivalent (CE) and probability equivalent (PE) via a sequence of binary valuation choices between a fixed bet (either the P-bet or the $-bet) and a varying reference option (a sure payment when eliciting CE and a reference lottery when eliciting PE). In Part II, we elicited subjects' intervals of possible CE and PE for each bet through a series of valuation choices. In each of them subjects faced a fixed bet and a varying reference option. They could 1) pay a small cost (¥0.10) and choose the fixed bet, 2) pay a small cost (¥0.10) and choose the reference option, or 3) choose the randomization option for free, according to which a computer would randomly select one of the two options.

We demonstrate that the intervals of possible values (CE or PE) are captured by the randomization ranges (referred to as R-ranges hereafter): The ranges of sure payments (or the ranges of winning probabilities in the reference lottery) in which subjects choose the randomization option. We further show how our experiment allows us to estimate one stochastic function from the nine-repeated direct choices between the two bets and a separate one from the randomization ranges. This allows us to examine the prevalence of conscious stochastic choice among subjects, and check whether stochastic choice is sensitive to the different elicitation procedures (direct choices vs. indirect valuations). Finally, we evaluate quantitatively the explanatory power of stochastic choice for PR by noting that, when subjects' direct choices and indirect valuations are potentially stochastic, an experimental outcome would be a random realization of the stochastic process. We use the estimated stochastic functions to perform one million simulations, as if having these subjects participating in the experiment one million times, and compare the simulated patterns with the actual patterns in the experiment. If PR arises from conscious stochastic choice alone, the actual choice patterns in the experiment should not differ significantly from the simulated patterns.

Our experimental results suggest prevalent stochastic choice among subjects and substantial PR comparable to previous studies. However, the analysis of our

experimental results also clearly shows that the estimated stochastic functions depend on elicitation procedures (the direct choices or the valuation choices), the type of bets (the P-bet or the $-bet), and the type of valuation (certainty equivalents or probability equivalents), inconsistent with theoretical predictions. Moreover, the simulated choice patterns based on one consistent stochastic function for both the direct choice and the indirect valuations fail to capture the magnitude of PR and the asymmetry between the standard PR and the non-standard PR in the experiment. In contrast, when using R-ranges to estimate the probability of valuing the P-bet higher than the $-bet and using the nine-repeated direct choices to estimate separately the probability of choosing the P-bet, the simulated proportions capture the magnitude and the asymmetry of PR well. We conclude that subjects' choices were not only stochastic but also procedure dependent, and that both are important in explaining PR.

Our study contributes to the literature in two ways. First, our novel method allows us to expand the experimental investigation of stochastic choice by also considering models of conscious stochastic choice, which have received limited attention compared to studies on unconscious stochastic choice, such as random utility models. In a pioneering study by Agranov and Ortoleva (2017), choices were repeated three times in a row and subjects were explicitly informed of the repetitions in advance. They found that 71% of their subjects selected different options, and attributed the observed pattern to conscious stochastic choice instead of random expected utility (Gul & Pesendorfer, 2006) or drift diffusion (Ratcliff, 1978). Similar to our design, Cettolin and Riedl (2019) offered subjects a randomization option in addition to standard binary choices to identify incomplete preferences. Their method, while suitable for their research purpose, may have underestimated the extent of conscious stochastic choice. This is because, without the choosing cost like ours, subjects could flip a virtual coin in their head to randomize instead of using the free external randomization option. In contrast, our design incentivizes subjects to use the external randomization option to fully reveal the extent of conscious stochastic choice. Close to our design, Bouacida (2021) elicited the set of "choosable" alternatives by offering subjects reward for keeping additional options and then randomly selecting one for payment. His focus is to identify choice correspondences. Our study is also related to the measurement of preference imprecision—a range of values for which subjects report unsureness about their preferences (Dubourg et al., 1994, 1997; Butler & Loomes, 2007, 2011; Cubitt et al., 2015) or the strength of preferences: "The relative degree of difference between the two options as perceived by the decision maker" (Butler et al., 2014b, p. 538).[2] Those studies relied on *unincentivized* self-reports (Bayrak & Hey, 2020, for a survey of the recent literature); in contrast, our method is incentivized. Furthermore, we contribute to the understanding of the deliberate stochastic model by Fudenberg et al. (2015). The investigation of this model is not straightforward because the cost function, a crucial element in their model, is not directly observable. We show that eliciting CE or PE randomization ranges offers a convenient way to identify the cost function. Our result shows that the cost parameter of the P-bet is significantly higher than the one of the $-bet, consistent with the interpretation that

[2] Alós-Ferrer et al. (2016, 2021) related stochastic choice to the strength of preferences and ingeniously infer stochasticity in choices from decision time.

subjects perceive significantly higher preference uncertainty for the $-bet. This result supports the hypothesis that the cost parameter depends on the elicitation environment, as in the two extensions (item-invariant and menu-invariant) of their working paper version (Fudenberg et al., 2014).

Second, we link preference reversals with a broader class of theories of stochastic choice. Building on the decision field theory (Busemeyer & Townsend, 1993), Johnson and Busemeyer (2005) proposed that valuations and choices are based on the same stochastic process, and the higher variance about the value of the $-bet than that of the P-bet drives PR. Eliaz and Ok (2006) proposed that PR may arise from preference incompleteness. Blavatskyy (2009) presented a particular stochastic specification of random utility models, later axiomatized in Blavatskyy (2012), and showed that this model can account for classical PR. These studies are theoretical. Butler and Loomes (2007) assumed that subjects may have imprecise understanding of their preferences, and showed experimentally that the systematic difference in the imprecision of preferences between the two bets and stochastic choice could explain PR. Collins and James (2015) elicited the probability equivalents for nine endogenously constructed pairs of lotteries using the dual-to-selling version of the Becker–DeGroot–Marschak mechanism (Becker et al., 1964; James, 2011), in addition to the original design of Lichtenstein and Slovic (1971). They found that subjects responded differently to probability equivalents than to certainty equivalents, and that reversals based on probability equivalents were less frequent and could be explained by Blavatskyy (2009). Our research is also related to two recent studies of Loomes and Pogrebna (2017) and Alós-Ferrer et al. (2020). Loomes and Pogrebna (2017) provided evidence that preference reversal arose from distinct cognitive processes in choices and valuations, inconsistent with Johnson and Busemeyer (2005). Their study deviated from standard PR studies in two significant ways: Mixing valuation choices for different bets and repeating them four times each, and defining the valuation of bets as the stochastic indifference (SI) point where the bet and the sure amount are chosen equally likely, which is the natural stochastic analogue of certainty equivalence in deterministic theories. Alós-Ferrer et al. (2020) demonstrated theoretically and experimentally that PR could arise from not just behavioral biases but also risk attitudes and experimental design factors like the parameterization of two bets. Our study complements these studies by assessing and differentiating also models of conscious stochastic choice for PR. Working with the estimated stochastic functions and the simulations based on them allows us to extend previous analyses to evaluate qualitatively and quantitatively the potential of the stochastic choice *alone* for PR.

The paper proceeds as follows. Section 2 presents the experimental design. Section 3 provides a theoretical analysis based on the design and each theory of conscious stochastic choice. Results are reported in Section 4. Section 5 concludes.

2 Experimental design

In the experiment, we use the P-bet and the $-bet as in Butler and Loomes (2007), except that the currency is Chinese Yuan instead of Australian dollar. Specifically, the P-bet gives ¥24 with a 70% probability and ¥0 otherwise; the $-bet offers

a 25% chance of earning ¥80 and ¥0 otherwise. The experiment mainly consists of two parts, plus some auxiliary tasks.

2.1 Part I: Standard PR

Similar to most PR experiments, subjects are confronted with a direct choice between the P-bet and the $-bet. The P-bet and the $-bet are presented in pie charts. Following Butler and Loomes (2007), we also ask subjects for their decision confidence after each choice. Subjects can state their confidence on three levels: "*Sure*", "*Somewhat sure*", and "*Unsure*". Figure 1 shows an example of the computer screen that subjects face in the direct choice. We repeat the direct choice nine times, with each repetition separated by other choices, and counter-balance the screen position (left or right) of the P-bet and the $-bet for each subject. This repetition allows us to observe whether subjects' choices between the P-bet and the $-bet are stochastic, and if so, to estimate the stochastic function from the nine repeated direct choices.

To make choice and valuation tasks as similar as possible, we elicit subjects' CE and PE of each bet with binary choices. We present subjects with a sequence of binary choices between a fixed bet and a varying reference option (a sure payment in CE valuation, and a reference lottery of paying ¥100 with probability p and ¥0 otherwise in PE valuation). Similar to the direct choice, subjects indicate

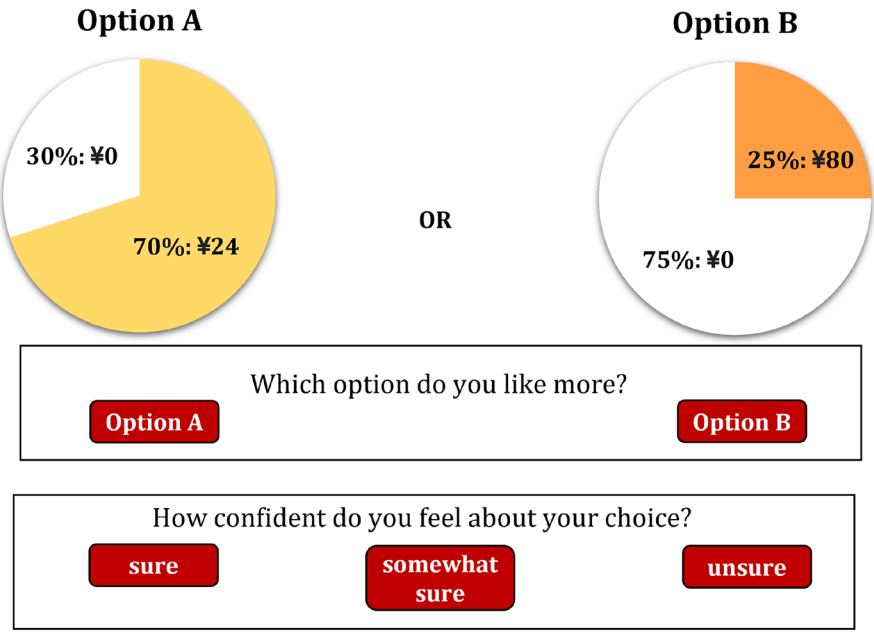

Fig. 1 The translated decision screen for the direct choice between the P-bet and the $-bet. Subjects first indicated their preferred bet, and then reported their confidence about their choice.

their preferred option and the confidence about their decision. Across choices the bet is fixed while the reference option varies according to a bisection process: Each time, the reference option is the mid-point of a certain interval. The original interval lies between 0 and the winning amount of money (or the winning probability) in the fixed bet, and in each succession the computer bisects the interval by using its mid-point as the new lower (upper) limit when subjects choose the fixed bet (or the reference option, respectively). Figure 2 illustrates the sequence of sure payment values used when eliciting CE for the P-bet. For example, suppose a subject chooses the sure payment in the first three choices and the bet in the last choice, her CE lies between 1.5 and 3, and we use the mid-point $2.25 = (1.5 + 3)/2$ as her CE. The bisection process allows us to find a relatively accurate CE or PE with a limited number of choices. For the elicitation of the CE, we implement the bisection process six times for the $-bet and four times for the P-bet, due to the larger payoff range of the $-bet than the P-bet (80 vs. 24). For the elicitation of the PE, we implement the bisection process four times for the $-bet and five times for the P-bet, due to the larger winning probability in the P-bet than in the $-bet (0.70 vs. 0.25).

While the bisection method has the advantage of avoiding biases in top-down or bottom-up methods such as choice lists (Békésy, 1947), it is well known that the bisection method is not fully incentive compatible. If subjects knew the complete procedure, they would have an incentive to misreport their preferences in the initial choices so that they face pairs of options with higher payoffs in later choices. However, as pointed in Bleichrodt et al. (2019, pp. 253): Manipulation is, "A theoretical possibility but it is practically impossible" (Bostic et al., 1990; Abdellaoui, 2000; Abdellaoui et al., 2008; van de Kuilen & Wakker, 2011; Qiu & Steiger, 2011, for more discussion and evidence). Nevertheless, to mitigate this concern, we included some extra choices at the beginning of each bisection

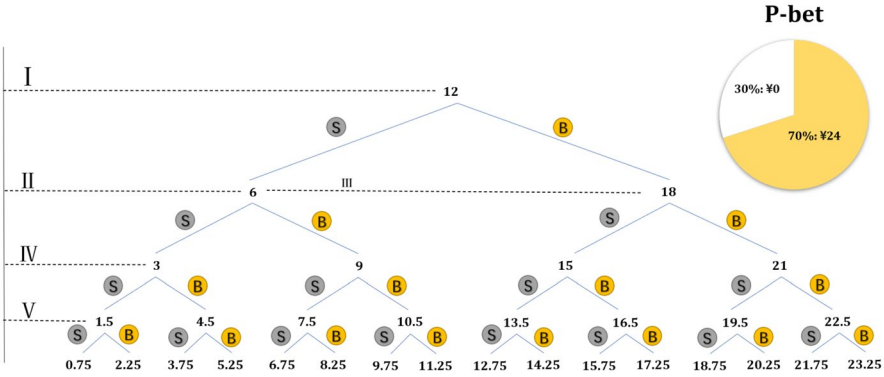

Fig. 2 The bisection process for eliciting CE of the P-bet. S stands for choosing the sure payment, B stands for choosing the bet. The numbers at the top four rows (I, II, III, IV) are the possible sure payments subjects faced in four binary choices. The numbers at the bottom are the elicited CE for the corresponding choices.

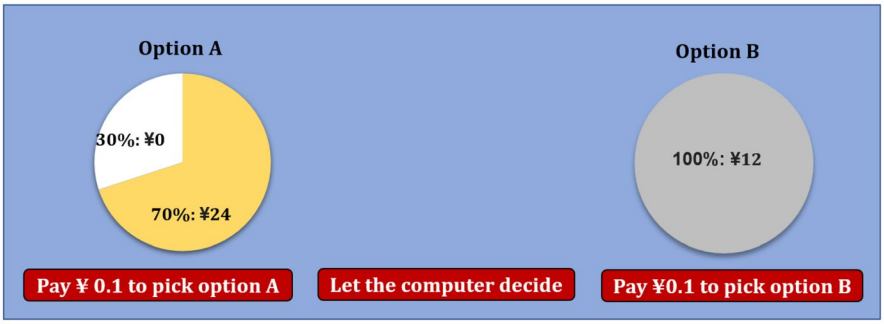

Fig. 3 The translated decision screen of eliciting the CE R-range of the P-bet.

process to make it less obvious. See Appendix "The bisection process" for more details. We will also specifically examine this potential in the data analysis.

2.2 Part II: Monetary and probability randomization ranges

In Part II of the experiment, we elicit monetary and probability randomization ranges (CE R-range and PE R-range) for the P-bet and the $-bet. The procedure of eliciting these ranges is similar to the aforementioned CE or PE valuation. The difference is that in each choice subjects have three instead of two options: They can 1) pay a small cost (¥0.10) and choose the fixed bet, 2) pay a small cost (¥0.10) and choose the reference option, or 3) choose a randomization option for free, according to which a computer randomly selects one of the two options. Figure 3 illustrates a decision screen (translated to English) to elicit the CE R-range for the P-bet. We use two bisection processes, one for the lower bound and another one for the upper bound, to determine the R-ranges. For more experimental details, please see Appendix "The elicitation of the randomization range (R-range)". In Section 3 we demonstrate the theoretical implications of the choice cost.

2.3 Additions: Risk and ambiguity attitudes

We include two additional tasks to separate the two main tasks in the experiment from each other. One task elicits risk attitudes, and we will use this to estimate subjects' utility function independent of the PR experiment in one of the analyses. Attitudes toward utility uncertainty play a key role in Fudenberg et al. (2015), but there is no established method to elicit them. Therefore, we use the attitude toward belief uncertainty as an exploratory substitute. Details about these two tasks can be found in Appendix "Risk and ambiguity attitudes".

2.4 Experimental procedure

Before the experiment began, the experimenter first read the experimental instructions aloud and asked the subjects if they had any questions about the tasks. To ensure that the subjects understood the experiment, we presented them with three incentivized control questions (¥2 for each correctly answered control question) before the real experimental tasks. As mentioned earlier, the experiment consisted of two main parts. To control for potential order effects, we counterbalanced the order of the different parts. Table 1 summarizes the experimental order. Finally, the subjects completed a questionnaire at the end of the experiment. Apart from standard demographic information such as gender, age, and field of study, the questionnaire also asked the subjects whether they ever chose the randomization option and, if so, why. Additionally, we administered a measure of control desirability, which elicits how subjects perceive the importance or benefits of maintaining control (Burger & Cooper, 1979; Gebhardt & Brosschot, 2002). See Appendix "Measuring the desirability of control" for more details. Subjects received a show-up fee of ¥10. Additionally, after the experiment, the computer randomly selected one choice from all the choices that the subjects made during the experiment and played it for real.

Despite COVID restrictions, we were able to conduct a laboratory experiment at the Lab of Experimental Economics and School of Finance, Dongbei University of Finance and Economics (DUFE) from May to July in 2021. We recruited 175 undergraduate and graduate students at DUFE through the laboratory's WeChat recruiting system. Of the subjects, 33.71% were male and 66.29% were female. There were eight experimental sessions, and the average duration of each session was about one hour. The average payment was approximately ¥37.82, which was about 6 US dollars. This payment is higher than the average students' hourly wage in China.

3 Theoretical analysis

In this section, we provide the theoretical analysis of our experiment. We mainly consider models of incomplete preferences (Eliaz & Ok, 2006) or imprecise preferences (Butler & Loomes, 2007, 2011; Cubitt et al., 2015), and deliberately stochastic preferences (Fudenberg et al., 2015).

Table 1 The four orders in the experiment. Subjects were randomly assigned into one of the four orders.

Order	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5
1	PR	Risk	CE R-range	Ambiguity	PE R-range
2	CE R-range	Risk	PE R-range	Ambiguity	PR
3	PR	Ambiguity	CE R-range	Risk	PE R-range
4	CE R-range	Ambiguity	PE R-range	Risk	PR

3.1 Models of incomplete/imprecise preferences

The behavioral predictions of imprecise preferences models and incomplete preferences models are qualitatively the same in our experiment. Below we present the analysis based only on the incomplete preference model of Eliaz and Ok (2006), which explicitly considers incomplete preferences as a potential to explain PR. In this model, there exist pairs of options that individuals are unable to compare (neither strict preference nor indifference). To incorporate incomplete preferences into revealed choices, Eliaz and Ok (2006) relax the standard Weak Axiom of Revealed Preferences (WARP) and consider instead the Weak Axiom of Revealed Non-Inferiority (WARNI). In WARP individuals choose a weakly preferred option, while in WARNI individuals may choose an option that is not clearly inferior to other options (e.g., according to the first-order stochastic dominance criterion), including options that individuals cannot compare. When there are multiple such options, subjects would state, "I don't know/care, I'd flip a coin, I guess" (Eliaz & Ok, 2006, p. 65). Preference incompleteness can arise from uncertainty in beliefs or preferences (Ok et al., 2012). Our study considers only lotteries with objective probabilities, and thus we focus on uncertainty in preferences: Individuals are uncertain about the true utility of a lottery and consider a range of valuations possible. Dubra et al. (2004) model this as individuals having a set of utility functions and ranking options only when all utility functions produce the same ordering.

We now demonstrate how R-ranges reveal preference incompleteness or imprecision. Intuitively, subjects are willing to pay a cost to select a particular option only when they strictly prefer that option over the other, and they should choose the free randomization option when they are unsure about their preference. Concretely, when subjects cannot compare the bet and the reference option or happen to be indifferent between them, in the spirit of Eliaz and Ok (2006), both options are choosable and subjects randomize. There are two ways for subjects to randomize: They can randomize subjectively in their head (e.g., by flipping a virtual coin), or they can choose the explicit randomization option. However, due to the cost of choosing, the explicit option is preferred over the subjective one since the payoff of the explicit randomization (receiving either the bet or the sure payment/the reference lottery) dominates that of the subjective randomization (receiving either the payoff for the bet – ¥0.10 or the sure payment/the reference lottery – ¥0.10). Thus, unlike previous studies such as Cettolin and Riedl (2019), our novel method incentivizes subjects to use the external randomization option to fully reveal preference incompleteness/imprecision. We further note that it is highly unlikely that subjects happen to be indifferent between the bet and one of the reference options in the experiment. Even if we allow for indifference, we can distinguish unsure preference from indifference by recognizing that, since the cost of choosing is small, indifference can occur for at most one value of the reference options (Cettolin & Riedl, 2019). Additionally, we expect that the elicited CE (PE) of a bet lies within the CE R-range (PE R-range). The idea is straightforward: If a subject is willing to pay a cost to pick the bet or the sure payment in the CE R-range task (the reference lottery in the PE R-range task), they should continue doing so when they can freely pick that option in the valuation tasks.

Using the R-ranges of the two bets, we can determine the probability of subjects valuing the P-bet more than the $-bet by assuming that the value of a bet is a random realization within the R-range. According to Eliaz and Ok (2006), when one bet's lower bound is higher than the upper bound of the other bet, subjects have a clear preference and will choose the bet with the higher valuation. When the R-ranges of the P-bet and the $-bet overlap, subjects should choose randomly because their preference is incomplete. Unfortunately, theory does not offer clear guidance on how subjects should randomize when they cannot compare. Therefore, in the analysis of the experimental results, we will consider a few alternative rules of randomization. By using the probabilities of valuing the P-bet higher than the $-bet and the probability of choosing the P-bet in the direct choice from the nine-repeated choices, we can calculate the probability of exhibiting each of the four PR patterns for each subject.

We then run one million simulations. The basic idea is that, since subjects' choices and valuations are stochastic, each experimental outcome can be treated as a random realization of a stochastic process. In each simulation we generate a random number between 0 and 1 for each subject. The randomly generated number and the calculated probabilities together determine the PR choice pattern that this subject would exhibit. For example, suppose a subject exhibits the consistent P-bet choice pattern with probability 20%, PR with 50%, non-standard PR with 10%, and the consistent $-bet choice pattern with 20%. Then the subject is said to exhibit the consistent P-bet pattern if the randomly generated number is below 0.2; PR if it is between 0.2 and 0.7; non-standard PR if it is between 0.7 and 0.8; and the consistent $-bet pattern if it is between 0.8 and 1.0. Aggregating across subjects, we obtain one possible realization of the experiment, e.g., k% of subjects exhibit the standard PR, j% the non-standard PR and so on. By running one million simulations, it is as if these subjects participated in the experiment one million times, and we obtain a distribution of proportions for each PR pattern, e.g., the proportions of exhibiting the standard PR across simulations range from a% to b%, the proportions of exhibiting the non-standard PR range from x% to y%, and so on. If PR arises from preference incompleteness/imprecision alone, we should expect the simulated proportions to match the actual proportions. We summarize the predictions of models of incomplete/imprecise preferences below.

Prediction 1 In models of incomplete/imprecise preferences: I) subjects may choose the randomization option frequently, and II) the elicited CE or PE must lie within the R-range. Subjects choose differently in the nine-repeated direct choices only when the R-ranges of the two bets overlap; III) the simulated proportions of the PR patterns are comparable to the actual proportions.

3.2 Hedging of preference uncertainty

In models of incomplete/imprecise preferences, individuals choose their preferred option when they have a clear preference and randomize by choosing equally likely options that are not dominated when they have incomplete/imprecise preferences.

This is a strong assumption and may be too restrictive. It is possible that individuals are not fully confident about their preferences between two options but may be more likely to choose one option over another. Below we perform the analysis based on the deliberately stochastic choice model of Fudenberg et al. (2015) to capture this idea. In this model individuals have uncertainty about their true preferences, and they randomize deliberately to hedge this preference uncertainty.[3] When facing two options A and B, individuals behaving according to Fudenberg et al.'s (2015) functional form choose the optimal randomization probability as

$$p^* = \text{argmax}_p \quad pU(A) + (1-p)U(B) - c(p) - c(1-p),$$

where $U(\cdot)$ is the Von Neumann-Morgenstern expected utility, and $c(p)$ is a convex and continuously differentiable cost function. The decision utility of choosing the lottery X for sure is $U(X) - c(1)$, and the decision utility of choosing the reference option y for sure is $U(y) - c(1)$. The first order condition is simply $c'(p) - c'(1-p) = U(A) - U(B)$, which predicts that the randomization probability p is close to 0.5 when $U(A) - U(B)$ is small. Fudenberg et al. (2015) show that their representation corresponds to a form of an uncertainty averse individual who is unsure about her true utility and uses randomization to hedge her preference uncertainty. We assume a cost function of $c(p) = plog(p)/\eta$, which leads to the popular logit choice function: $p = \frac{e^{\eta U(A)}}{e^{\eta U(A)} + e^{\eta U(B)}}$. The sensitivity parameter η can be related to preference uncertainty, with a larger η corresponding to lower preference uncertainty.

We now show that we can infer η, the critical parameter in Fudenberg et al. (2015), from the R-ranges. Let X denote the bet and y the sure payment. When the cost is sufficiently small, as in our experiment, we can ignore it and define the smallest ($\underline{y}$) and largest ($\overline{y}$) values of y such that an individual prefers the randomization option as

$$\begin{aligned} u(\underline{y}) + 0.5[U(X) - u(\underline{y})] - 2c(0.5) &\geq U(X) - c(1.0) - c(0) \\ &\Rightarrow u(\underline{y}) = U(X) - 2[c(1.0) + c(0) - 2c(0.5)], \\ u(\overline{y}) + 0.5[U(X) - u(\overline{y})] - 2c(0.5) &\geq u(\overline{y}) - c(1.0) - c(0), \\ &\Rightarrow u(\overline{y}) = U(X) + 2[c(1.0) + c(0) - 2c(0.5)]. \end{aligned}$$

It follows that $U(X) = [u(\overline{y}) + u(\underline{y})]/2$. Since $c(0)$ or $c(1)$ approaches to infinity when the cost function is $c(p) = plog(p)/\eta$, we approximate these values by $c(0.01)$ and $c(0.99)$, and accordingly $\eta = \frac{2.55}{u(\overline{y}) - u(\underline{y})}$. Similarly, we can also infer η from the PE R-range and obtain $\eta = \frac{2.55}{(\overline{p} - \underline{p})u(100)}$. We assume a parametric utility function $u(x) = \frac{x^\gamma}{\gamma}$ ($u(x) = log(x)$ when $\gamma = 0$), and estimate the parameter γ from the separate risk

[3] Cerreia-Vioglio et al. (2019) is also an important model in this direction, where they consider a class of cautious expected utility preferences (Cerreia-Vioglio et al., 2015). While it could generate preferences for randomization over non-degenerate lotteries, by the axiom of Weak Stochastic Certainty Effect, individuals do not randomize when one of the two options is a sure payment.

attitude elicitation tasks.[4] If subjects behave according to the deliberately stochastic model of Fudenberg et al. (2015), we expect the estimated η from the direct choices to be the same as those from valuations. With the estimated stochastic function for each subject, we can easily calculate the probability of choosing the P-bet and the probability of valuing the P-bet higher than the $-bet. We can then perform the same analysis as in Subsection 3.1, and we should expect the simulation proportions of PR patterns to match the actual proportions.

The above analysis allows for the possibility that the elicited CE or PE is outside the R-range because the optimal randomization probability p^* can be positive even when $y < \underline{y}$ and p^* can be less than 1 even when $y > \overline{y}$. We can derive a further hypothesis. In light of Fudenberg et al. (2015), the cost function may relate to a measure of control desirability because, as Fudenberg et al. (2015) point out, a preference for randomization arises because the individual trades off the probability of errors against the cost of making the desired choice. An individual who prefers to maintain control may perceive the implementation cost to be small and is less willing to randomize. In addition, we explore the relationship between the cost function and the attitude toward ambiguity, bearing in mind that this attitude may differ from attitude toward preference uncertainty. We summarize the predictions below.

Prediction 2 In light of Fudenberg et al. (2015): I) subjects' R-ranges relate positively to uncertainty aversion and negatively to the desirability of control, II) the estimated η does not differ between the stochastic function estimated from direct choices or from valuations, and III) the simulated proportions should be comparable to the actual proportions in the experiment.

4 Experimental results and analysis

In Appendix 1 we demonstrate that our experiment replicates the results in Butler and Loomes (2007), and the bisection method did not cause any significant distortion in subjects' behavior. Moving on to our main findings, we first present the prevalence of conscious stochastic choice among subjects. The following result summarizes the finding.

Result 1 The majority of subjects exhibited conscious stochastic choice: They chose the randomization option more than twice in at least one task, and their R-ranges related to self-reported decision confidence systematically.

Support: Table 2 presents the proportions of subjects who chose the randomization option zero times, once, and twice or more. On aggregate, a significant

[4] This is possible despite stochastic choice in Fudenberg et al. (2015) because, in binary choices without the randomization option, $u(\underline{y}) < U(X) < u(\overline{y})$, where $\underline{y}$ and $\overline{y}$ are the switching values in the multiple price list. We can then follow the standard practice and use the midpoint of the switching values as the certainty equivalent of the lottery.

Table 2 The proportions of subjects who chose the randomization option zero times, once, and twice or more. The aggregate numbers across the four tasks (CE_P, $CE_\$$, PE_P, and $PE_\$$) are the proportions of subjects who never chose the randomization option, chose the randomization at most once in any of the four tasks, and twice or more in at least one of the four tasks. CE aggregate (PE aggregate) considers only the CE R-range task (PE R-range task, respectively).

	0 times	1 time	≥ 2 times
Aggregate	39 (22.3%)	23 (13.1%)	113 (64.6%)
CE_P	90 (51.4%)	38 (21.7%)	47 (26.9%)
$CE_\$$	58 (33.1%)	33 (18.9%)	84 (48.0%)
CE aggregate	50 (28.6%)	29 (16.6%)	96 (54.9%)
PE_P	75 (42.9%)	31 (17.7%)	69 (39.4%)
$PE_\$$	76 (43.4%)	23 (13.1%)	76 (43.4%)
PE aggregate	61 (34.9%)	23 (13.1%)	91 (52.0%)

proportion of subjects (77.7%) chose the randomization option at least once during the experiment. As it is unlikely that subjects would happen to be indifferent between the bet and one of the presented reference options, this suggests that many subjects may not always have clear preferences. Even when we narrow our focus to subjects who chose the randomization option twice or more in at least one of the four tasks, this proportion is still substantial (64.6%). The proportions of subjects who chose the randomization option twice or more are similar across the CE R-range and the PE R-range tasks, with 54.9% and 52.0%, respectively.

R-ranges differ significantly between bets and across the type of valuation. The mean size of the CE R-range for the P-bet is 1.35, which is about 8% of the expected value of the P-bet. It is 4.74 for the $-bet, which is about 24% of the expected value of the $-bet. Thus, the range of CE values that subjects do not have a clear preference is significantly larger for the $-bet than for the P-bet, consistent with the finding in Butler and Loomes (2007). Similar results hold for the PE R-range, with the mean size of 0.06 for the P-bet and 0.04 for the $-bet, corresponding to the larger winning probability of the P-bet (70%) than the $-bet (25%). These differences between the two bets are statistically significant (Wilcoxon signed rank tests, $p < 0.01$ for both CE and PE R-ranges). The histogram in Fig. 7 illustrates the distribution of the size of the CE and PE R-range for both bets.

We further find that the R-range relates systematically to the interval of less than full confidence ("Somewhat sure" or "Unsure" according to self-reported confidence statement). The average width of the CE interval of less than full decision confidence is 1.84 for the P-bet and 6.23 for the $-bet. A series of Spearman rank correlation tests suggest that the size of the CE R-range is positively correlated with the interval of limited confidence ($\rho = 0.30$ for the P-bet, $\rho = 0.32$ for the $-bet, and $p < 0.01$ for both tests). In addition, the two bounds of the CE R-range positively and significantly correlate with the two bounds of the interval of limited confidence (the lower bound: $\rho = 0.53$ for the P-bet and $\rho = 0.37$ for the $-bet. The upper bound: $\rho = 0.63$ for the P-bet, $\rho = 0.43$ for the $-bet, and $p < 0.01$ for all tests). Similar results hold when the valuation is based on PE. The average size of the interval of

limited decision confidence is 0.09 for the P-bet and 0.03 for the $-bet, which is positively correlated with the size of the PE R-range (Spearman rank correlation $\rho = 0.32$ for the P-bet, $\rho = 0.26$ for the $-bet, and $p < 0.01$ for both tests). Moreover, the two bounds of the PE R-range positively and significantly correlate with the two bounds of the interval of limited confidence (the lower bound: $\rho = 0.57$ for the P-bet, $\rho = 0.49$ for the $-bet, and $p < 0.01$ for both tests. The upper bound: $\rho = 0.70$ for the P-bet and $p < 0.01$, $\rho = 0.23$ for the $-bet, and $p > 0.10$). These results are consistent with the interpretation that the CE R-ranges and the PE R-ranges capture the range in which subjects do not have clear preferences, and they chose the randomization option consciously.

We proceed to examine the explanatory power of conscious stochastic choice for PR. Under each conscious stochastic choice theory, we perform two analyses:

Analysis 1 We test whether the stochastic functions from different elicitation procedures differ significantly from each other by estimating the stochastic function separately from the nine-repeated direct choices and the CE (PE) R-ranges.

Analysis 2 We evaluate whether it is possible to use one consistent stochastic function for both the direct choice and the CE (or PE) valuation to generate the PR pattern in the experiment. We also check whether allowing for two stochastic functions—one for the direct choice and a different one for the CE valuation—brings simulated proportions closer to the actual proportions.

4.1 Models of incomplete/imprecise preferences

There is some potential for the role of preference incompleteness/imprecision for PR. As we can see from Table 3, subjects with overlapping R-ranges were more likely to exhibit the standard PR than those without overlapping R-ranges (50.0% vs. 39.5%) when the valuation is based on CE, although the difference

Table 3 The proportions of each PR pattern for two groups of subjects based on the incomplete/imprecise preferences. The numbers in parentheses are the number of subjects in each group. Group 1 includes the subjects who had positive (CE or PE) R-ranges and the R-range for the P-bet overlapped with that of the $-bet. Group 2 includes other subjects: those who had a positive (CE or PE) R-range for at least one of the bets but their R-ranges did not overlap, or those who had no positive (CE or PE) R-ranges.

The CE PR patterns				
	Consis. P	Std. PR	Nonstd. PR	Consis. $
G1 (56)	16.1%	50.0%	0.0%	33.9%
G2 (119)	12.6%	39.5%	6.7%	41.2%
The PE PR patterns				
	Consis. P	Std. PR	Nonstd. PR	Consis. $
G1 (34)	35.3%	14.7%	23.5%	26.5%
G2 (141)	50.4%	7.8%	19.1%	22.7%

is not statistically significant ($p = 0.7252$ according to a two-sided proportional test). Similarly, when the valuation is based on PE, the subjects with overlapping R-ranges exhibit higher non-standard PR than the subjects without overlapping R-ranges (23.5% vs. 19.1%), but the difference is not statistically significant ($p = 0.738$ in a two-sided proportional test).

To evaluate more rigorously whether incomplete/imprecise preferences can account for PR patterns, we follow Eliaz and Ok (2006) and assume that subjects choose randomly when they cannot compare the two bets. This happens when their CE or PE R-range overlaps. They select the bet with the higher valuation otherwise.

In the main text we focus on the PR patterns when the valuation is based on CE (referred to as CE PR patterns hereafter), which are the focus of most previous studies. Results on the PR patterns when the valuation is based on PE (referred to as PE PR patterns) are in general similar to those of the CE PR patterns. For conciseness, we report them in Appendix "Additional results about models of incomplete/imprecise preferences". Here our Analysis 1 consists of three parts: (i) we compare subjects' CE with their CE R-range. Subjects' CE should fall in the CE R-range if their behavior arises from preference incompleteness/imprecision; (ii) we use subjects' CE R-ranges for the P-bet and the $-bet to predict their direct choices. Subjects who had no overlapping CE R-ranges should not choose differently in the nine-repeated choices according to models of incomplete/imprecise preferences, and (iii) we use subjects' direct choices to predict whether their CE R-ranges for the P-bet and the $-bet overlap. Subjects who chose differently in the nine-repeated choices must have overlapping CE R-ranges for the P-bet and the $-bet. The result below summarizes the findings:

Result 2 Inconsistent with models of incomplete/imprecise preferences, (i) the elicited CE of a non-negligible proportion of subjects (28.0%) is outside the CE R-range, (ii) among the subjects who had no overlapping CE R-ranges in the P-bet and the $-bet, 50.4% (60 out of 119) chose differently in the nine-repeated direct choices, and (iii) among the subjects who chose differently in the nine-repeated direct choices, 68.2% (60 out of 88) had no overlapping R-ranges in the P-bet and the $-bet.

Support: As we can see from the sub-table a) of Table 4, among the 96 subjects who have a positive CE R-range for at least one bet, 51.0% (49 out of 96) had CE outside the R-range for at least one of the two bets. Examining each bet separately, we see that this occurs more often with the $-bet than the P-bet (52.4% vs. 21.3%). As we explained in Subsection 3.1, when subjects behave according to preference incompleteness/imprecision, their CE should fall within their CE R-range. Thus, these subjects do not behave according to models of incomplete/imprecise preferences.

Sub-table b) of Table 4 reports the number of subjects who had overlapping CE R-ranges or non-overlapping CE R-ranges as well as who chose consistently or differently in the nine-repeated direct choices. As we can see, among subjects who had no overlapping CE R-ranges, 50.4% (60 out of 119) chose differently in the nine-repeated choices. Similarly, among subjects who chose differently in the

Table 4 Sub-table a) reports the numbers of subjects who had no R-range, who had positive R-ranges and with CE in the R-range, and who had positive R-range but whose CE was outside the R-range. The group "No R-range" includes subjects who had no positive R-range in any of the tasks. The group "In the R-range" includes subjects who had positive R-range in at least one task, and their elicited CE or PE were in the R-range. The group "Outside the R-range" includes subjects who had positive R-range in at least one task and their elicited CE or PE were outside the R-range in at least one task. Sub-table b) reports the number of subjects who had overlapping CE R-ranges or non-overlapping CE R-ranges as well as who chose consistently or differently in the nine-repeated direct choices.

Sub-table a)

	No R-range	In the R-range	Outside the R-range	
			Larger than the upper bound	Lower than the lower bound
CE_P	128 (73.1%)	37 (21.1%)	6 (3.4%)	4 (2.3%)
$CE_\$$	91 (52.0%)	40 (22.9%)	36 (20.6%)	8 (4.6%)
CE aggregate	79 (45.1%)	47 (26.9%)	49 (28.0%)	

Sub-table b)

	In the nine-repeated direct choices	
	Chose consistently	Chose differently
Overlapping CE R-ranges	28	28
Non-overlapping CE R-ranges	59	60

nine-repeated choices, 68.2% (60 out of 88) had no overlapping CE R-ranges for the P-bet and the $-bet. Relaxing the group "Chose consistently" by including also the subjects who chose the same bet eight out of nine times, interpreting the one different choice as a mistake, does not improve these proportions, as we can see in Table 12 in Appendix "Additional results about models of incomplete/imprecise preferences". Preference incompleteness/imprecision alone cannot explain the behavior of these subjects.

The above result suggests that a large proportion of subjects' behavior in direct choices and valuations cannot be explained by preference incompleteness or imprecision. To quantitatively evaluate the potential of preference incompleteness/imprecision in the explanation of CE PR patterns, we calculate the probability of valuing the P-bet higher than the $-bet by assuming subjects randomly drawing a value in the CE R-range. To calculate the choice probability, we assume subjects choose randomly (50%:50%) when their R-ranges for the P-bet and the $-bet overlap. As a robustness check, we additionally calculate the choice probability from CE R-ranges, assuming that subjects' choices are determined by randomly sampling the potential values of the bet in the CE R-ranges. By combining the valuation probability with the choice probability, we calculate the probability of each of the four choice patterns for each subject and perform one million simulations, as described in Subsection 3.1. This leads to:

Result 3 In light of preference incompleteness/imprecision, simulated proportions based on one consistent stochastic function overestimate consistent preferences and fail to capture the magnitude as well as the asymmetry of CE PR patterns. In contrast, when we estimate the probability of valuing the P-bet higher than the $-bet using CE R-ranges and separately estimate the probability of choosing the P-bet through nine-repeated direct choices, the simulated proportions provide a more accurate representation of the magnitude and asymmetry of PR.

Support: As we can see from M1-M3 in Fig. 4 (more details in Table 10 in Appendix "Additional results about models of incomplete/imprecise preferences"), simulated proportions based on one consistent stochastic function tend to overestimate the proportion of consistent preferences (ranging from 74.9% to 90.1%), which is significantly higher than the actual proportion of 52.5% (Wilcoxon signed rank tests, $p < 0.10$ for all comparisons). The estimated mean proportions of standard and non-standard PR ranges from 9.8% to 25.1%, which is significantly and substantially lower than the actual proportion of 47.5% (Wilcoxon signed rank

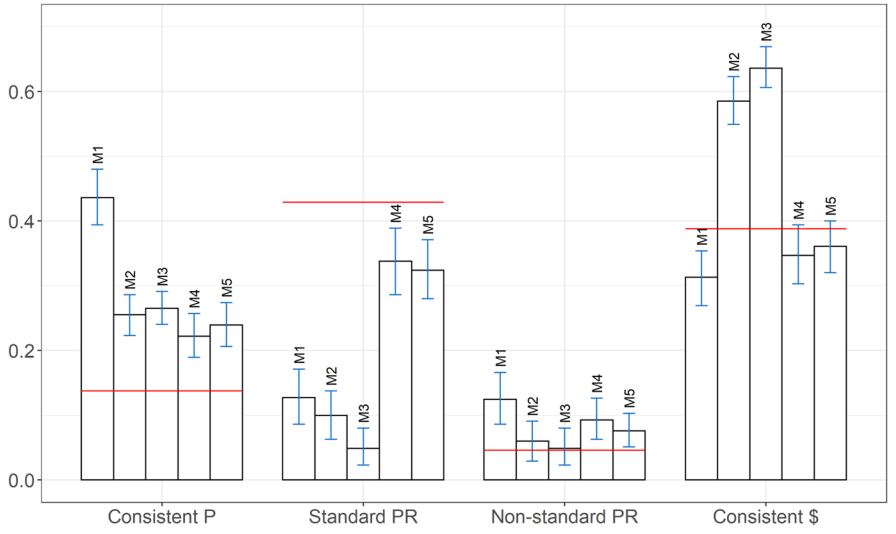

Fig. 4 The comparison of means and 95% confidence intervals of simulated proportions of CE PR patterns under incomplete/imprecise preferences models across different simulation methods. The error bars are the 95% confidence intervals of the simulated proportions. The horizontal red lines are the actual proportions in the experiment. In M1, the probability of valuing the P-bet higher than the $-bet is 50% if subjects chose different bets in the nine repeated direct choices and it is 1 (or 0) if subjects chose the P-bet (or the $-bet) for nine times. In M2, subjects choose randomly (50%:50%) when their R-ranges for the P-bet and the $-bet overlap. In M3, the choice probability is calculated from CE R-ranges by assuming that subjects' choices are determined by randomly sampling the potential values of the bet in the CE R-range. In M4 and M5, we calculate the probability of choosing the P-bet in the direct choice from the nine-repeated direct choices and use CE R-ranges to calculate the probability of valuing the P-bet higher than the $-bet. In M4, we assume random choice (50%:50%) if subjects chose differently in the nine-repeated direct choices. In M5, we use the actual choice ratio in the nine repeated direct choices as the probability of choosing the P-bet in the direct choice.

tests, p<0.10 for all comparisons). The simulated proportions underestimate the asymmetry between the standard PR and non-standard PR as well: The asymmetry ratio (standard PR/non-standard PR) ranges from 1 to 1.67 according to the simulated proportions, which is significantly and substantially lower than the actual ratio of 9.32 (Wilcoxon signed rank test, p<0.010).

When allowing for two separate stochastic functions, one estimated from R-ranges and another from the nine-repeated choices, the overall simulated proportions of the CE PR patterns now match some of the important properties of the actual proportions in the experiment. As we can see from M4 and M5 in Fig. 4, the simulated proportions estimate the proportion of PR (the standard and non-standard) to be 40.0% or 43.1%, which is close in magnitude to the actual proportion of 47.5%. Importantly, while still significantly different, the simulated proportions capture the asymmetry of the standard and non-standard PR: The mean asymmetry ratio of the standard PR divided by the non-standard PR is 3.63 and 4.26, although it is still statistically different from the actual ratio of 9.32. This result suggests further that choices are sensitive to the elicitation procedures, and we need separate stochastic functions for choices and valuations in order to account for PR.

4.2 Hedging of preference uncertainty (Fudenberg et al., 2015)

To examine whether hedging of preference uncertainty as in Fudenberg et al. (2015) can explain subjects' behavior, we first investigate the relationship between R-range and ambiguity attitude, as well as the desirability of control. Table 15 in Appendix "Additional results about hedging of preference uncertainty" Fudenberg et al. (2015) provides some supporting evidence, as R-ranges are positively associated with weaker desirability of control and ambiguity aversion, although the relationship is statistically significant only for the former ($p < 0.05$ for weaker desirability of control and $p > 0.10$ for ambiguity aversion).

As we show in the deliberately stochastic model of Fudenberg et al. (2015) in Subsection 3.2, individuals randomize when they are uncertain about their preferences over the two bets. In particular, subjects could have CE or PE values outside the R-ranges because these ranges do not capture the full extent of their preference uncertainty. To infer subjects' stochastic function for the CE valuations, we follow the theoretical analysis and make use of their R-ranges: $\eta_X = \frac{2.55}{u(\overline{y})-u(\underline{y})}$ when the valuation is based on CE, and $\eta_X = \frac{2.55}{(\overline{p}-\underline{p})u(100)}$ when the valuation is based on PE, where X is the P-bet or the \$-bet. To infer the stochastic function for the direct choice, we estimate the η by using the actual choice ratio in the nine-repeated direct choices and assuming the Logit stochastic function $P(P-bet \succ \$-bet) = \frac{e^{\eta U(P)}}{e^{\eta U(P)}+e^{\eta U(\$)}}$. To calculate $u(\overline{y})$, $u(\underline{y})$, $U(\$)$, and $U(P)$ we use the mean estimated risk parameter $\bar{\gamma}$ from the risk measurement task. Below we report the results about the CE PR pattern. The results about PE PR pattern, which are generally similar to those of CE PR pattern, are reported in Appendix "Additional results about hedging of preference uncertainty" Fudenberg et al. (2015). The following result summarizes the finding:

Result 4 The estimated η from CE R-ranges based on Fudenberg et al. (2015) differs significantly from the η estimated from the nine-repeated direct choices.

Support: Table 5 reports the mean and the median estimated η of the stochastic functions from the direct choices and CE R-ranges. As we can see, the mean estimated η from CE R-ranges is 10.85 for the P-bet and 10.96 for the \$-bet (medians are 6.45 and 4.61, respectively). In contrast, the mean estimated η from the nine-repeated direct choices is –0.24 (median of 0.28). Wilcoxon signed rank tests suggest that the difference in the estimated η from CE R-ranges and nine-repeated direct choices is statistically significant ($p < 0.01$ for both the P-bet and the \$-bet comparisons). The η estimated from the choice ratio of the nine-repeated direct choices is significantly lower, even with a negative mean. This suggests that the stochastic function for the direct choice does not respond sensitively (or even wrongly) to the expected utility difference of the two bets. The low values of η estimated from the direct choices thus suggest that subjects may behave differently in the direct choice and in the risk measurement tasks.

With the estimated stochastic functions from the CE R-ranges of the P-bet or the \$-bet as well as from the nine-repeated direct choices for each subject, following the analysis in Subsection 3.2, we use each of these stochastic functions to calculate the probability of choosing the P-bet in the direct choice and the probability of valuing the P-bet higher than the \$-bet.

Result 5 Under Fudenberg et al. (2015), simulated proportions based on one consistent stochastic function overestimate consistent preferences and fail to capture the magnitude as well as the asymmetry of CE PR patterns. In contrast, when using the CE R-ranges to estimate valuation probabilities and using nine-repeated direct choices to estimate the probability of choosing the P-bet, the simulated proportions capture the magnitude and the asymmetry of PR well.

Table 5 The fitted η in the deliberately stochastic models. Standard deviations are in parentheses. The η of the stochastic function from the direct choices is estimated from the equation $P(P-bet \succ \$-bet) = \frac{e^{\eta U(P)}}{e^{\eta U(P)}+e^{\eta U(\$)}}$. The η of the stochastic function from the CE R-ranges is estimated from the equation $\eta_X = \frac{2.55}{u(\overline{y})-u(\underline{y})}$, where X is the P-bet or the \$-bet. Wilcoxon signed rank tests check the difference in the estimated η from the direct choices and the CE R-ranges. * denotes $p < 0.10$, ** denotes $p < 0.05$, *** denotes $p < 0.01$.

	The estimated η		
	from the direct choices	from the CE R-ranges	
		P-bet	\$-bet
Mean	-0.24	10.85***	10.96***
	(3.80)	(9.51)	(15.65)
Median	0.28	6.45	4.61

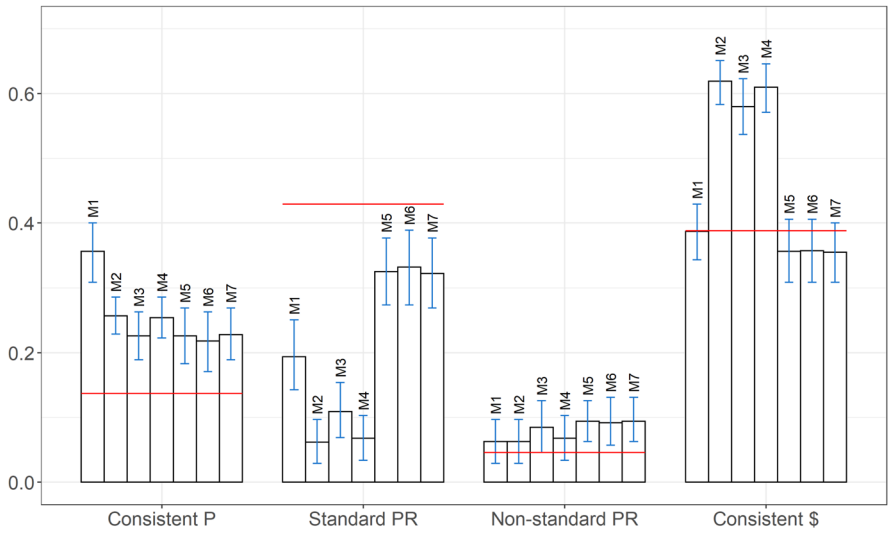

Fig. 5 The comparison of means and 95% confidence intervals of simulated proportions of CE PR patterns under the deliberately stochastic model of Fudenberg et al. (2015) across different simulation methods. The error bars are the 95% confidence intervals of the simulated proportions. The red horizontal lines are the actual proportions in the experiment. The simulations M1 to M4 use a consistent stochastic function estimated from the nine-repeated direct choices, from the CE R-ranges of the P-bet, the $-bet, and the mean η across the P-bet and the $-bet, respectively. The simulations M5 to M7 use the stochastic function estimated from the nine-repeated direct choices for the direct choice and the stochastic function for valuations estimated from the CE R-ranges of the P-bet, the $-bet, and the mean η across the P-bet and the $-bet.

Support: Fig. 5 (more details in Table 14 in Appendix "Additional results about hedging of preference uncertainty") Fudenberg et al. (2015) reports the means and 95% confidence intervals of these simulated proportions of the CE PR patterns for each stochastic function. As we can see, the mean simulated proportions in M1 to M4 overestimate the proportions of consistent preferences (choosing the P-bet or the $-bet consistently): The simulated proportion ranges from 74.3% to 87.6%, which is significantly higher than the actual proportion of 52.5% (Wilcoxon signed rank test, $p < 0.10$). Further, they fail to capture the magnitude and the asymmetry of standard PR and non-standard CE PR. The simulated proportions in M1 to M4 predict 12.5% to 25.7% of standard and non-standard PR, which is significantly lower than the actual proportion of 47.5% (Wilcoxon signed rank test, $p < 0.10$). Most importantly, the simulated proportions do not fully capture the asymmetry of standard PR and non-standard PR (asymmetry ratio of standard PR divided by non-standard PR ranging from 0.98 to 3.08 according to simulated proportions in M1 to M4, which is different from the ratio of 9.33 according to the actual proportions).

We now use the stochastic function from CE R-ranges to calculate the valuation probabilities, and use the actual choice ratio as the choice probability of choosing the P-bet in the direct choice. We find that, while the actual levels are

not in the 95% confidence intervals of the simulated proportions, the magnitude of the simulated levels in M5 to M7 are close to the actual proportions. The proportion of PR is 41.6% to 42.4% according to the simulated proportions, which is close to the actual proportion of 47.5%. Importantly, the simulated proportions are able to capture the asymmetry of the standard PR and non-standard PR, although not fully (asymmetry ratio of standard PR vs. non-standard PR 3.43 to 3.61 according to simulated proportions and 9.33 according to the actual proportions). Figure 5 illustrates directly the comparison of these two approaches.

5 Concluding remarks

In this paper we implemented a novel method in a PR experiment to examine the prevalence of conscious stochastic choice and its potential to explain PR. In this method, subjects need to pay a small cost for choosing a particular option or they can choose a free randomization option. We find that the majority of subjects exhibited conscious stochastic choices. While we also find substantial PR, our experimental results do not coincide with the theoretical predictions. Our results suggest that choices are not only stochastic but also depend on the elicitation procedures.

Our study highlights the importance of understanding the stochastic nature of choices and the influence of elicitation procedures in generating values and choices. The findings that choices can be consciously stochastic and violate procedural invariance have several important implications. First, they suggest that, when subjects' choices are stochastic, the standard method of asking subjects to state just one value may not be particularly informative. Instead, eliciting a range of possible values, as demonstrated in earlier works (Butler & Loomes, 2007; Cubitt et al., 2015; Loomes & Pogrebna, 2017) and our study, can offer richer insights. Together with earlier findings in the literature (Tversky et al., 1988, 1990), the violation of procedural invariance poses a significant challenge to the theory of revealed preferences. For example, when different elicitation tasks reveal systematically different levels of risk tolerance (see e.g., Loomes, 1988; Friedman et al., 2022), it is unclear which task we should rely on to capture risk preferences. Therefore, it is advisable to elicit values in the setting in which they are intended to be used, as this can also greatly influence the obtained value.

Second, the procedure dependency of stochastic choice underscores the need for caution when extrapolating preferences inferred from one elicitation procedure to represent those obtained through another. For example, a recent body of literature makes an insightful observation that noise can bias choices in paired tasks differently (Ballinger & Wilcox, 1997; Loomes, 2005; Blavatskyy, 2007, 2010; Bhatia & Loomes, 2017), but not valuations (Bernheim & Sprenger, 2020; Carrera et al., 2022). Consequently, McGranaghan et al. (2024) suggest to use valuations over choices to uncover preferences. Our finding that choices are stochastic but violate procedural invariance suggests such an interpretation may not always be warranted. While the violation of procedural invariance may not be a concern in McGranaghan et al. (2024) because the payoff ranges did not change across their lottery pairs, in general,

individuals may have context-dependent utility functions (Loomes & Sugden, 1983; Tversky et al., 1988, 1990), and preferences revealed through valuations may not be informative of those underlying choices (Butler et al., 2014a).

Third, our finding that stochastic choice can be conscious opens new ways for revealing and measuring the stochasticity in choices. Instead of repeating similar choices multiple times as in random utility models, we may be able to directly elicit subjects' subjective assessments of randomness and use them to capture stochastic choice. For example, we could allow subjects to assign randomization probabilities to various options in a choice set and pay them according to these randomization probabilities (Feldman & Rehbeck, 2022; Agranov & Ortoleva, 2023; Ong & Qiu, 2023; Halevy et al., 2023). Arts et al. (2020) provided some indicative evidence for the potential of such an approach for capturing stochastic choices: Across subjects and decisions, a higher randomization probability for an option corresponded to choosing that option more frequently (but not always) in binary choices. Another promising direction is to directly ask subjects for their belief about their intended choices out of a set and use that belief as a proxy for stochastic choice probabilities.

Appendix 1: Comparison with results in Butler and Loomes (2007)

Table 6 The percentages of subjects in each choice pattern. The classification of the four choice patterns in sub-table a) is based on subjects' decisions in the first direct choice and the indirect valuations, as in Butler and Loomes (2007). The classification of the four choice patterns in sub-table b) is based on subjects' more frequent (5 times or more) choice in the nine repeated direct choices and the indirect valuations. In both tables, numbers in parentheses are from Butler and Loomes (2007).

Sub-table a)

	$CE_P > CE_\$$	$CE_P \leq CE_\$$	$PE_P > PE_\$$	$PE_P \leq PE_\$$
Choose P-bet	13.7% (11.2%)	42.9% (57.3%)	47.4% (64.0%)	9.1% (4.5%)
Choose $-bet	4.6% (1.1%)	38.8% (30.3%)	20.0% (20.2%)	23.4% (11.2%)

Notes: The categorization of choosing the P-bet or the $-bet is based on the first direct choice.

Sub-table b)

	$CE_P > CE_\$$	$CE_P \leq CE_\$$	$PE_P > PE_\$$	$PE_P \leq PE_\$$
Choose P-bet	15.3% (11.2%)	38.6% (57.3%)	49.7% (64.0%)	4.6% (4.5%)
Choose $-bet	2.8% (1.1%)	42.6% (30.3%)	17.7% (20.2%)	28.0% (11.2%)

Notes: The categorization of choosing the P-bet or the $-bet is based on the more frequent (5 times or more) choice in the nine repeated direct choices.

Table 7 The choice patterns based on each of the direct choice between the P-bet and the $-bet. In the last row about the aggregate pattern, subjects who choose the P-bet no less than 5 times are defined as choosing the P-bet and as choosing the $-bet otherwise.

		$CE_P > CE_\$$	$CE_P \leq CE_\$$	$PE_P > PE_\$$	$PE_P \leq PE_\$$
Choice #1	Chose P	13.7%	42.9%	47.4%	9.1%
	Chose $	4.6%	38.9%	20.0%	23.4%
Choice #2	Chose P	13.7%	44.6%	49.7%	8.6%
	Chose $	4.6%	37.1%	17.7%	24.0%
Choice #3	Chose P	16.6%	40.6%	50.3%	6.9%
	Chose $	1.7%	41.1%	17.1%	25.7%
Choice #4	Chose P	13.7%	40.6%	47.4%	6.9%
	Chose $	4.6%	41.1%	20.0%	25.7%
Choice #5	Chose P	14.9%	44.0%	51.4%	7.4%
	Chose $	3.4%	37.7%	16.0%	25.1%
Choice #6	Chose P	15.4%	38.9%	47.4%	6.9%
	Chose $	2.9%	42.9%	20.0%	25.7%
Choice #7	Chose P	16.0%	39.4%	49.1%	6.3%
	Chose $	2.3%	42.3%	18.3%	26.3%
Choice #8	Chose P	15.4%	40.6%	50.3%	5.7%
	Chose $	2.9%	41.1%	17.1%	26.9%
Choice #9	Chose P	15.4%	40.0%	49.7%	5.7%
	Chose $	2.9%	41.7%	17.7	26.9%
Aggregate	Chose P	15.3%	38.6%	49.7%	4.6%
	Chose $	2.8%	42.6%	17.7%	28.0%

Result 6 Our experiment replicates the results in Butler and Loomes (2007). When the valuation is based on CE, a substantial proportion (47.5%) of subjects exhibited (standard or non-standard) preference reversals, of which majority are standard reversals (42.9% vs. 4.6%). When the valuation is based on PE, a less but non-negligible proportion (29.1%) of subjects exhibited (standard or non-standard) preference reversals, of which majority are non-standard reversals (20% vs. 9.1%).

In the experiment, subjects faced the direct choice between the P-bet and the $-bet nine times, and 50.3% of them did not always choose the same bet. Among the nine repeated choices, the average ratio of choosing the P-bet ($-bet) is 56.6% (43.4%).[5] We classify subjects' choice patterns in two ways: Following Butler and Loomes (2007) and using subjects' decision in the first direct choice or classifying choice patterns according to subjects' more frequent choices in the nine repeated direct choices. Table 6 shows the proportions of subjects in each of the four choice patterns.

[5] Butler and Loomes (2007) repeated the direct choice three times in their experiment. The average ratio of choosing the P-bet in their experiment is 72.3%.

In sub-table a) of Table 6, the four choice patterns are classified via subjects' decisions in the first direct choice and the indirect valuation choices. A substantial proportion of subjects exhibited (standard or non-standard) PR (47.5% when the valuation is based on CE and 29.1% when the valuation is based on PE). More importantly, there is a clear asymmetry, with the majority of subjects exhibiting standard PR rather than non-standard PR when the valuation is based on CE (42.9% vs. 4.6%), while more subjects exhibiting non-standard PR than standard PR when the valuation is based on PE (9.1% vs. 20%).

We also classified subjects' choice patterns according to the more frequent choices in the nine repeated direct choices and the indirect valuations. More specifically, subjects are regarded as choosing the P-bet over the $-bet if they selected the P-bet five times or more, and they are categorized as choosing the $-bet over the P-bet otherwise. According to sub-table b) of Table 6, there remain non-negligible proportions of subjects exhibiting (standard and non-standard) preference reversals (41.4% when the valuation is based on CE and 22.3% when the valuation is based on PE). Likewise, the majority of subjects exhibited standard PR rather than non-standard PR when the valuation is based on CE (38.6% vs.

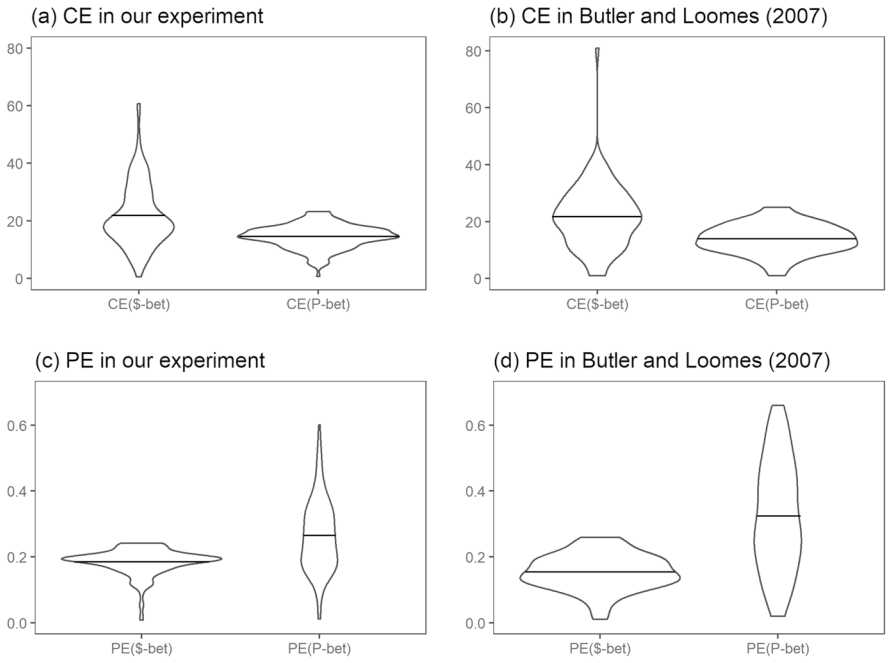

Fig. 6 The comparison of the elicited CE and PE between our study and Butler and Loomes (2007). The lines in the violins are the mean equivalents.

2.8%), and more subjects exhibited non-standard PR than standard PR (17.7% vs. 4.6%) when the valuation is based on PE. Findings from both classifications indicate that the overall choice patterns of our experiment are qualitatively similar to those of Butler and Loomes (2007). For the reader's convenience, Table 7 also reports the choice patterns classified according to subjects' decisions in each of the nine repeated direct choices and the indirect valuation choices. Qualitatively similar results hold.

We compare the experimental results in the two studies more closely by looking at the distributions of the elicited CE and PE. The average CE of the P-bet and the $-bet in our experiment are 14.6 and 21.9 separately, close to the corresponding values of 14 and 21.8 in Butler and Loomes (2007). The upper panel in Fig. 6 displays two violin plots of the distributions of CE in our experiment and those in Butler and Loomes (2007). It shows that, in both studies, the CE distribution of the $-bet is right-skewed, while the CE distribution of the P-bet is roughly symmetric around the mean value. The distributions of PE in the two studies are less comparable due to the use of different reference lotteries. Our reference lottery consisted of paying ¥100 with probability p and 0 otherwise, whereas the reference lottery in Butler and Loomes (2007) offered subjects $160 with probability p and 0 otherwise. The average PE for the P-bet and the $-bet are 26.5% and 18.6% in our experiment, while they are 32.4% and 15.5% in Butler and Loomes (2007). Nevertheless, the lower panel in Fig. 6 demonstrates that the distribution of PE in the two studies share similar features: Both are left-skewed for the $-bet while close to symmetry for the P-bet. Note additionally that changing the payment amount from 160 to 100 in the reference lottery should lead to a smaller asymmetry that favors non-standard reversals according to Butler and Loomes (2007), which is exactly what we find.

Result 7 There is no evidence that the bisection method caused any significant distortion in our experimental results.

Butler and Loomes (2007) used the incentive compatible incremental choice method. The comparability of our results with theirs thus suggests that our subjects were unlikely to behave strategically in the bisection method. We further checked the potential incentive incompatibility of the bisection method by comparing the CEs elicited in our experiment with the resulting CEs if subjects knew the complete procedure of the bisection method and behaved strategically to maximize their expected payoffs. We find that the average CE in our experiment is 14.6 for the P-bet and 21.9 for the $-bet, which are significantly smaller than the corresponding optimal CE of 18.75 and 61.625, respectively (one-sided Wilcoxon rank sum tests, $p < 0.01$ for both tests).

Appendix 2: Additional figures and tables

Table 8 The number (and percentage in parentheses) of subjects who chose the P-bet or the $-bet 0, 1, ..., 9 times in the nine repeated direct choices

	The number of times (percentages) choosing the bet in 9 direct choices									
	0	1	2	3	4	5	6	7	8	9
The P-bet	33	14	14	8	11	5	4	15	17	54
	(18.9%)	(8.0%)	(8.0%)	(4.6%)	(6.3%)	(2.9%)	(2.3%)	(8.6%)	(9.7%)	(30.9%)
The $-bet	54	17	15	4	5	11	8	14	14	33
	(30.9%)	(9.7%)	(8.6%)	(2.3%)	(2.9%)	(6.3%)	(4.6%)	(8.0%)	(8.0%)	(18.9%)

Table 9 The overview of subjects' risk attitudes according to two risk measures. Risk measure 1 is estimated from the risk attitude measurement task where we vary the winning amount of money in the reference lottery. Risk measure 2 is estimated from the measurement task where we vary the winning probability in the reference lottery

		Risk measure 1			
		Averse	Neutral	Seeking	Aggregate
Risk measure 2	Averse	138	2	0	140
	Neutral	20	5	3	28
	Seeking	5	0	2	7
	Aggregate	163	7	5	175

Additional results about models of incomplete/imprecise preferences

Results about PE PR patterns

The following result summarizes the findings about PE PR patterns:

Result 8 While preference incompleteness/imprecision explains the PE PR patterns better than CE PR patterns, in general, results on PE PR patterns are similar to those in CE PR patterns.

In general, results on PE PR patterns are similar to those in CE PR patterns. A sizeable proportion of subjects (37.4%) have PE outside their PE R-range in at least one bet. Among subjects who had no overlapping PE R-ranges, 49.6% chose differently in the nine-repeated choices; among those who chose differently in the nine-repeated choices, 79.5% had no overlapping PE R-ranges. For more details please refer to Table 11. Table 13 suggests that the simulated proportions capture the general PE PR patterns better than the CE PR patterns by correctly estimating the PE PR pattern of choosing the P-bet consistently as

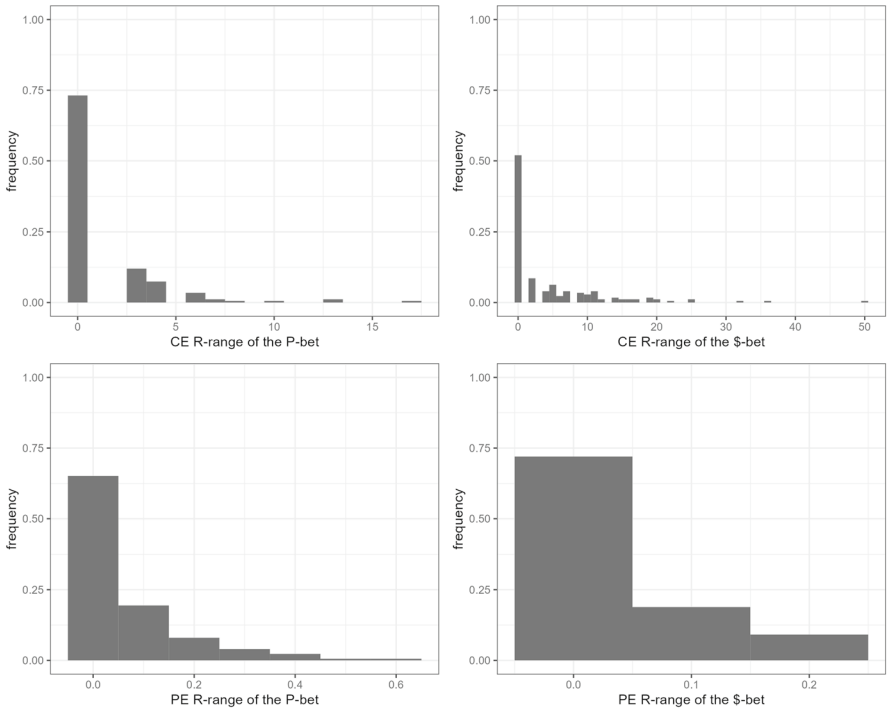

Fig. 7 The histogram of the size of CE R-range of the P-bet and the $-bet. The average size of the CE R-range is 1.35 for the P-bet and 4.47 for the $-bet. The average size of the PE R-range is 0.06 for the P-bet and 0.04 for the $-bet

the most frequent and choosing the $-bet consistently as the second most frequent. However, like the simulations in CE PR patterns, the simulated proportions overestimate the proportion of consistent choices and fail to capture the magnitude as well as the asymmetry of PE PR. Assuming those who chose differently in the nine-repeated direct choices randomly select between the two bets(50%:50%) and use the PE R-ranges to calculate the probability of valuing the P-bet higher than the $-bet, the simulated proportions are then close to the actual proportions. They capture the ranking of the four PE PR patterns as well as the asymmetry of the non-standard PR versus the standard PR. Moreover, for all PE PR patterns the actual proportions fall in the 95% confidence intervals of the simulated proportions. The same result holds when we use the actual ratio of choosing the P-bet in the nine-repeated choices as the probability of choosing the P-bet in the direct choice.

Table 10 Means and 95% confidence intervals (in square brackets) of simulated proportions of CE PR patterns according to models of preference incompleteness/imprecision. In M1: Direct choices, we assume the probability of valuing the P-bet higher than the $-bet is 50% if subjects chose different bets in the nine-repeated direct choices and it is 1 (or 0) if subjects chose the P-bet (or the $-bet) for nine times. In M2: CE R-ranges 1, we assume subjects choose randomly (50%:50%) when their CE R-ranges for the P-bet and the $-bet overlap. In M3: CE R-ranges 2, we calculate the choice probability from CE R-ranges by assuming that subjects' choices are determined by randomly sampling the potential values of the bet in the CE R-range. In M4 and M5: Two stochastic functions 1 and 2, we calculate the probability of choosing the P-bet in the direct choice from the nine-repeated direct choices and use CE R-ranges to calculate the probability of valuing the P-bet higher than the $-bet. More specifically, in M4: Two stochastic functions 1, we assume random choice (50%:50%) if subjects chose differently in the nine-repeated direct choices. In M5: Two stochastic functions 2, we use the actual choice ratio in the nine-repeated direct choices as the probability of choosing the P-bet in the direct choice

	The CE PR patterns			
	Consis. P	Sth. PR	Nonsth. PR	Consis. $
Actual	13.7%	42.9%	4.6%	38.8%
M1: Direct	43.6%	12.7%	12.4%	31.3%
choices	[39.4%, 48.0%]	[8.6%, 17.1%]	[8.6%, 16.6%]	[26.9%, 35.4%]
M2: CE R-ranges 1	25.5%	10.0%	6.0%	58.5%
(50%:50%)	[22.3%, 28.6%]	[6.3%, 13.7%]	[2.9%, 9.1%]	[54.9%, 62.3%]
M3: CE R-ranges 2	26.5%	4.9%	4.9%	63.6%
(valuation)	[24.0%, 29.1%]	[2.3%, 8.0%]	[2.3%, 8.0%]	[60.6%, 66.9%]
M4: Two stochastic functions 1	22.2%	33.8%	9.3%	34.7%
(50%:50% from choices)	[18.9%, 25.7%]	[28.6%, 38.9%]	[6.3%, 12.6%]	[30.3%, 39.4%]
M5: Two stochastic	23.9%	32.4%	7.6%	36.1%
functions 2 (actual ratios)	[20.6%, 27.4%]	[28.0%, 37.1%]	[5.1%, 10.3%]	[32.0%, 40.0%]

Table 11 Sub-table a) reports the number of subjects who had no R-range, who had positive R-ranges with PE falling in the R-range, and who had positive R-ranges but with PE outside the R-range. The group "No R-range" includes subjects who had no positive ranges in any of the tasks. The group "In the R-range" includes subjects who had positive R-ranges in at least one task, and their elicited PE were in the R-range. The group "Outside the R-range" includes subjects who had positive R-ranges in at least one task and their elicited PE were outside the R-range in at least one task. Sub-table b) reports the number of subjects who had overlapping PE R-ranges, who had non-overlapping PE R-ranges as well as those who chose consistently or who chose differently in the nine-repeated direct choices

Sub-table a)				
	No	In the	Outside the R-range	
	R-range	R-range	larger than	lower than
			the upper bound	the lower bound
PE_P	106 (60.6%)	43 (24.6%)	6 (3.4%)	20 (11.4%)
$PE_\$$	99 (56.6%)	65 (37.1%)	4 (2.3%)	7 (4.2%)
PE aggregate	84 (48.0%)	57 (32.6%)	34 (19.4%)	

Sub-table b)		
	In the nine-repeated direct choices	
	Chose consistently	Chose differently
Overlapping PE R-ranges	16	18
Non-overlapping PE R-ranges	71	70

Table 12 Sub-table a) shows the number of subjects who had overlapping CE R-ranges or non-overlapping CE R-ranges, as well as those who chose consistently or differently in the nine-repeated direct choices. Sub-table b) displays the number of subjects who had overlapping PE R-ranges or non-overlapping PE R-ranges, as well as those who chose consistently or differently in the nine-repeated direct choices

Sub-table a)		
	In the nine-repeated direct choices	
	Chose consistently	Chose differently
Overlapping CE R-ranges	40	16
Non-overlapping CE R-ranges	78	41

Notes: Subjects who chose the same bet in eight of the nine repeated direct choices were also included in the "Chose consistently" group.

Sub-table b)		
	In the nine-repeated direct choices	
	Chose consistently	Chose differently
Overlapping PE R-ranges	22	12
Non-overlapping PE R-ranges	96	45

Notes: Subjects who chose the same bet in eight of the nine repeated direct choices were also included in the "Chose consistently" group.

Table 13 The comparison of means and 95% confidence intervals (in square brackets) of simulated proportions of PE PR patterns according to models of preference incompleteness/imprecision. In M1: Direct choices, we assume the probability of valuing the P-bet higher than the $-bet is 50% if subjects chose different bets in the nine repeated direct choices and it is 1 (or 0) if subjects chose the P-bet (or the $-bet) for nine times. In M2: PE R-ranges 1, we assume subjects choose randomly (50%:50%) when their R-ranges for the P-bet and the $-bet overlap. In M3: PE R-ranges 2, we calculate the choice probability from PE R-ranges by assuming that subjects' choices are determined by randomly sampling the potential values of the bet in the PE R-range. In M4 and M5: Two stochastic functions 1 and 2, we calculate the probability of choosing the P-bet from the nine-repeated direct choices and use the PE R-ranges to calculate the probability of valuing the P-bet higher than the $-bet. In M4: Two stochastic functions 1, we assume random choice (50%:50%) if subjects chose differently in the nine-repeated direct choices. In M5: Two stochastic functions 2, we use the actual ratio of choosing the P-bet in the nine repeated direct choices as the probability of choosing the P-bet in the direct choice

	The PE PR patterns			
	Consis. P	Sth. PR	Nonsth. PR	Consis. $
Actual	47.4%	9.1%	20.0%	23.4%
M1: Direct	43.6%	12.7%	12.4%	31.3%
choices	[39.4%, 48.0%]	[8.6%, 17.1%]	[8.6%, 16.6%]	[26.9%, 35.4%]
M2: PE R-range 1	63.3%	3.0%	6.7%	27.0%
(50%:50%)	[60.6%, 66.3%]	[1.1%, 5.1%]	[4.0%, 9.7%]	[25.1%, 29.1%]
M3: PE R-range 2	67.2%	2.8%	2.8%	27.3%
(valuation)	[64.6%, 69.7%]	[0.6%, 5.1%]	[0.6%, 5.1%]	[25.7%, 29.1%]
M4: Two stochastic functions 1	46.5%	9.5%	23.4%	20.6%
(50%:50% from choices)	[41.7%, 51.4%]	[6.3%, 12.6%]	[18.9%, 28.0%]	[17.7%, 24.0%]
M5: Two stochastic	48.8%	7.5%	21.1%	22.6%
functions 2 (actual ratios)	[45.1%, 52.6%]	[5.1%, 10.3%]	[17.1%, 25.1%]	[20.0%, 25.7%]

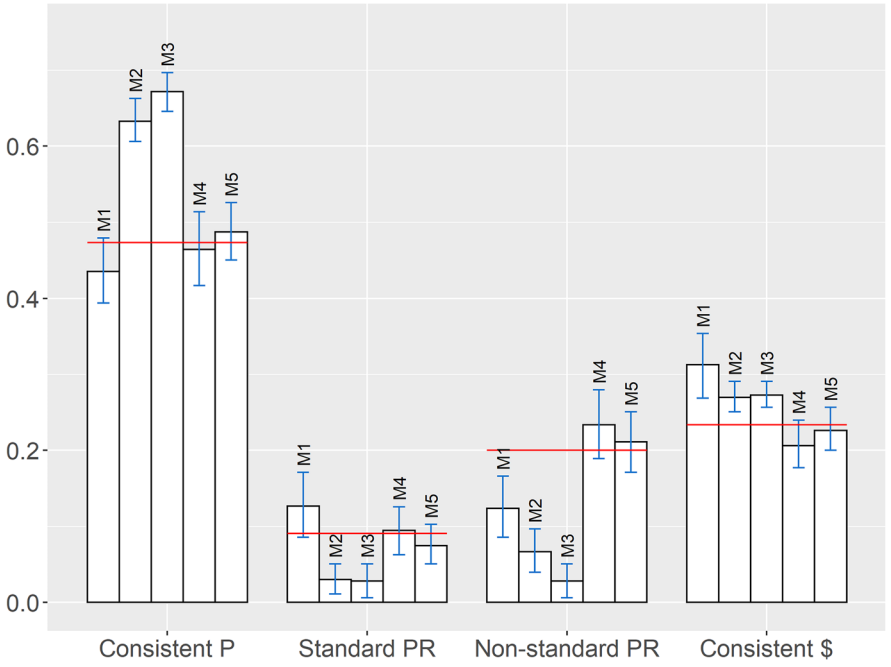

Fig. 8 The comparison of means and 95% confidence intervals of simulated proportions of PE PR patterns according to models of preference incompleteness/imprecision across different simulation methods. The error bars are the 95% confidence intervals of the simulated proportions. The red horizontal lines are the actual proportions in the experiment. The horizontal red line is the actual proportion in the experiment. In M1, we assume the probability of valuing the P-bet higher than the $-bet is 50% if subjects chose different bets in nine repeated direct choices and it is 1 (or 0) if subjects chose the P-bet (or the $-bet) for nine times. In M2, we assume subjects choose randomly (50%:50%) when their R-ranges for the P-bet and the $-bet overlap. In M3, we calculate the choice probability from PE R-ranges by assuming that subjects' choices are determined by randomly sampling the potential values of the bet in the PE R-range. In M4 and M5, we calculate the probability of choosing the P-bet in the direct choice from the nine-repeated direct choices and use PE R-ranges to calculate the probability of valuing the P-bet higher than the $-bet. In M4, we assume random choice (50%:50%) if subjects chose differently in the nine-repeated direct choices. In M5, we use the actual choice ratio in the nine repeated direct choices as the probability of choosing the P-bet in the direct choice

Additional results about hedging of preference uncertainty Fudenberg et al. (2015)

Results about PE PR patterns

Result 9 In the deliberately stochastic model of Fudenberg et al. (2015), results on PE PR patterns are in general similar to those in CE PR patterns.

The median η estimated from the PE of the P-bet is 4.37 (mean=7.41) and from the PE of the $-bet is 4.43 (mean=9.81). The median η estimated from the nine-repeated direct choices is -0.24 (mean=3.80). Wilcoxon signed rank tests that compare η estimated from the direct choices and η estimated from the PE for both the P-bet and the $-bet suggest that the difference is significant in both comparisons ($p < 0.01$). See Table 16 for more details. Table 17 reports the simulated proportions of the PE PR patterns, indicating that the simulated proportions capture the general PE PR patterns. When we use the actual choice ratio from the nine-repeated direct choice as the choice probability of choosing the P-bet in the direct choice and the probabilities calculated from the PE R-ranges

Table 14 Means and 95% confidence intervals (in square brackets) of the simulated proportions of the CE PR patterns under the deliberately stochastic model of Fudenberg et al. (2015). The simulations M1 to M4 use a consistent stochastic function estimated from the nine-repeated direct choices, from the CE R-ranges of the P-bet, the CE-Ranges of the $-bet, and the mean η across the P-bet and the $-bet, respectively. The simulations M5 to M7 use the stochastic function estimated from the nine-repeated direct choices for the direct choice and the stochastic function estimated from the CE R-ranges of the P-bet, the CE R-ranges of the $-bet, and the mean η across the P-bet and the $-bet for valuation choices, respectively

	The CE PR patterns			
	Consis. P	Std. PR	Nonstd. PR	Consis. $
Actual	13.7%	42.9%	4.6%	38.8%
M1: Direct	35.6%	19.4%	6.3%	38.7%
Choices	[30.9%, 40.0%]	[14.3%, 25.1%]	[2.9%, 9.7%]	[34.3%, 42.9%]
M2: CE R-range	25.7%	6.2%	6.3%	61.9%
P-bet	[22.9%, 28.6%]	[2.9%, 9.7%]	[2.9%, 9.7%]	[58.3%, 65.1%]
M3: CE R-range	22.6%	10.9%	8.5%	58.0%
$-bet	[18.9%, 26.3%]	[6.9%, 15.4%]	[4.6%, 12.6%]	[53.7%, 62.3%]
M4: CE R-range	25.5%	6.7%	6.8%	61.0%
mean	[22.3%, 28.6%]	[3.4%, 10.3%]	[3.4%, 10.3%]	[57.1%, 64.6%]
M5: Two stochastic	22.6%	32.5%	9.4%	35.6%
functions (P-bet)	[18.3%, 26.9%]	[27.4%, 37.7%]	[6.3%, 12.6%]	[30.9%, 40.6%]
M6: Two stochastic	21.8%	33.2%	9.2%	35.7%
functions ($-bet)	[17.1%, 26.3%]	[27.4%, 38.9%]	[5.7%, 13.1%]	[30.9%, 40.6%]
M7: Two stochastic	22.8%	32.2%	9.4%	35.5%
functions (mean)	[18.9%, 26.9%]	[26.9%, 37.7%]	[6.3%, 13.1%]	[30.9%, 40.0%]

Table 15 Regression results of CE R-range and PE R-range on ambiguity attitude and the desirability of control, with standard errors in parentheses. * denotes $p < 0.10$, ** denotes $p < 0.05$, *** denotes $p < 0.01$

	Dependent variable:					
	CE R-range			PE R-range		
	(1)	(2)	(3)	(4)	(5)	(6)
Amb. aversion	0.12	-	0.08	0.01*	-	0.005
	(0.20)	-	(0.20)	(0.00)	-	(0.00)
Less control	-	0.08**	0.08**	-	0.002***	0.002***
	-	(0.03)	(0.03)	-	(0.00)	(0.00)
Risk aversion	0.03	−0.07	−0.04	0.01***	0.004*	0.01**
	(0.18)	(0.16)	(0.18)	(0.00)	(0.00)	(0.00)
Constant	2.00	5.83***	5.12**	−0.04	0.08***	0.04
	(2.11)	(1.71)	(2.47)	(0.03)	(0.03)	(0.04)
Adj. R^2	−0.01	0.01	0.01	0.02	0.07	0.07

as valuation probabilities, the simulated proportions are quite close to the actual proportions. They capture the ranking of the four PE PR patterns as well as the asymmetry of the non-standard PR versus standard PR. Furthermore, the actual proportions fall in the 95% confidence intervals of the simulated proportions.

Table 16 The fitted η in the deliberately stochastic models. Standard deviations are in parentheses. The η of the stochastic function from the direct choices is estimated from the equation $P(P-bet \succ \$-bet) = \frac{e^{\eta U(P)}}{e^{\eta U(P)}+e^{\eta U(\$)}}$. The η of the stochastic function from the PE R-ranges is estimated from the equation $\eta_X = \frac{2.55}{(\bar{p}-\underline{p})u(100)}$, where X is the P-bet or the \$-bet. Wilcoxon signed rank tests check the difference in the estimated η from the direct choices and the PE R-ranges. * denotes $p < 0.10$, ** denotes $p < 0.05$, *** denotes $p < 0.01$

	The estimated η		
	from the direct choices	from the PE R-ranges	
		P-bet	\$-bet
Mean	-0.24	7.41***	9.81***
	(3.80)	(8.59)	(11.59)
Median	0.28	4.37	4.43

Table 17 Means and 95% confidence intervals (in square brackets) of simulated proportions of the PE PR patterns under the deliberately stochastic model of Fudenberg et al. (2015). The simulations M1 to M4 use a consistent stochastic function estimated from the nine-repeated direct choices, from the PE R-ranges of the P-bet, the PE R-ranges of the $-bet, and the mean η across the P-bet and the $-bet, respectively. The simulations M5 to M7 use the stochastic function estimated from the nine-repeated direct choices for the direct choice and the stochastic function estimated from the PE R-ranges of the P-bet, the PE R-ranges of the $-bet, and the mean η across the P-bet and the $-bet for valuation choices respectively

	The PE PR patterns			
	Consis. P	Std. PR	Nonstd. PR	Consis. $
Actual	47.4%	9.1%	20.0%	23.4%
M1: Direct	51.4%	3.6%	34.8%	10.2%
Choices	[46.3%, 56.6%]	[1.1%, 6.3%]	[29.1%, 40.6%]	[6.9%, 13.7%]
M2: PE R-range	62.6%	5.4%	10.3%	21.6%
P-bet	[58.9%, 66.3%]	[2.3%, 8.6%]	[6.9%, 14.3%]	[18.3%, 24.6%]
M3: PE R-range	65.6%	4.2%	7.8%	22.5%
$-bet	[62.3%, 68.6%]	[1.7%, 6.9%]	[4.6%, 10.9%]	[19.4%, 25.1%]
M4: PE R-range	64.9%	4.6%	8.3%	22.2%
mean	[61.7%, 68.0%]	[1.7%, 7.4%]	[5.1%, 12.0%]	[19.4%, 25.1%]
M5: Two stochastic	47.8%	7.3%	25.2%	19.8%
functions (P-bet)	[42.9%, 52.6%]	[4.0%, 10.9%]	[20.0%, 30.3%]	[16.0%, 23.4%]
M6: Two stochastic	48.3%	6.8%	25.1%	19.9%
functions ($-bet)	[43.4%, 53.1%]	[4.0%, 10.3%]	[20.0%, 30.3%]	[16.0%, 23.4%]
M7: Two stochastic	48.1%	6.9%	25.1%	19.9%
functions (mean)	[43.4%, 53.1%]	[4.0%, 10.3%]	[20.0%, 30.3%]	[16.0%, 23.4%]

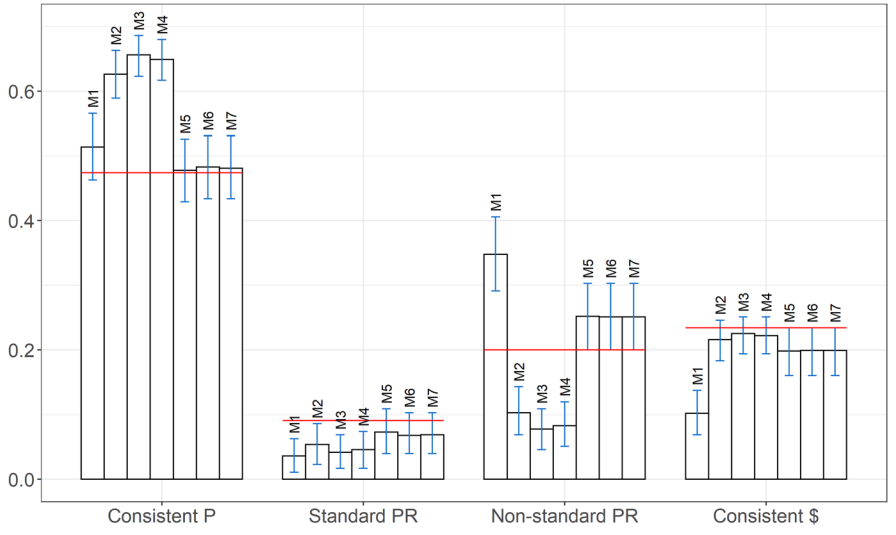

Fig. 9 The comparison of means and 95% confidence intervals of simulated proportions of PE PR patterns under the deliberately stochastic model of Fudenberg et al.

(2015) across different simulation methods. The error bars are the 95% confidence intervals of the simulated proportions. The red horizontal lines are the actual proportions in the experiment. The simulations M1 to M4 use a consistent stochastic function estimated from the nine-repeated direct choices, from the PE R-ranges of the P-bet, the PE R-ranges of the $-bet, and the mean η across the P-bet and the $-bet, respectively. The simulations M5 to M7 use the stochastic function estimated from the nine-repeated direct choices for the direct choice and the stochastic function estimated from the PE R-ranges of the P-bet, the PE R-ranges of the $-bet, and the mean η across the P-bet and the $-bet for valuation choices respectively

Appendix 3: Experimental materials

Translated decision screens in the CE valuation and PE valuation

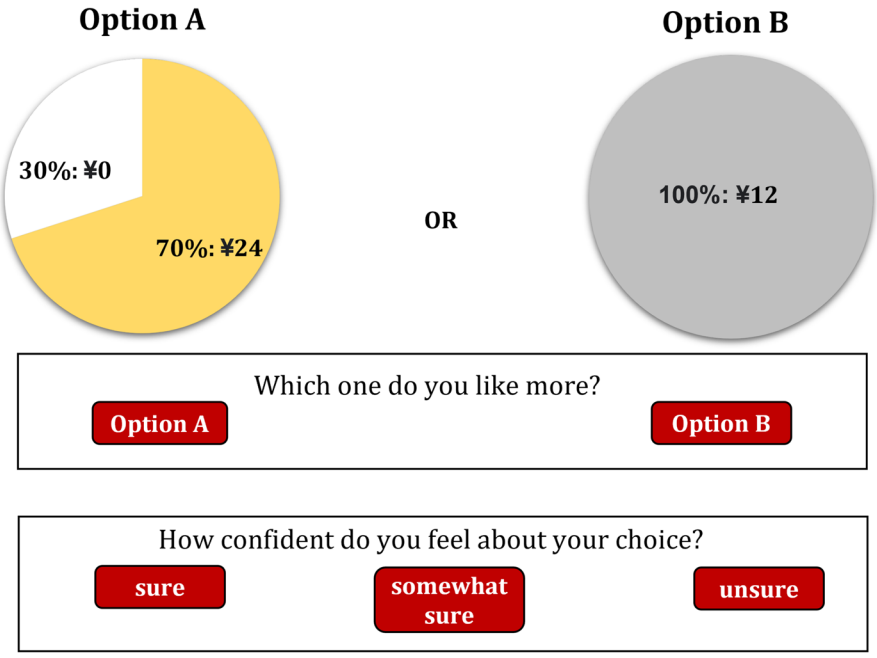

Fig. 10 Display of binary choice in CE valuation

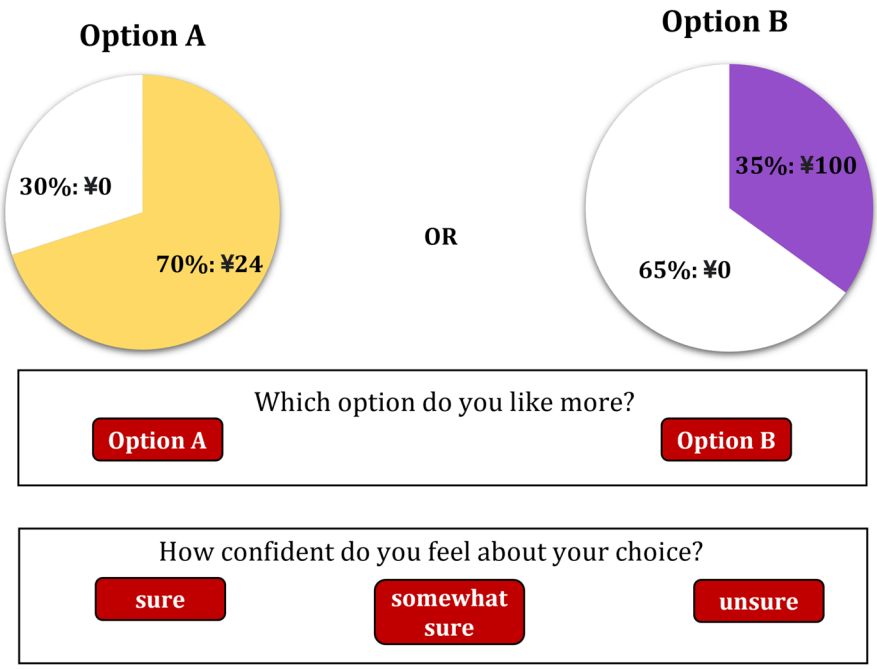

Fig. 11 Display of binary choice in PE valuation

Risk and ambiguity attitudes

In the elicitation of the risk attitudes, subjects face two multiple price lists (MPL). In each row of the lists, subjects face two options: a sure payment and a lottery. In the first MPL, subjects face a sure payment of ¥60 and a lottery of receiving ¥110 with probability 50% and ¥X otherwise. The amount of ¥X increases from ¥0 to ¥60 from the top to the bottom rows. In the second MPL, subjects face a sure payment of ¥50 and a lottery of receiving ¥90 with probability p% and ¥10 otherwise. Moving down the rows p% increases from 10% to 100%. Note that the first MPL is close to the CE valuation where we vary the monetary payoff across choices, while the second MPL is close to the PE valuation where we vary the winning probabilities across choices. We ask subjects to indicate the row they switch from one option to another. Tables 18 and 19 illustrate the tasks.

The literature documents substantial order effects: the elicited risk attitude may depend on the presentation order (ascending or descending) of the reference option. However, using one order in MPL is unlikely to bias our results. This is because PR is about the disparity between valuations and choices, and we use the same risk attitude for valuations and choices in our estimation of stochastic functions. Consequently, even if the elicited risk attitude were influenced by the order, its impact on the difference between valuations and choices would be limited.

Table 18 Risk task 1 (varying the monetary amounts in the reference lottery)

	Option 1	Option 2	Your choice
situation 1	100%: 60	50%: 110, 50%: 0	prefer option 2
situation 2	100%: 60	50%: 110, 50%: 5	prefer option 2
situation 3	100%: 60	50%: 110, 50%: 10	prefer option 2
situation 4	100%: 60	50%: 110, 50%: 15	prefer option 2
situation 5	100%: 60	50%: 110, 50%: 20	prefer option 2
situation 6	100%: 60	50%: 110, 50%: 25	prefer option 2
situation 7	100%: 60	50%: 110, 50%: 30	prefer option 2
situation 8	100%: 60	50%: 110, 50%: 35	prefer option 2
situation 9	100%: 60	50%: 110, 50%: 40	prefer option 2
situation 10	100%: 60	50%: 110, 50%: 45	prefer option 2
situation 11	100%: 60	50%: 110, 50%: 50	prefer option 2
situation 12	100%: 60	50%: 110, 50%: 55	prefer option 2
situation 13	100%: 60	50%: 110, 50%: 60	prefer option 2

We use the classical Ellsberg paradox to elicit ambiguity attitude. There are two boxes, box U and box K, each of which contained 100 balls of either red or black. The composition of red and black balls is known in box K while unknown in box U. Subjects first choose their preferred color – red or black – as the winning color. They then face an MPL with 11 rows. In each of them they chose to let the computer randomly draw a ball from box U or box K. Moving down the rows, box U remains the same while the proportion of winning colored balls in box K increases from 0% to 100%. We ask subjects to indicate the row they switch from drawing a ball from box U to drawing a ball from box K. For experimental payment, one out of the 11 rows will be randomly chosen and subjects receive ¥100 if the randomly drawn ball from their chosen box at that row had the same color as their winning color. We take the average of the proportions of winning color balls in box K at the switching row and the above row as the matching probability. Table 20 illustrates the task.

Table 19 Risk task 2 (varying the probabilities in the reference lottery)

	Option 1	Option 2	Your choice
situation 1	100%: 50	10%: 90, 90%: 10	prefer option 2
situation 2	100%: 50	20%: 90, 80%: 10	prefer option 2
situation 3	100%: 50	30%: 90, 70%: 10	prefer option 2
situation 4	100%: 50	40%: 90, 60%: 10	prefer option 2
situation 5	100%: 50	50%: 90, 50%: 10	prefer option 2
situation 6	100%: 50	60%: 90, 40%: 10	prefer option 2
situation 7	100%: 50	70%: 90, 30%: 10	prefer option 2
situation 8	100%: 50	80%: 90, 20%: 10	prefer option 2
situation 9	100%: 50	90%: 90, 10%: 10	prefer option 2
situation 10	100%: 50	100%: 90, 0%: 10	prefer option 2

Table 20 Ambiguity task

	Box U	Box K	Your choice
situation 1	? red balls, (100-?) black balls	0 red balls, 100 black balls	prefer Box K
situation 2	? red balls, (100-?) black balls	20 red balls, 80 black balls	prefer Box K
situation 3	? red balls, (100-?) black balls	30 red balls, 70 black balls	prefer Box K
situation 4	? red balls, (100-?) black balls	40 red balls, 60 black balls	prefer Box K
situation 5	? red balls, (100-?) black balls	45 red balls, 55 black balls	prefer Box K
situation 6	? red balls, (100-?) black balls	50 red balls, 50 black balls	prefer Box K
situation 7	? red balls, (100-?) black balls	55 red balls, 45 black balls	prefer Box K
situation 8	? red balls, (100-?) black balls	60 red balls, 40 black balls	prefer Box K
situation 9	? red balls, (100-?) black balls	70 red balls, 30 black balls	prefer Box K
situation 10	? red balls, (100-?) black balls	80 red balls, 20 black balls	prefer Box K
situation 11	? red balls, (100-?) black balls	100 red balls, 0 black balls	prefer Box K

Note: you chose color Red as your wining color

The bisection process

To illustrate, consider the elicitation of CE via the bisection process. For each choice, we start by presenting the mid-point of the possible range of CE as the sure payment option. Let Y denote the positive payment in the bet ($Y = 24$ for the P-bet and $Y = 80$ for the \$-bet). The CE of a bet should be between 0 and Y considering first-order stochastic dominance. The sure payment in the first choice should be $Y/2$, which is the mid-point of the possible range $(0, Y)$. If a subject chooses the bet in the first choice, the possible range would be adjusted to $(Y/2, Y)$ and she will face the mid-point $3Y/4$ as the sure payment in the second choice. Instead, if she chooses the sure payment in the first choice, the possible range will be adjusted to $(0, Y/2)$ and she will face the mid-point $Y/4$ in the second choice. We continue this process four times for the P-bet and six times for the \$-bet due to the larger payment amount in the \$-bet.

Subjects may make mistakes, and these mistakes could have a large impact on CE or PE if they occur in the initial choices of the bisection process. To mitigate this impact, we add an additional choice after the second choice in the bisection process. For example, a subject that chooses the P-bet for $Y/2$ will not only face $3Y/4$, but also $Y/4$ as additional choice before continuing with the next step of the bisection process. If subjects' decisions in the first three choices are inconsistent, e.g., choosing the bet over $Y/2$ (implying the bet $\succ Y/2$) but choosing $Y/4$ over the bet (implying $Y/4 \succ$ the bet), they return to the initial choice with the sure payment of $Y/2$.

The elicitation of the randomization range (R-range)

We illustrate the elicitation of the R-range with the CE R-range task. In each choice subjects have three options: they can 1) pay a small cost (¥0.10) and choose the fixed bet, 2) pay a small cost (¥0.10) and choose the sure payment, or 3) choose a

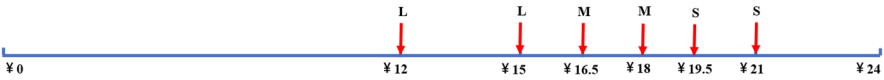

Fig. 12 An example of eliciting the CE R-range for the P-bet. S stands for choosing the sure payment, L stands for choosing the P-bet, and M stands for the randomization option. We use $\underline{y}$ and $\overline{y}$ to represent the lowest and the highest sure payments of the R-range separately. In the example, a subject chooses the P-bet when the sure payment is ¥12, and picks the randomization option when the sure payment increases to ¥18. We find $\underline{y}$ in the range [¥12, ¥18]. In the following choices, the subject chooses the P-bet when the sure payment is ¥15 and selects the randomization option when the sure payment is ¥16.5. We therefore obtain a $\underline{y}$ of ¥15.75 for the subject. Similarly, we find $\overline{y}$ in the range [¥18, ¥24]. In the choices of finding $\overline{y}$, the subject first selects the sure payment when it is ¥21 and picks the sure payment again when it is ¥19.5. We thus obtain a $\overline{y}$ of ¥18.75 for the subject. Finally, the elicited monetary randomization range for the subject is [¥15.75, ¥18.75]

randomization option for free, according to which a computer randomly selects one of the two options. Across choices, the sure payment changes according to the bisection method. Let Y denote the positive payment of the bet, and y^r the sure payment for which subjects chose the randomization option for the first time. Let $\underline{y}$ and $\overline{y}$ denote the lower and the upper bound of the R-range separately. It is clear that $0 \leq \underline{y} \leq y^r$ and $y^r \leq \overline{y} \leq Y$. To find $\underline{y}$, we raise (lower) the sure payment in the next choice if subjects choose the bet (the randomization option) according to the bisection procedure with the same number of iterations as in CE valuation. Similarly, to find $\overline{y}$ we increase (decrease) the sure payment in the next choice if subjects choose the randomization option (the sure payment) according to the bisection procedure. If subjects never choose the randomization option, their CE R-range is zero. Figure 12 provides an example decision screen for the elicitation of the CE R-range for the P-bet.

Measuring the desirability of control

The questions to elicit the control desirability in the questionnaire (Burger & Cooper, 1979; Gebhardt & Brosschot, 2002):

1. When I see a problem, I prefer to do something about it rather than sit by and let it continue.
2. I wish I could push many of life's daily decisions off on someone else.
3. There are many situations in which I would prefer only one choice rather than having to make a decision.
4. I like to wait and see if someone else is going to solve a problem so that I don't have to be bothered by it.
5. I prefer a job where I have a lot of control over what I do and when I do it.
6. I try to avoid situations where someone else tells me what to do.
7. Others usually know what is best for me.
8. I enjoy making my own decisions.
9. I enjoy having control over my own destiny.
10. I prefer to avoid situations where someone else has to tell me what it is I should be doing.

Data availibility The authors confirm that all data generated or analysed during this study are included in this published article. Furthermore, primary and secondary sources and data supporting the findings of this study were all publicly available at the time of submission.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Abdellaoui, M. (2000). Parameter-free elicitation of utility and probability weighting functions. *Management Science, 46*(11), 1497–1512.

Abdellaoui, M., Bleichrodt, H., & L'Haridor, O. (2008). A tractable method to measure utility and loss aversion under prospect theory. *Journal of Risk and Uncertainty, 36*(3), 245–266.

Agranov, M., & Ortoleva, P. (2017). Stochastic choice and preferences for randomization. *Journal of Political Economy, 125*(1), 40–68.

Agranov, M., & Ortoleva, P. (2023). Ranges of preferences and randomization. *Review of Economics and Statistics, Forthcoming*.

Alós-Ferrer, C., Buckenmaier, J., & Garagnani, M. (2020). *Stochastic choice and preference reversals*. ECON Working Papers 370, Department of Economics, University of Zurich.

Alós-Ferrer, C., Fehr, E., & Netzer, N. (2021). Time will tell: Recovering preferences when choices are noisy. *Journal of Political Economy, 129*(6), 1828–1877.

Alós-Ferrer, C., Granić, D.-G., Kern, J., & Wagner, A. K. (2016). Preference reversals: Time and again. *Journal of Risk and Uncertainty, 52*(1), 65–97.

Apesteguia, J., & Ballester, M. (2018). Monotone stochastic choice models: The case of risk and time preferences. *Journal of Political Economy, 126*(1), 74–106.

Arts, S., Ong, Q., & Qiu, J. (2020). *Measuring subjective decision confidence*. MPRA Paper 106811, University Library of Munich, Germany.

Aumann, R. J. (1962). Utility theory without the completeness axiom. *Econometrica, 30*(3), 445–462.

Ballinger, T. P., & Wilcox, N. T. (1997). Decisions, error and heterogeneity. *Economic Journal, 107*(443), 1090–1105.

Bayrak, O. K., & Hey, J. D. (2020). Understanding preference imprecision. *Journal of Economic Surveys, 34*(1), 154–174.

Becker, G. M., Degroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science, 9*(3), 226–232.

Bernheim, B. D., & Sprenger, C. D. (2020). On the empirical validity of cumulative prospect theory: Experimental evidence of rank-independent probability weighting. *Econometrica, 88*(4), 1363–1409.

Bhatia, S., & Loomes, G. (2017). Noisy preferences in risky choice: A cautionary note. *Psychological Review, 124*(5), 678–687.

Blavatskyy, P. R. (2007). Stochastic expected utility theory. *Journal of Risk and Uncertainty, 34*(3), 259–286.

Blavatskyy, P. R. (2009). Preference reversals and probabilistic decisions. *Journal of Risk and Uncertainty, 39*(3), 237–250.

Blavatskyy, P. R. (2010). Reverse common ratio effect. *Journal of Risk and Uncertainty, 40*(3), 219–241.

Blavatskyy, P. R. (2012). Probabilistic choice and stochastic dominance. *Economic Theory, 50*(1), 59–83.

Bleichrodt, H., Doctor, J. N., Gao, Y., Li, C., Meeker, D., & Wakker, P. P. (2019). Resolving Rabin's paradox. *Journal of Risk and Uncertainty, 59*(3), 239–260.
Bostic, R., Herrnstein, R., & Luce, R. (1990). The effect on the preference-reversal phenomenon of using choice indifferences. *Journal of Economic Behavior and Organization, 13*(2), 193–212.
Bouacida, E. (2021). *Identifying Choice Correspondences*. Working Papers 327800275, Lancaster University Management School, Economics Department.
Burger, J. M., & Cooper, H. M. (1979). The desirability of control. *Motivation and Emotion, 3*(4), 381–393.
Busemeyer, J., & Townsend, J. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review, 100*(3), 432–459.
Butler, D., Isoni, A., Loomes, G., & Navarro-Martinez, D. (2014a). On the measurement of strength of preference in units of money. *The Economic Record, 90*(s1), 1–15.
Butler, D., Isoni, A., Loomes, G., & Tsutsui, K. (2014b). Beyond choice: Investigating the sensitivity and validity of measures of strength of preference. *Experimental Economics, 17*(4), 537–563.
Butler, D., & Loomes, G. (2011). Imprecision as an account of violations of independence and betweenness. *Journal of Economic Behavior & Organization, 80*(3), 511–522.
Butler, D., & Loomes, G. (2007). Imprecision as an account of the preference reversal phenomenon. *American Economic Review, 97*(1), 277–297.
Békésy, G. (1947). A new audiometer. *Acta Oto-Laryngologica, 35*(5–6), 411–422.
Camerer, C. F. (1989). An experimental test of several generalized utility theories. *Journal of Risk and Uncertainty, 2*(1), 61–104.
Carrera, M., Royer, H., Stehr, M., Sydnor, J., & Taubinsky, D. (2022). Who chooses commitment? Evidence and welfare implications. *The Review of Economic Studies, 89*(3), 1205–1244.
Cerreia-Vioglio, S., Dillenberger, D., & Ortoleva, P. (2015). Cautious expected utility and the certainty effect. *Econometrica, 83*(2), 693–728.
Cerreia-Vioglio, S., Dillenberger, D., Ortoleva, P., & Riella, G. (2019). Deliberately stochastic. *American Economic Review, 109*(7), 2425–2445.
Cettolin, E., & Riedl, A. (2019). Revealed preferences under uncertainty: Incomplete preferences and preferences for randomization. *Journal of Economic Theory, 181*, 547–585.
Chew, S. (1989). Axiomatic utility theories with the betweenness property. *Annals of operations Research, 19*(1), 273–298.
Collins, S. M., & James, D. (2015). Response mode and stochastic choice together explain preference reversals. *Quantitative Economics, 6*(3), 825–856.
Cubitt, R. P., Navarro-Martinez, D., & Starmer, C. (2015). On preference imprecision. *Journal of Risk and Uncertainty, 50*(1), 1–34.
Dubourg, W. R., Jones-Lee, M. W., & Loomes, G. (1994). Imprecise preferences and the WTP-WTA disparity. *Journal of Risk and Uncertainty, 9*(2), 115–133.
Dubourg, W. R., Jones-Lee, M. W., & Loomes, G. (1997). Imprecise preferences and survey design in contingent valuation. *Economica, 64*(256), 681–702.
Dubra, J., Maccheroni, F., & Ok, E. A. (2004). Expected utility theory without the completeness axiom. *Journal of Economic Theory, 115*(1), 118–133.
Eliashberg, J., & Hauser, J. (1985). A measurement error approach for modeling consumer risk preference. *Management Science, 31*(1), 1–25.
Eliaz, K., & Ok, E. A. (2006). Indifference or indecisiveness? Choice-theoretic foundations of incomplete preferences. *Games and Economic Behavior, 56*(1), 61–86.
Feldman, P., & Rehbeck, J. (2022). Revealing a preference for mixtures: An experimental study of risk. *Quantitative Economics, 13*(2), 761–786.
Friedman, D., Habib, S., James, D., & Williams, B. (2022). Varieties of risk preference elicitation. *Games and Economic Behavior, 133*, 58–76.
Fudenberg, D., Iijima, R., & Strzalecki, T. (2014). *Stochastic choice and revealed perturbed utility*. Working paper, Massachusetts Institute of Technology.
Fudenberg, D., Iijima, R., & Strzalecki, T. (2015). Stochastic choice and revealed perturbed utility. *Econometrica, 83*(6), 2371–2409.
Gebhardt, W. A., & Brosschot, J. F. (2002). Desirability of control: Psychometric properties and relationships with locus of control, personality, coping, and mental and somatic complaints in three Dutch samples. *European Journal of Personality, 16*(6), 423–438.
Gul, F., & Pesendorfer, W. (2006). Random expected utility. *Econometrica, 74*(1), 121–146.

Halevy, Y., Walker-Jones, D., & Zrill, L. (2023). *Difficult decisions*. Working papers, University of Toronto, Department of Economics.
Harless, D., & Camerer, C. F. (1994). The predictive utility of generalized expected utility theories. *Econometrica, 62*(6), 1251–1289.
Hey, J., & Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica, 62*(6), 1291–1326.
Holt, C. A. (1986). Preference reversals and the independence axiom. *The American Economic Review, 76*(3), 508–515.
James, D. (2011). *Incentive compatible elicitation procedures*. Working papers, 19th International Congress on Modelling and Simulation, Perth, Australia, 1421-1427, Modelling and Simulation Society of Australia and New Zealand, Canberra, Australia.
Johnson, J. G., & Busemeyer, J. R. (2005). A dynamic, computational model of preference reversal phenomena. *Psychological Review, 112*(4), 841–861.
Karni, E., & Safra, Z. (1987). Preference reversal and the observability of preferences by experimental methods. *Econometrica, 55*(3), 675–685.
Lichtenstein, S., & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of experimental psychology, 89*(1), 46.
Loomes, G. (1988). Different experimental procedures for obtaining valuations of risky actions: Implications for utility theory. *Theory and Decision, 25*(1), 1–23.
Loomes, G. (2005). Modelling the stochastic component of behaviour in experiments: Some issues for the interpretation of data. *Experimental Economics, 8*(4), 301–323.
Loomes, G., & Pogrebna, G. (2017). Do preference reversals disappear when we allow for probabilistic choice? *Management Science, 63*(1), 166–184.
Loomes, G., Starmer, C., & Sugden, R. (1989). Preference reversal: Information-processing effect or rational non-transitive choice? *Economic Journal, 99*(395), 140–151.
Loomes, G., Starmer, C., & Sugden, R. (1991). Observing violations of transitivity by experimental methods. *Econometrica, 59*(2), 425–439.
Loomes, G., & Sugden, R. (1983). A rationale for preference reversal. *American Economic Review, 73*(3), 428–432.
Luce, R. D., & Suppes, P. (1965). *Preference, utility and subjective probability*. In R. D. Luce, Bush, R. R., & Galanter, E. H. (Eds.), Handbook of Mathematical Psychology, Vol. 3 (Vol. 1-3, Vol. 3, pp. 249-410). New York: Wiley.
MacCrimmon, K., & Smith, M. (1986). *Imprecise equivalences: Preference reversals in money and probability*. University of British Columbia working paper, 1211.
Machina, M. (1985). Stochastic choice functions generated from deterministic preferences over lotteries. *Economic Journal, 95*(379), 575–594.
McGranaghan, C., Nielsen, K., O'Donoghue, T., Somerville, J., & Sprenger, C. D. (2024). Distinguishing common ratio preferences from common ratio effects using paired valuation tasks. *American Economic Review, 114*(2), 307–347.
Mosteller, F., & Nogee, P. (1951). An Experimental Measurement of Utility. *Journal of Political Economy, 59*(5), 371–371.
Ok, E. A., Ortoleva, P., & Riella, G. (2012). Incomplete preferences under uncertainty: Indecisiveness in beliefs versus tastes. *Econometrica, 80*(4), 1791–1808.
Ong, Q., & Qiu, J. (2023). Paying for randomization and indecisiveness. *Journal of Risk and Uncertainty, 67*(1), 45–72.
Qiu, J., & Steiger, E.-M. (2011). Understanding the two components of risk attitudes: An experimental analysis. *Management Science, 57*(1), 193–199.
Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior and Organization, 3*(4), 323–343.
Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*(2), 59–108.
Rieskamp, J., Busemeyer, J. R., & Mellers, B. A. (2006). Extending the bounds of rationality: Evidence and theories of preferential choice. *Journal of Economic Literature, 44*(3), 631–661.
Segal, U. (1988). Does the preference reversal phenomenon necessarily contradict the independence axiom? *American Economic Review, 78*(1), 233–236.
Starmer, C., & Sugden, R. (1989). Probability and juxtaposition effects: An experimental investigation of the common ratio effect. *Journal of Risk and Uncertainty, 2*(2), 159–178.
Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review, 95*(3), 371–384.

Tversky, A., Slovic, P., & Kahneman, D. (1990). The causes of preference reversal. *The American Economic Review*, 80(1), 204–217.
van de Kuilen, G., & Wakker, P. P. (2011). The midweight method to measure attitudes toward risk and ambiguity. *Management Science, 57*(3), 582–598.