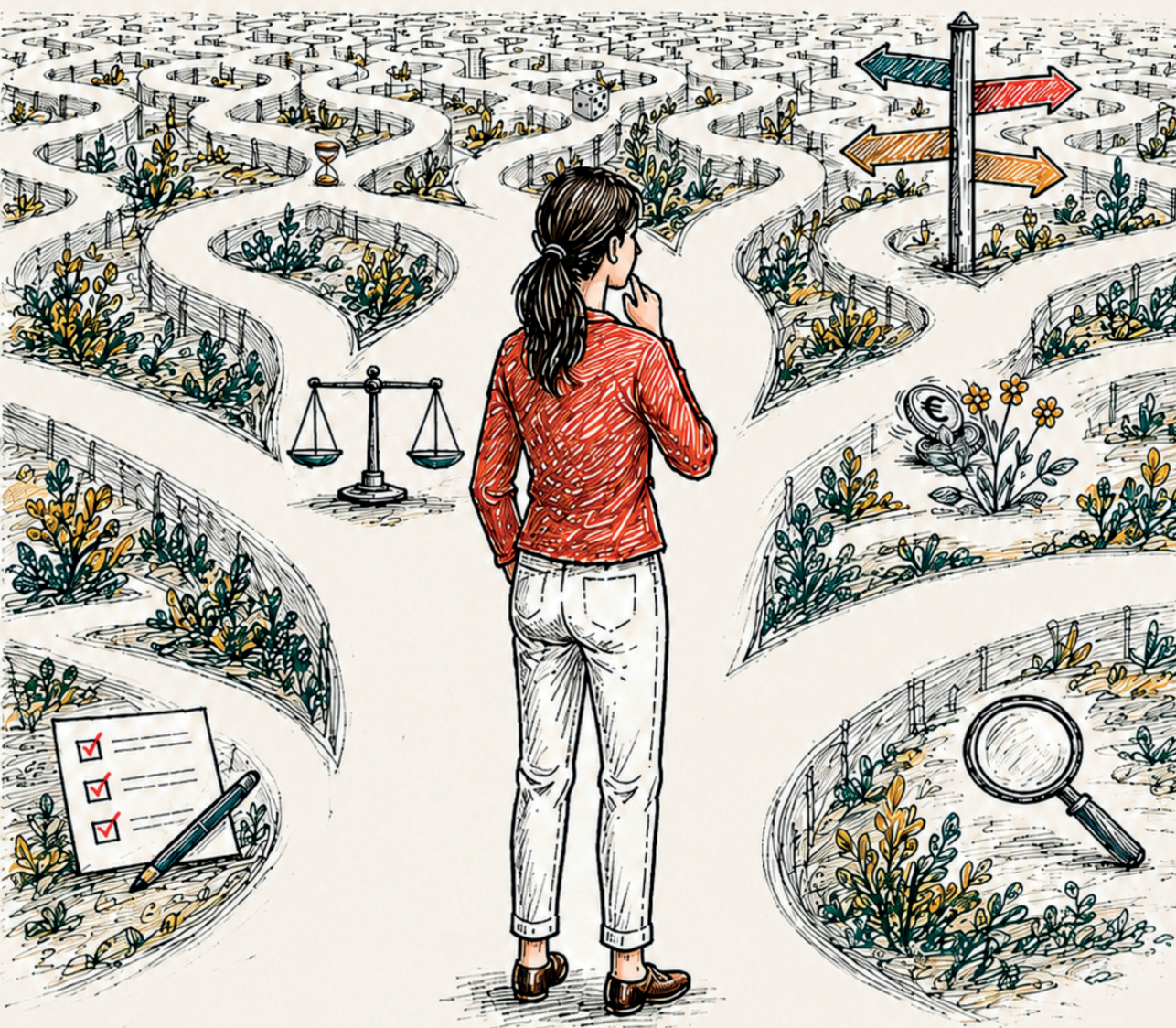


The role of **complexity**

in individual
decision making



Institute for
Management Research

Sara Arts

**RADBOD
UNIVERSITY
PRESS**

The role of complexity in individual decision making

Sara Arts

Sara Arts

The role of complexity in individual decision making

Radboud Dissertation Series

ISSN: 2950-2772 (Online); 2950-2780 (Print)

Published by RADBOUD UNIVERSITY PRESS

Postbus 9100, 6500 HA Nijmegen, The Netherlands

www.radbouduniversitypress.nl

Design: Sara Arts

Cover image: Ludo Arts

Printing: DPN Rikken/Pumbo

ISBN: 9789465152400

DOI: 10.54195/9789465152400

Free download at: <https://doi.org/10.54195/9789465152400>

© 2026 Sara Arts

**RADBOUD
UNIVERSITY
PRESS**

This is an Open Access book published under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives International license (CC BY-NC-ND 4.0). This license allows reusers to copy and distribute the material in any medium or format in unadapted form only, for noncommercial purposes only, and only so long as attribution is given to the creator; see <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

The role of complexity in individual decision making

Proefschrift ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.M. Sanders,
volgens besluit van het college voor promoties
in het openbaar te verdedigen op

maandag 28 september 2026
om 10.30 uur precies

door

Sara Dirkje Lisa Arts
geboren op 11 februari 1995
te Boxmeer

Promotor:

Prof. dr. S. M. Zeisberger

Copromotor:

Dr. J. Qiu

Manuscriptcommissie:

Prof. dr. E.A.J.A. Rouwette

Prof. dr. N. Jacquemet (Université Paris 1 Panthéon-Sorbonne, Frankrijk)

Prof. dr. Z. Li (Beijing Normal – Hong Kong Baptist University, China)

Contents

1	Introduction	1
1.1	Complexity in decision making	1
1.2	Overview and contribution	2
1.2.1	Overview	2
1.2.2	Contribution	4
2	Measuring decision confidence	7
2.1	Introduction	8
2.2	Experimental design	11
2.2.1	General structure of the experiments	11
2.2.2	Manipulating decision confidence	13
2.2.3	Design considerations	14
2.2.4	Sample and procedure	15
2.3	Theoretical analysis	16
2.4	Experimental results	17
2.4.1	Randomization probabilities and self-reported confidence	18
2.4.2	Manipulating decision confidence	20
2.4.3	Order effects	23
2.4.4	Alternative interpretations of randomization probabilities	24
2.5	Conclusion	25
3	The WTA-WTP gap and Complexity in a Representative Sample	27
3.1	Introduction	28
3.2	Experimental Design	32
3.2.1	Eliciting within-subject WTA and WTP	32
3.2.2	Measuring mechanism complexity	33
3.2.3	Measuring valuation complexity	34
3.2.4	Other measures	35
3.2.5	Sample and procedure	36
3.3	Experimental Results	36
3.3.1	Mechanism complexity	38
3.3.2	Valuation complexity	40
3.3.3	Controlling for mechanism and valuation complexity	42
3.3.4	Complexity versus preference-based explanations	47
3.3.5	Robustness checks	49
3.4	Concluding Remarks	50

4	Choice (in)consistency and household wealth	53
4.1	Introduction	54
4.2	Experimental sample, design, and data	58
4.2.1	Experimental sample	58
4.2.2	Experimental design and data	59
4.2.3	Household wealth and socioeconomic variables	63
4.3	Results	64
4.3.1	Relating MPI to demographic characteristics	64
4.3.2	The relationship between MPI and household wealth	66
4.3.3	Controlling for cognitive ability and risk attitudes	68
4.3.4	Self-reported data vs administrative records	69
4.3.5	Why does MPI relate to household wealth?	71
4.4	Conclusion	72
5	Conclusion and discussion	75
5.1	General conclusion	75
5.2	Implications	76
5.2.1	Methodological implications	76
5.2.2	Theoretical implications	77
5.2.3	Policy implications	78
5.3	Limitations and future research	78
	Bibliography	81
	Appendices to chapter 2	89
	Appendices to chapter 3	133
	Appendices to chapter 4	165
	Acknowledgements	185
	Nederlandse Samenvatting	187
	English Summary	189
	Research Data Management	191
	Curriculum Vitae	192

Introduction

Everyday individuals have to make countless decisions. Standard economics assumes that individuals are able to make optimal decisions based on the available information, budget constraints, and their preferences. This implies that individuals can fully understand the decision environment, and that their preferences are complete, stable, and well-defined. However, casual reflection and a vast amount of academic research suggests that in many situations individuals have difficulty to make optimal choices.

This difficulty can arise from various sources. For instance, it could be that the decision options are complex to compare due to trade-offs between multiple, conflicting objectives such as risk versus return, fairness versus efficiency, etc. It could also be that the decision environment is complex, making individuals uncertain about the impact different decision options have on the desired outcomes. These forms of complexity and the uncertainty it leads to can have significant impact on the quality of decisions, and in turn on the consequences of the decisions later on. In this dissertation we investigate the measurement of decision complexity individuals experience, the different forms of complexity in decision-making, and the potential consequences of experiencing decision complexity.

There are various ways to conduct such an investigation. In the chapters that follow we will integrate theoretical analysis with experimental investigation. Theoretically, we propose conceptual measures of decision complexity and categorize different forms of complexity that individuals perceive when making their decisions. Experimentally, we develop and validate our measures of decision complexity, link it to established experimental observations, and provide evidence of the impact of complexity on decision-making and its real-life consequences.

1.1 Complexity in decision making

Complexity in decision-making has received more and more attention in recent years and it can take various forms. We focus on two forms of complexity: The first considers the decision context. Specifically, the elicitation mechanism that is used to infer preferences from choices, an important experimental method in economic analyses (Plott and Zeiler, 2005, 2007; Cason and Plott, 2014; Smitizsky et al., 2021). Understanding the elicitation mechanism takes cognitive capacity and effort. If individuals are unfamiliar with a (complex) elicitation mechanism, they may fail to understand the task or confuse it with another mechanism. For example, the literature suggests that many individuals have difficulty to understand the second-price auction and confuse it with the first-price auction.

The second form of complexity concerns the difficulty individuals perceive when it comes to their own preferences. Individuals may find it difficult to come to an exact valuation of the choice options. As a result of this difficulty, individuals may, for example, choose

stochastically (Agranov and Ortoleva, 2017; Cettolin and Riedl, 2019; Halevy et al., 2023) or postpone their decision (Danan and Ziegelmeyer, 2006). This form of complexity is closely related to the notions of incomplete preferences (Bewley, 2002; Masatlioglu and Ok, 2005; Eliaz and Ok, 2006; Ok et al., 2012; Nishimura, 2018), preference imprecision (Dubourg et al., 1994; Butler and Loomes, 2007; Cubitt et al., 2015), and cognitive noise (Khaw et al., 2021; Woodford, 2020; Frydman and Jin, 2021; Enke and Graeber, 2023; Oprea, 2024).

When experiencing complexity, individuals may rely on heuristics to form decisions (Simon, 1955; Kahneman and Tversky, 1974; Gigerenzer et al., 1999; Ortoleva, 2010; Iyengar and Kamenica, 2010). Heuristics are cognitive shortcuts that reduce the mental effort required to make a decision. For instance, individuals may anchor their choices on salient values or defaults (Tversky and Kahneman, 1974; Johnson and Goldstein, 2003), adopt a satisficing strategy, selecting the first option that meets a minimum threshold of acceptability rather than optimizing across all available alternatives (Simon, 1955), or defer to the status quo option or avoid making a choice altogether (Samuelson and Zeckhauser, 1988).

While the use of heuristics can be adaptive in certain decision situations, it can also lead to decision-making that has systematic negative consequences. For instance, choice inertia can keep individuals from switching to potentially beneficial alternatives when making portfolio choices (Sautua, 2017). Acting overly cautiously in decisions can offer protection from potential unknown risks, but can lead to hedging excessively and overpaying for insurance (Rabin and Bazerman, 2019). Other heuristics such as deferring decisions, delegating decisions to experts, or focusing on certain option attributes while ignoring others can have similar negative long-term consequences while offering a way to minimize uncertainty or complexity in the moment.

1.2 Overview and contribution

1.2.1 Overview

The first step in understanding the role of complexity in decision making is being able to elicit and quantify the extent to which individuals experience complexity. In Chapter two we focus on measuring the second form of complexity, the difficulty about the valuation of the choice options. One manifestation of the decision difficulty is the lack of decision confidence when individuals make decisions. The dominant approach in the literature is the self-reported decision confidence (Cohen et al., 1987; Dubourg et al., 1994, 1997; Butler and Loomes, 2007, 2011; Cubitt et al., 2015; Enke et al., 2025; Enke and Graeber, 2023). In contrast to this, we obtain a quantitative and incentivized measure of decision confidence by relating the confidence individuals have in their decisions to their tendency to randomize between the choice options in the decisions.

In two experiments, participants faced pairs of options and were asked to choose their preferred one. We obtained two measures of self-reported decision confidence (Butler and Loomes, 2007; Enke and Graeber, 2023). Additionally, we allowed participants to

assign randomization probabilities to each choice option, building on the literature on preferences for randomization (Rubinstein, 2002; Dwenger et al., 2018; Miao and Zhong, 2018; Feldman and Rehbeck, 2022; Agranov and Ortoleva, 2023; Ong and Qiu, 2023). As a further test, we exogenously manipulated decision confidence by varying the complexity of the choice options and the experience subjects gained with the choice before making their final decision.

We found a strong and positive correlation between randomization probabilities and self-reported decision confidence. The exogenous manipulations affected both self-reported decision confidence and randomization probabilities in similar ways. These results indicate that randomization probabilities are an appropriate measure of decision confidence.

The next step is understanding to what extent each form of complexity affects individuals' decision making. In Chapter three we quantify both forms of complexity in an experiment on the willingness-to-accept and willingness-to-pay gap (WTA-WTP gap; Thaler, 1980; Knetsch, 1989; Kahneman et al., 1990). The WTA-WTP gap has been a widely researched phenomenon. And while loss aversion has always been the main explanation, different alternative theoretical and empirical explanations have been provided (Ericson and Fuster, 2014). Both complexity surrounding the elicitation mechanism (Loomes et al., 2003; Plott and Zeiler, 2005, 2007, 2011; Isoni et al., 2011; Cason and Plott, 2014) and complexity around the valuation of choice options (Dubourg et al., 1994; Bayrak and Hey, 2020; McGranaghan and Otto, 2022) have been suggested as (part of) the source of the gap.

In an experiment with a representative sample of the Dutch population, we elicited the WTA and WTP of participants for two monetary lotteries using a price list mechanism. Additionally, we obtained measures of each form of complexity. We found that mechanism complexity explains about 40% of the observed WTA-WTP gap, valuation complexity about 15%, and together both forms of complexity can account for about 50% of gap. And while these results holds across the sample, the extent to which complexity affects the WTA-WTP gap across demographics are largest for participants with lower cognitive abilities. These results suggest that decision complexity could be an important factor in phenomena that are often considered 'irrational'.

Finally, in Chapter four we focus on the real-life consequences of decision complexity observed in experimental settings. In this chapter, we take choice inconsistency as a measure of decision complexity. While inconsistent choices are very commonly observed in economic experiments, it is less clear to what extend they translate to real-life decision making. We link choice inconsistency to real-life economic outcomes through the idea of exploitation by a third party as seen in "Dutch books" (Ramsey, 1931; Raiffa, 1968; De Finetti, 1974; Eells, 1982; Yaari, 1985; Green, 1987; Kelsey, 1995). Specifically, we compute a money pump index (MPI; Echenique et al., 2011) based on the decisions participants made in an experiment. In this case we aimed to minimize the complexity related to the experimental elicitation mechanism, by making the choices simple binary decisions between a monetary lottery, a certain amount, or letting the computer randomly choosing either option. We

combined the experimental measures with administrative data from Statistics Netherlands (CBS) on wealth and income.

The results show that the MPI is related to our main real-life outcome of household wealth. In our baseline regression model we find that one standard deviation increase in MPI is associated with a 14.5% decrease in household wealth when controlling for socioeconomic characteristics. This relationship remains when including various controls, such as cognitive ability or risk attitudes, and in various subgroups of our sample. Together the findings from this chapter suggest that the decision making observed in an artificial and relatively abstract setting could be insightful when it comes to real-life economic behavior.

1.2.2 Contribution

This dissertation contributes to the growing work on complexity in decision-making in several ways. Chapter two makes a conceptual and methodological contribution to the study of decision-making under complexity by introducing an incentivized, behavioral measure of decision confidence based on subjects' chosen randomization probabilities. The measure is simple to elicit within standard binary choice tasks, incentive-compatible, and yields both ordinal and cardinal information about confidence. Conceptually, the paper clarifies how deliberate randomization can arise as a hedge under preference uncertainty and links this behavior to the literature on incomplete/imprecise preferences, cognitive uncertainty, and stochastic choice. While theoretical and empirical advancements have been made separately in this field, our findings provide a direct link between these concepts.

Chapter three builds on this by combining the literature on the two sources of decision complexity. It contributes to the vast amount of studies trying to understand the WTA-WTP gap and shows the role of decision complexity in this well-known behavioral anomaly. The main contribution of this chapter is quantifying the potential role of decision complexity. While previous studies have shown the presence of complexity in decision making, our experimental method allows us to show how substantial the role of each form of complexity is. Furthermore, the use of a large representative sample contributes to the external validity of the findings and allows us to differentiate the effects of complexity on behavior among different demographic groups within our sample.

Finally, the findings in Chapter four further suggests that decision complexity has real-life consequences, contributing to the demonstration of the external validity of our findings. By linking an experimental measure of decision complexity to real-life economic outcomes such as household wealth across a representative sample, we show that differences in experimentally observed behavior corresponds to differences in financial outcomes. While the link between experimental and real-life data has been made before, the quantitative measure of choice inconsistency in combination with the administrative records from Statistics Netherlands (CBS) allows us to draw more robust conclusions than previous studies that have relied mostly on categorical measures of inconsistency and self-reported outcomes.

To summarize, by focusing on all facets of the decision-making process, from theoretical analyses to experimental investigation, from lab behavior to real-life consequences, we provide a broad overview of the role that complexity plays in decision-making. After the chapters outlined in the previous paragraphs, the implications, limitations, and directions for future research on decision complexity are discussed in Chapter five.

Measuring decision confidence

Abstract: We examine whether the way individuals randomize between options captures their decision confidence. In two experiments in which subjects faced pairs of options (a lottery and a varying sure payment), we allowed subjects to choose randomization probabilities according to which they would receive each option. Separately, we obtained two measures of self-reported confidence - confidence statements and probabilistic confidence - for choosing between the two options. Consistent with the predictions of two theoretical frameworks incorporating preference uncertainty, the randomization probabilities correlated strongly with both self-reported measures (median Spearman correlations between 0.86 to 0.89) and corresponded in absolute levels to probabilistic confidence. This relationship is robust to two exogenous manipulations of decision confidence, where we varied the complexity of the lottery and subjects' experience with the lottery.

This chapter has been published as: Arts S, Ong Q, Qiu J. Measuring decision confidence. *Experimental Economics*. 2024; 27(3):582-603. doi:10.1007/s10683-024-09837-x

2.1 Introduction

There are many decisions in life that people may not be able to make with full confidence. These decisions often involve difficult trade-offs among conflicting objectives, such as price vs. quality when buying goods, risk vs. return when investing, and efficiency vs. equality when making policy decisions. As more studies suggest that decision confidence has the potential to explain a wide range of behavioral anomalies, there is growing interest in eliciting and accounting for decision confidence when studying people's choices.¹

Past studies so far have mostly relied on non-incentivized self-reports to elicit decision confidence. For example, some studies (Cohen et al., 1987; Dubourg et al., 1994, 1997; Cubitt et al., 2015) allowed subjects to indicate whether they were unsure of their choices. Butler and Loomes (2007, 2011) had subjects indicate their decision confidence using the ordinal terms “definitely” and “probably.” More quantitatively, Enke et al. (2025); Enke and Graeber (2023) had subjects rate how certain (from 0% to 100% in increments of 5%) they were that their actual valuation of an option was within the interval obtained from the choices they made earlier in a price list.

While asking people explicitly about how confident they are about their decisions directly elicits decision confidence, incentivized behavioral measures that elicit decision confidence without referring to confidence may encourage people to contemplate their decisions carefully and thereby reduce behavioral biases such as priming effects and experimenter demand effects (Camerer and Hogarth, 1999). Yet, finding behavioral measures of decision confidence “in a form simple and transparent enough to work without creating additional uncertainty” is not trivial (Butler and Loomes, 2011, p. 516).

Building on earlier studies, we propose to use the randomization probability assigned to an option in a choice pair as a behavioral measure of decision confidence, and test its validity in two experiments. For each pair of options (a lottery x and a sure payment y), we obtained subjects' binary choice, self-reported decision confidence, and randomization probabilities. Self-reported decision confidence was elicited by having subjects select a confidence statement from “Surely x ,” “Probably x ,” “Unsure,” “Probably y ,” “Surely y ,” after they chose an option and by having them report how confident they were in choosing either options as a probabilistic confidence of $p\%$ x and $100-p\%$ y , where p ranged from 0 to 40 and 60 to 100 in increments of 10. Separately, subjects chose a randomization probability $0 \leq \lambda \leq 1$ with which they would receive x (and with probability $1 - \lambda$ receive y) for each pair of options (Miao and Zhong, 2018; Agranov and Ortoleva, 2023; Feldman and Rehbeck, 2022; Ong and Qiu, 2023). Unlike the two self-reported confidence measures, the elicitation of randomization probabilities made no reference to decision confidence.

¹These anomalies include the willingness to accept (WTA) - willingness to pay (WTP) gap (Dubourg et al., 1994), preference reversals (Butler and Loomes, 2007), stochastic choices (Agranov and Ortoleva, 2017), insensitivity to variation in probabilities (Enke and Graeber, 2023), anomalies in intertemporal choices (Enke et al., 2025), small-stakes risk aversion (Khaw et al., 2021), and many other violations of standard decision theory (Butler and Loomes, 2011).

We exogenously manipulated decision confidence by a) having a simple lottery with two outcomes and a complex lottery with the same expected value but with more payoff outcomes over a wider range of possible values (Fudenberg and Puri, 2021), and by b) increasing subjects' experience with the lottery by allowing them to either observe the outcome draws of the lottery or to make hypothetical choices and observe the payoffs of their choices and the counterfactual (Myagkov and Plott, 1997; Plott and Zeiler, 2005; van de Kuilen and Wakker, 2006; van de Kuilen, 2009).

To ensure that experimenter demand effects and order effects were not driving our results, in one of our experiments, we elicited the three measures separately over three sessions (at least seven days apart). Subjects were randomly assigned to one of the three decision orders, which differed by whether the confidence statements, the probabilistic confidence, or the randomization probabilities were elicited in the first session.

We structure the analyses of our experiments through two theoretical frameworks, one based on Klibanoff et al. (2005) and Cerreia-Vioglio et al. (2015) and the other based on Fudenberg et al. (2015). We illustrate how randomization emerges from the optimization behavior of an individual who faces uncertainty regarding her preference between the two options. Based on our expectations that randomization probabilities serve as a good proxy for decision confidence, and that complexity increases preference uncertainty while experience decreases it, we have four hypotheses. First, randomization probabilities correlate positively with both self-reported measures; second, subjects choose randomization probabilities around 0.5 for choice pairs in which the sure payment has a similar decision utility as the lottery (and one does not dominate the other); third, subjects randomize over a wider range of sure payments and choose randomization probabilities closer to 0.5 for the complex lottery than for the simple lottery; finally, subjects randomize over a smaller range of sure payments and choose randomization probabilities further away from 0.5 in the experience treatments.

Our experimental results support all four hypotheses. Subjects' randomization probabilities were strongly and positively correlated with both confidence statements and probabilistic confidence (median Spearman correlation between 0.86 to 0.89). In line with our expectations, the two exogenous manipulations affected self-reported decision confidence. Increasing the complexity of the lottery led to a decrease in self-reported decision confidence, while increasing experience with lotteries led to an increase in self-reported decision confidence. These exogenous changes in self-reported decision confidence were met with corresponding changes in randomization probabilities: subjects randomized over a larger range of sure payments and the randomization probabilities were closer to 0.5 for the complex lottery than for the simple lottery, while the opposite occurred when subjects had more experience. As a result, the correlations between randomization probabilities and self-reported decision confidence measures were robust to the exogenous manipulations of decision confidence.

Our study builds on the growing literature on preferences for randomization, implying preference functionals that are convex with respect to probabilistic mixing, which is a

violation of the betweenness axiom (Chew, 1983; Dekel, 1986; Chew, 1989). Preferences for randomization have been documented over wide ranges, across different domains, in experimental settings as well as in real life decisions. In a multiple-decision setting, Rubinstein (2002) suggested that randomization (diversification in his term) by choosing differently across five independent and identical decisions is “an expression of a more general phenomenon in which people tend to diversify their choices when they face a sequence of similar decision problems and are uncertain about the right action” (Rubinstein, 2002, p.1370). Dwenger et al. (2018) found that their experimental subjects preferred to randomize via an external randomization device rather than making choices themselves, and the authors reported similar behavior among German university applicants. Miao and Zhong (2018) showed that randomization could be used to balance ex-ante and ex-post social preferences. Feldman and Rehbeck (2022) elicited individuals’ attitudes toward reduced mixtures over two lotteries in the space of three-outcome lotteries (the Marschak-Machina triangle) and found pervasive evidence of a preference for non-degenerate mixing over lotteries. The studies closest to ours are Agranov and Ortoleva (2023) and Ong and Qiu (2023), who also allowed subjects to choose randomization probabilities when deciding between two options. Both studies found that subjects often randomized and did so over large ranges. Ong and Qiu (2023) further found that subjects were willing to pay to randomize, suggesting that randomization was deliberate and not merely a result of indifference. Popular explanations for convex preferences include hedging in the face of preference uncertainty (Cerreia-Vioglio et al., 2015; Fudenberg et al., 2015; Cerreia-Vioglio et al., 2019), non-linear probability weighting (Kahneman and Tversky, 1979; Quiggin, 1982; Tversky and Kahneman, 1992), and responsibility aversion (Dwenger et al., 2014). Our study is the first to provide experimental evidence linking preference uncertainty and randomization behavior.²

We also contribute to the literature on stochastic choice, which examines why individuals change their decisions when they face the same decision situation repeatedly. The relationship we found between randomization probabilities and sure payments bears a remarkable resemblance to results reported in studies on stochastic choice, for example, Mosteller and Nogee (1951, Figure 2) and Loomes and Pogrebna (2017, Table 1).³ The similarity between the choice proportion in repeated choices and the randomization probability in a one-shot decision suggests that decision confidence may have the potential to explain stochastic choices. Consistent with this interpretation, we find that, across subjects and decisions, higher decision confidence in an option corresponded to choosing that option more fre-

²Agranov and Ortoleva (2023) also explicitly discuss this link. Based on reports from the end-of-experiment questionnaire, they found that many of their subjects randomized because they were unsure of their preferences (Agranov and Ortoleva, 2023, Appendix A.8).

³Note that these results come from entirely different designs. In Mosteller and Nogee (1951) and Loomes and Pogrebna (2017) individuals repeatedly faced a lottery and a sure payment, with the sure payment varying from one question to another, and the results are about the proportion of accepting the lottery across decisions, whereas in our experiment subjects faced the lottery and a sure payment once and chose the randomization probability of receiving the lottery.

quently (but not always) in binary choices. Meanwhile, random (expected) utility models (see, e.g., Eliashberg and Hauser, 1985; Loomes and Sugden, 1995; Gul and Pesendorfer, 2006; Apesteguia and Ballester, 2018), which are the standard explanations for stochastic choices, do not predict randomization in a one-shot decision as observed in our experiment. This is because, while individuals may be considered to have a set of utility functions in this literature, at the moment of decision-making, they rely on one utility function randomly realized from the set.

Overall, our study provides direct evidence to the connections between some important concepts in the literature, such as decision confidence, cognitive uncertainty, preference uncertainty/imprecision, incomplete preference, preference for randomization, and stochastic choice. While there have been notable theoretical advancements and accumulating empirical evidence in this field, the precise interplay and relationships between these concepts as well as how they are connected to choices remain ambiguous. Our finding of a systematic relationship between the randomization, alternative measures of decision confidence, and stochastic choice suggests that there may exist a common psychological underpinning for these various concepts.

The rest of the paper proceeds as follows. Section 2.2 describes the experimental procedure. Section 2.3 provides the theoretical basis for how randomization probabilities may be linked to decision confidence. The results are reported in Section 2.4. Finally, Section 2.5 concludes the paper.

2.2 Experimental design

We had two experiments. We first describe the general structure of the experiments before detailing the differences.

2.2.1 General structure of the experiments

In each decision, subjects faced a pair of options: a lottery x and a sure payment y . The lottery was paired with 13 values of sure payments (0, 2, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 8, and 10 euros). For each type of decisions that we will describe below, subjects faced these pairs in a random sequence. Each decision was made on a separate screen, and subjects were not allowed to review or change their decisions once they were made. Each subject made three types of decisions: binary choices, self-reported decision confidence, and randomized choices.

Binary choices

The binary choices required subjects to choose either x or y . If x was chosen, the computer would draw a random number to determine x 's outcome. For example, for x that has a 50% chance of paying 9 euro and a 50% chance of paying 1 euro, if the randomly drawn number falls between 1 and 50, the subject would receive 1 euro, and if the randomly drawn number falls between 51 and 100, the subject would receive 9 euros.

Option x:
Gain €9 with a chance of **50%**, and
gain €1 with a chance of **50%**.

Option y:
Gain €4 for sure.

Please move the slider to determine the chance according to which you want to receive option x and option y.

100% x  100% y

You will be paid according to **Option x** with a chance of: **50%**
 You will be paid according to **Option y** with a chance of: **50%**

Figure 1: An example of the randomized choice decision screen, where option x is a lottery with a 50% chance of gaining 9 euro and a 50% chance of gaining 1 euro. Option y is a sure payment and varies across choices.

Two measures of self-reported decision confidence

After making the binary choices, we asked subjects how confident they felt about their choices. The confidence statements they could choose were “Surely x,” “Probably x,” “Unsure,” “Probably y,” or “Surely y.” Similar statements were used in Dubourg et al. (1994), Butler and Loomes (2007), and Butler and Loomes (2011). Confidence statements were not incentivized and could not affect payoffs.

In addition to the confidence statements, subjects in Experiment 2 also had to report their probabilistic confidence in a separate experimental decision. Instead of making a direct binary choice, subjects had to choose how confident they felt about choosing x versus y. They had to choose between ten levels of probabilistic confidence: “100% x, 0% y,” “90% x, 10% y,” ... “60% x, 40% y,” “40% x, 60% y,” ..., “0% x, 100% y.” Subjects were considered to have chosen the option for which they indicated more than 50% probabilistic confidence. For example, if a subject chose “60% x, 40% y,” she was considered to have chosen x over y in that decision. To use the probabilistic confidence as a measure of decision confidence as well as an indicator of a subject’s choice between x and y, we omitted the option “50% x, 50% y”.

Randomized choices

The randomized choices required subjects to choose a randomization probability λ , based on which they would receive x (and hence with a probability $1 - \lambda$ of receiving y). For example, choosing $\lambda = 0.40$ means the subject would receive x with a chance of 40% and y with a chance of 60%. The subjects used a slider from 0% to 100% with increments of 1% to choose the randomization probability in each choice. In experiment 1, the slider was set in the middle at the start. To reduce anchoring bias, in experiment 2, the slider had no initial position, and subjects needed to click on the slider and move the bar to determine the randomization probability. If the randomized choice was chosen for payment, the computer would draw a random number between 1 and 100. If the drawn number was between 1 and

100%, x would be chosen over y in that decision. If x was chosen for payment, a second random draw determined the outcome of the lottery. Figure 1 shows the decision screen for the randomized choice. To ensure that subjects understand the payoff structure of randomized choices, we provided two examples as well as reminders in the lower part of the decision screens on how randomized choices affect payment (see Figure C.4 and C.5 in Appendix C).

2.2.2 Manipulating decision confidence

In the baseline treatment, the subjects faced a simple lottery with two outcomes (a 50% chance of 9 euro and a 50% chance of 1 euro). They received a complete description of the lottery before making their decisions.

Varying the complexity of the lottery: We manipulated, within subject, the complexity of the lottery by asking all subjects to make decisions involving a complex lottery with four outcomes in addition to decisions involving the simple lottery. The complex lottery has the same expected value as the simple lottery. It offers 9.75 euros with a chance of 20%, 7.50 euros with a chance of 30%, 2.50 euros with a chance of 30%, and 0.25 euros with a chance of 20%. The order of the lotteries was randomized at individual subject level: some subjects proceeded from the simple lottery to the complex lottery, while others completed the decisions in the reverse order.⁴

Varying subjects' experience with the lottery: For this manipulation, we adopted a between-subject design, where subjects were randomly assigned to either the baseline treatment or the experience treatment. In Experiment 1, after learning the probability distribution of the lottery and prior to making actual decisions, subjects in the partial-experience treatment had to click and view 20 draws of the lottery. As the subjects viewed each lottery draw, an accompanying bar chart which recorded each lottery outcome was updated. By the 20th draw, the bar chart reflected the probability distribution of the lottery. Figure 2(a) shows an example of the partial-experience treatment.

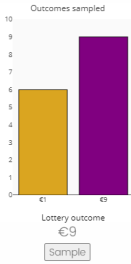
In Experiment 2, subjects in the full-experience treatment had to make five hypothetical decisions per lottery, with each decision involving a different sure payment (3, 4, 5, 6 or 7 euros in a random sequence). For each decision, they observed four potential realizations of the payoff of their choice as well as the counterfactual in a payoff table. Figure 2(b) shows the decision screen and the payoff table viewed by a subject who chose the lottery over the sure payment of 4 euros in the hypothetical decision.

⁴The subjects in Experiment 1 also made decisions involving a loss lottery and a mixed lottery. We included these lotteries because they might lead to larger preference uncertainty due to the additional uncertainty in attitudes toward losses. We omitted these two lotteries in Experiment 2 because, as pointed out by one reviewer, the theoretical analysis of these two lotteries requires a more general approach than what we had relied on. The main results involving these lotteries are provided in Appendix B.4. Further details can be found in the working paper version (Arts et al., 2020).

Before you are asked to make the decisions we want to give you the opportunity to **experience the different outcomes of option A**. For this, you can click the button below. Each time you click the button a possible outcome of option A will be shown. You will get to sample **20** outcomes.

The outcomes that you obtain by clicking the button do not influence your payoff but are only presented to make you experience the possible outcomes.

To keep track of the sampled outcomes, they will be presented in a bargraph.



(a) The partial-experience treatment

Option A:
Receive €9 with a chance of **50%**, and
Receive €1 with a chance of **50%**.

Option B:
 Receive a **€4** for sure.

Please indicate which option you chose. After you made your choice you can see the outcomes of your decision by clicking on the trial buttons. This allows you to experience the possible consequences of your decision. The outcomes of the option you selected are highlighted in the table.

	Trial 1	Trial 2	Trial 3	Trial 4
Choose Option A	€1	€9	€9	€1
Choose Option B	€4	€4	€4	€4

(b) The full-experience treatment

Figure 2: Panel (a) illustrates what subjects in the partial-experience treatment saw when they generated the outcomes of the lottery. Panel (b) illustrates what subjects saw in the full-experience treatment. The numbers highlighted in blue in the table show a subject's hypothetical decision and her four potential payoffs, and the not highlighted numbers show the counterfactuals.

2.2.3 Design considerations

A few design features deserve some discussion. First, our within-subject design of eliciting self-reported confidence measures and randomization probabilities for each subject raises the concern that experimenter demand effects or order effects may unintentionally influence subjects to give similar responses, resulting in a systematic relationship between them. We took several measures in Experiment 2 to make it more obscure and costly for subjects to connect self-reported confidence measures and randomization probabilities (Zizzo, 2010), such as spreading the decisions over three sessions (at least seven days apart), including a cost for randomizing, and randomly assigning subjects to one of the three different orders. The key features of the two experiments are summarized in Figure 3.

Second, varying the complexity of the lottery could induce changes in behavior through channels other than decision confidence. For example, subjects may value the two lotteries differently, as found in studies documenting complexity seeking and complexity averse behaviors (see e.g. Abdellaoui et al., 2020, and the references therein). This difference is less relevant for our purpose because our focus is on decision confidence (e.g., the range of sure payments that subjects do not have full confidence) rather than the average valuation of the lotteries. Another concern is that varying the complexity of the lottery induces different randomization behavior. While this may occur in some non-EUT models (Machina, 1985; Kahneman and Tversky, 1979; Tversky and Kahneman, 1992; Quiggin, 1982), they do not

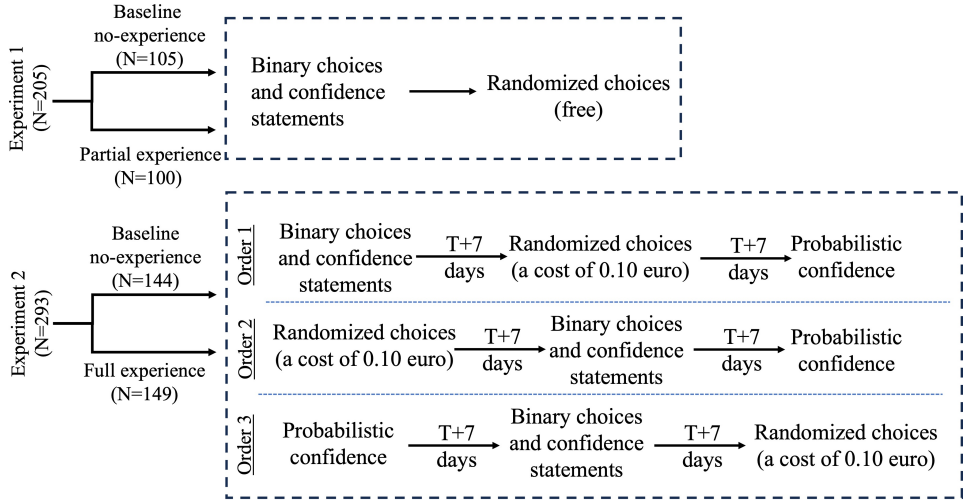


Figure 3: Summary of the treatments and experimental procedure in Experiment 1 and 2. The dotted rectangle highlights the types of decisions that subjects in each experiment made for both lotteries. The sequence of the simple lottery and the complex lottery in each type of decision was randomized at individual subject level.

predict a close relationship between randomization probability and self-reported decision confidence, or how this relationship changes with the complexity of the lottery. We will return to this point in subsection 2.4.4 to discuss other interpretations of randomization probabilities that could predict different randomization behavior between the two lotteries.

A final concern is the choice of the experience treatments. The description-experience gap literature suggests that description and experience may induce different risk preferences (see e.g., Hertwig et al., 2004; Wulff et al., 2018), which could reduce decision confidence. Our partial experience design is unlikely to have this issue because subjects' experience of the lottery realizations were equivalent to its description. Recent studies by Aydogan and Gao (2020) and Cubitt et al. (2022) show that the description-experience gap should be small in this case. Examining more closely, our experience treatments differ from the standard design in the Description-Experience gap literature as we provided subjects with the full description of the lottery's probability distribution in addition to the opportunity to experience the realizations of the lottery. In this sense, our experience treatments are closer to studies showing that experience in addition to a full description of the lottery could help subjects develop a better understanding about their preference (e.g., van de Kuilen and Wakker, 2006; van de Kuilen, 2009).

2.2.4 Sample and procedure

The data were collected from a sample of 498 subjects of the ID lab at Radboud University. A total of 205 subjects participated in Experiment 1 and 293 in Experiment 2. Invitations

were sent in batches via ORSEE (Greiner, 2015). The experiment was conducted using Qualtrics and lasted approximately 20 minutes for Experiment 1 and about 30 minutes in total for Experiment 2. In Experiment 2, subjects were also asked to answer a post-experiment questionnaire at the end of each of the three sessions based on the type of decision confidence they reported in that session. Appendix C contains the experimental instructions and decision screens. Each subject received a participation fee of 1 euro and an additional payment based on one of the decisions they made in the experiment. In Experiment 1, the additional monetary compensation was based on a decision randomly selected from their binary choices or randomized choices. In Experiment 2, it was based on a decision randomly selected from their binary choices, randomized choices, or probabilistic confidence decisions. The average additional payment was 6.28 euros. We made the payment via bank transfers.

2.3 Theoretical analysis

Under expected utility theory (EUT) which ignores decision confidence and assumes that a unique utility function (subject to positive affine transformation) captures an individual's preference, it is straightforward to show that the individual chooses $\lambda^* \in (0, 1)$ for at most one value of the sure payment in the 13 choice pairs. Thus, under EUT, strict randomization ($\lambda^* \in (0, 1)$) rarely occurs, and randomization probabilities do not contain additional information beyond indifference.

We present two theoretical analyses of our experiments that provide an explicit link between randomization probabilities and decision confidence. Both analyses assume that the individual is uncertain about her preference. Appendix A.1 presents a theoretical framework based on Klibanoff et al. (2005) and Cerreia-Vioglio et al. (2015). In this framework, the individual has multiple utility functions that we call multiple selves. She is not fully confident about her choice when some selves prefer one option while others prefer the other. In such instances, the individual prefers randomization over selecting a particular option because it offers a "fair" way to resolve internal conflicts among her different selves. This approach of capturing the lack of decision confidence from unsureness about preferences is closely related but is different from models of ambiguity (e.g., Gilboa and Schmeidler, 1989; Klibanoff et al., 2005), which focus on unsureness about beliefs (e.g., Halevy, 2007; Chew et al., 2017; Cubitt et al., 2020, and the references therein).

Appendix A.2 presents the extension of Fudenberg et al.'s (2015) model. Fudenberg et al. (2015) axiomatized a choice rule of deliberate randomization called additive perturbed utility (APU). Their representation corresponds to a form of ambiguity-averse preferences for an individual who is uncertain about her true utility function. The individual randomizes to balance the probability of errors due to preference uncertainty against the cost of avoiding them (Fudenberg et al., 2015, p. 2373).

Both analyses suggest that the preference for randomization is motivated by the hedging of preference uncertainty. In particular, randomization probabilities are affected by the

perceived preference uncertainty of the options, attitudes towards preference uncertainty, as well as the utility difference between the options. The two theoretical analyses suggest that our proposed link between randomization probabilities and decision confidence could hold under a broad class of decision models that incorporate preference uncertainty.

We expect that subjects perceive more preference uncertainty with the complex lottery than with the simple lottery, and experience with the lottery reduces preference uncertainty regarding the lottery. With these expectations, our theoretical analyses show that randomization probabilities share three important properties of decision confidence: a) subjects choose randomization probabilities close to 0.5 when the sure payment has a similar decision utility to the lottery; b) they randomize over a wider range of sure payments, with randomization probabilities closer to 0.5, when they face the complex lottery compared to the simple lottery; and c) with experience and less preference uncertainty about the lottery, subjects' randomization probabilities may be stretched away from 0.5 as they randomize over a smaller range of sure payments. If randomization probabilities and the two self-report measures both capture decision confidence, we expect the following:

Hypotheses

1. *Randomization probabilities are positively correlated with the self-reported confidence measures.*
2. *When two choice options are more similar, for example, around the switching choices where subjects switch between the lottery and the sure payment, the subjects have lower decision confidence, a higher likelihood of randomizing, and randomization probabilities closer to 0.5.*
3. *Compared to the decisions about the simple lottery, the decisions about the complex lottery exhibit lower decision confidence, as measured by the self-reported confidence measures, and randomization probabilities are affected in the same direction, maintaining a strong association between them.*
4. *Compared to the decisions in the no-experience treatment, the decisions in the experience treatments exhibit higher decision confidence, as measured by the self-reported confidence measures, and randomization probabilities are affected in the same direction, maintaining a strong association between them.*

2.4 Experimental results

We report the results in two steps. We begin by showing the systematic link between randomization probabilities and the two measures of self-reported confidence in the baseline no-experience treatment for decisions about the simple lottery (Hypotheses 1 and 2). We then show that decision confidence responded to our treatment manipulations in the expected direction by comparing the two measures of self-reported decision confidence

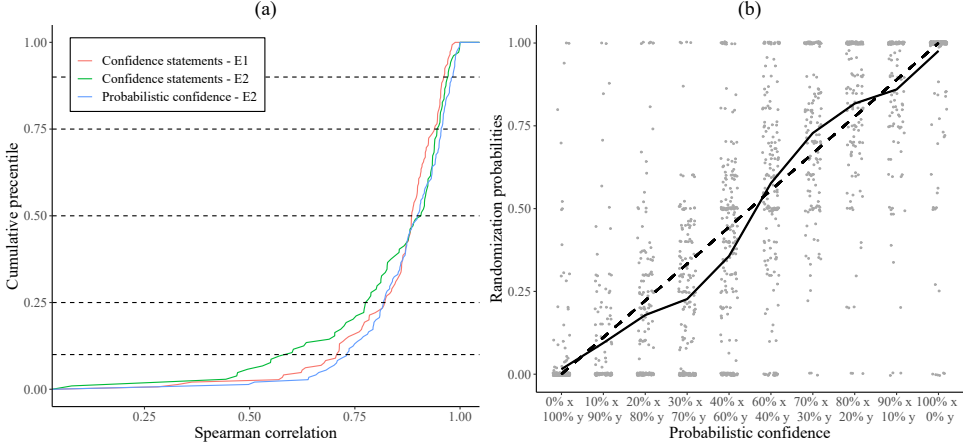


Figure 4: (a) Cumulative distributions of subjects’ Spearman correlations between randomization probabilities and confidence statements or probabilistic confidence in the baseline no-experience treatment for decisions about the simple lottery. E1 and E2 refer to data from Experiment 1 and 2 respectively. (b) Scatter plot of randomization probabilities, with the mean randomization probability (in solid line) at each probabilistic confidence level in Experiment 2 in the baseline no-experience treatment for decisions about the simple lottery. The dashed line is a 45-degree line.

across treatments. We demonstrate that exogenous shifts in self-reported decision confidence are paired with corresponding shifts in randomization probabilities, maintaining their systematic link (hypotheses 3 and 4). We pool subjects in different orders in our main analyses and discuss order effects in subsection 2.4.3.

2.4.1 Randomization probabilities and self-reported confidence

Below, we report two empirical observations that are consistent with Hypothesis 1 and 2.

Result 1. *In the baseline no-experience-simple-lottery treatment, randomization probabilities were significantly and positively correlated with confidence statements and probabilistic confidence among the large majority of subjects. Further, on average, randomization probabilities corresponded to probabilistic confidence in absolute levels.*

To obtain the correlation between randomization probabilities and confidence statements, we transformed the confidence statements to a scale of 1 to 5, with “Surely y” taking the value of 1 and “Surely x” taking the value of 5 to represent one’s decision confidence in choosing x. We computed for each subject the nonparametric Spearman correlation between confidence statements and randomization probabilities in Experiment 1 and 2, and between probabilistic confidence and randomization probabilities in Experiment 2. The results for the baseline treatment are illustrated in Figure 4, Panel (a). Table B.1 in Appendix B summarizes the cross-measure correlations across treatments. Consistent with Hypothesis 1, confidence statements and randomization probabilities have a high and positive correlation. Moderate

to strong correlations of 0.60 in Experiment 1 and 0.71 in Experiment 2 were found at the 10th percentile level, which increased to 0.91 in Experiment 1 and 0.89 in Experiment 2 at the median level. Since the subjects in Experiment 2 reported confidence statements and randomization probabilities in different sessions separated by at least seven days and in different orders, the similarities between the correlations found in Experiments 1 and 2 suggest that confidence statements and randomization probabilities are associated in ways beyond experimenter demand effects and order effects. In Experiment 2, we also found high correlation between self-reported probabilistic confidence and randomization probabilities: the correlation is 0.73 at the 10th percentile and 0.90 at the median.

As correlations do not describe the correspondence between randomization probabilities and self-reported decision confidence in absolute levels, we also computed the mean randomization probability at each level of probabilistic confidence for each subject and took the mean across subjects. This is shown in Panel (b) of Figure 4. Overall, the mean randomization probability for x is close to the probabilistic confidence of choosing x : subjects who chose a randomization probability of, for example, 0.7 would report probabilistic confidence of 70% on average. Examining the absolute correspondence between randomization probabilities and confidence statements gives a similar result, as summarized in Figure B.1 and Table B.2 in Appendix B. These results suggest randomization probabilities can be used as a direct proxy for probabilistic confidence.

Next, we turn to Hypothesis 2 and examine the randomization probabilities around the switching choices. Intuitively, x and y are harder to compare around the switching choices. Reflecting this, subjects reported lower decision confidence and chose randomization probabilities close to 0.5 around the switching choices, as indicated in Result 2.

Result 2. *On average, the subjects reported lower decision confidence around the switching choices based on the self-reported confidence measures and were more likely to randomize and chose randomization probabilities close to 0.5 around the switching choices.*

We study the switching choice of each subject by considering two levels of sure payments: $\underline{y}_b$ and $\bar{y}_b$. We let $\underline{y}_b$ denote the highest sure payment at and below which the subject always preferred x over y , and $\bar{y}_b$ denote the lowest sure payment amount at and above which the subject always chose y over x in the binary choices. We henceforth refer to the values of y between $\underline{y}_b$ and $\bar{y}_b$ as the subject's switching range. This approach allows us to accommodate subjects who switched once as well as those who switched multiple times between lottery x and the sure payments (for the simple and complex lotteries, respectively, 19% and 25% in Experiment 1 and 14% and 23% in Experiment 2).⁵

As expected, decision confidence was lower within the switching range than outside it. In Experiment 2 (Experiment 1), 88% (85%) of the confidence statements within the switching

⁵It is important to include these subjects, because when subjects are not fully confident about their choices, they may switch between x and y multiple times. For the subjects who switched from the lottery to the sure payments once, the switching range simply includes the two sure payments around the switching choice (e.g., if a subject chooses the lottery at $y = 4$ and switches to the sure payment at $y = 4.5$ euros, this means that $\underline{y}_b = 4$, $\bar{y}_b = 4.5$, and the switching range is $[4, 4.5]$).

Experiment	Median behavior around the switching range					
	Confidence statements		Probabilistic Confidence		Randomization probabilities	
	y_b	$\bar{y}_b$	$\underline{y}_b$	$\bar{y}_b$	$\underline{y}_b$	$\bar{y}_b$
Experiment 1 (N = 105)	Probably x	Probably y	-	-	0.67	0.46
Experiment 2 (N = 145)	Probably x	Probably y	60% x	40% x	0.65	0.45

Table 1: Median behavior around the switching choices in the baseline no-experience treatment for decisions about the simple lottery.

range were “Probably x,” “Unsure,” or “Probably y,” compared to 41% (40%) outside the switching range. In Experiment 2, “60% x, 40% y” and “40% x, 60% y” were selected for 53% of the values within the switching range, compared to 13% outside the switching range. Table 1 shows the median randomization probabilities, probabilistic confidence, and confidence statements around the switching range. The median responses to the self-reported confidence measures indicate a lack of confidence around the switching range.

The randomization probabilities within the switching range resemble the two self-reported confidence measures. In Experiment 2 (Experiment 1), 67% (85%) of randomization probabilities reported for values of y within the switching range fell between 0.1 and 0.9, whereas this only holds for 33% (47%) outside the switching range. Further, in Experiment 2 (Experiment 1), the subjects assigned a median randomization probability of 0.65 (0.67) to x at $\underline{y}_b$, and a median randomization probability of 0.45 (0.46) to x at $\bar{y}_b$. The median randomization probability for all the choices that fell within the switching range is 0.5. These results are consistent with Hypothesis 2: subjects were more likely to choose randomization probabilities close to 0.5 for choices that they found difficult to compare.

2.4.2 Manipulating decision confidence

In this section, we examine whether our exogenous manipulations of the decision situation affected self-reported decision confidence in the expected direction and whether randomization probabilities were affected in similar ways to maintain a systematic relationship with the self-reported confidence measures.

2.4.2.1 The complex lottery versus the simple lottery

Consistent with Hypothesis 3, we find that subjects had lower decision confidence when making decisions about the complex lottery compared to the simple lottery. Result 3 summarizes our finding.

Result 3. *Compared to decisions about the simple lottery, the subjects revealed less than full decision confidence over a wider range of sure payments for decisions about the complex lottery, and their reported decision confidence were more compressed toward “Unsure” or (50% x, 50% y). Likewise, the subjects randomized over a wider range of sure payments and chose randomization probabilities closer to 0.5 for decisions about the complex lottery.*

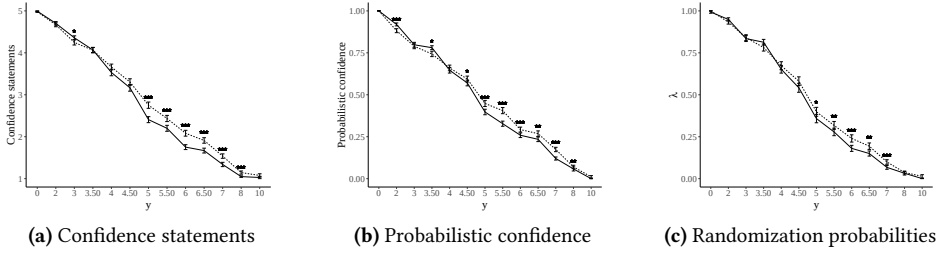


Figure 5: The mean self-reported decision confidence and randomization probabilities for each value of y for the simple lottery (solid line) and complex lottery (dashed line). Wilcoxon signed-rank tests were performed to test the treatment difference for each value of y : * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We find that the range of sure payments over which the subjects chose confidence statements “Unsure” or “Probably” is larger for decisions about the complex lottery than for decisions about the simple lottery in both Experiment 1 and 2, and it is statistically significant in Experiment 2 (Experiment 1: 3.62 vs 3.36, Wilcoxon signed-rank test $p = 0.150$; Experiment 2: 3.58 vs 3.15, Wilcoxon signed-rank test $p < 0.01$). Comparing the range of sure payments for which subjects did not indicate probabilistic confidence of (100% x , 0% y) or (0% x , 100% y) gives similar results: the subjects were not fully confident over a wider range of sure payments for decisions about the complex lottery than decisions about the simple lottery (4.63 vs. 4.37, Wilcoxon signed-rank test $p < 0.01$).

Panel (a) and Panel (b) of Figure 5 illustrate how confidence statements and probabilistic confidence varied with different sure payment amounts across the two lotteries in Experiment 2. Compared to the simple lottery, self-reported decision confidence measures were more compressed towards “Unsure” or (50% x , 50% y) when the subjects faced the complex lottery. The difference in decision confidence across the two lotteries is statistically significant between sure payments of 5 and 8 euros, and is less often statistically significant for lower sure payment amounts. The results for Experiment 1 are similar, albeit weaker, and can be found in Figure B.2 of Appendix B.

We proceed to examine the randomization probabilities chosen for each lottery. As we can see from Table B.3 in Appendix B, in both experiments, the range of sure payments over which subjects chose a randomization probability between 0.1 to 0.9 was significantly larger for decisions about the complex lottery than for decisions about the simple lottery (Experiment 1: 4.06 vs 3.63, Wilcoxon signed-rank test $p < 0.01$; Experiment 2: 3.19 vs 3.03, Wilcoxon signed-rank test $p < 0.10$). This is consistent with the findings from self-reported decision confidence reported above.

Further, the randomization probabilities were also more compressed towards 0.5 when the subjects faced the complex lottery compared to the simple lottery, as shown in Panel (c) of Figure 5. The difference in randomization probabilities across the two lotteries is statistically significant between sure payments of 5 and 7 euros, coinciding with the range obtained

from probabilistic confidence. Panel (c) of Figure 5 also shows asymmetric treatment effects on randomization probabilities for sure payments above 5 euros and sure payments below 5 euros. We show in Appendix A.3 that this asymmetric treatment effect can be consistent with the theoretical analysis when the treatment manipulation affects both preference uncertainty and the average valuation of the lotteries.

Despite our manipulation, the correlations between the two decision confidence measures and randomization probabilities remain similar. Comparing decisions about the simple lottery with those about the complex lottery, the median correlations between randomization probabilities and confidence statements are 0.86 vs 0.82 in Experiment 1 and 0.89 vs 0.88 in Experiment 2. The median correlations between randomization probabilities and probabilistic confidence are 0.90 vs 0.89 in Experiment 2. More results can be found in Table B.1, B.2, and B.4 in Appendix B.

2.4.2.2 Experience and no experience

Hypothesis 4 states that, compared to the baseline no-experience treatment, gaining experience with the lotteries increases decision confidence. The results about Hypothesis 4 are summarized in Result 4.

Result 4. *Decisions in the partial-experience treatment and decisions about the simple lottery in the full-experience treatment did not exhibit significant treatment effects. Comparing decisions about the complex lottery in the full-experience treatment and the no-experience treatment, the subjects (1) revealed less than full decision confidence over a narrower range of sure payments, and their self-reported decision confidence were stretched further away from “Unsure” or (50% x, 50% y); and (2) they randomized over a narrower range of sure payments and chose randomization probabilities further away from 0.5.*

We report the results about the partial-experience treatment and the simple lottery in Figure B.3, Figure B.4, and Table B.5 in Appendix B, and report the comparison between the full-experience treatment and the no-experience treatment about the complex lottery here. We find that the range of sure payments over which subjects reported confidence statements of “Probably” or “Unsure” is significantly narrower (3.16 vs. 3.58, Wilcoxon rank-sum test $p < 0.05$). The range of sure payments over which subjects chose probabilistic confidence between 0.1 and 0.9 did not differ significantly between the full-experience treatment and the no-experience treatment (4.58 vs. 4.63, Wilcoxon rank-sum test $p = 0.590$).

Panels (a) and (b) of Figure 6 show how self-reported decision confidence differs between the full-experience and the no-experience treatment. Compared to the no-experience treatment, self-reported decision confidence was stretched further away from “Unsure” or (50% x, 50% y) for the subjects in the full-experience treatment and these differences in decision confidence were significantly different for sure payments between 2 and 4.5 euros. This implies that the subjects in the full-experience treatment were more confident about which option they preferred than subjects in the no-experience treatment.

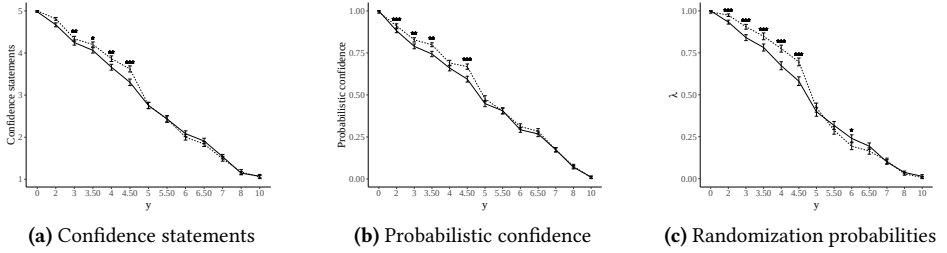


Figure 6: The mean self-reported decision confidence and randomization probabilities for each value of y for the complex lottery in Experiment 2. The graphs show the baseline no-experience treatment (solid line) compared to the full-experience treatment (dashed line). Wilcoxon rank-sum tests were performed to test the treatment difference for each value of y : * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Next, we examine these treatment effects on randomization probabilities. Like decision confidence, we find that the range of sure payments over which subjects chose randomization probabilities between 0.1 to 0.9 was significantly narrower in the full-experience treatment than in the no-experience treatment for decisions about the complex lottery (2.67 vs. 3.19, Wilcoxon rank-sum test $p < 0.05$). Panel (c) of Figure 6 shows that the difference in mean randomization probabilities across sure payments between the full-experience treatment and the no-experience treatment resembles that shown in Panel (a) and Panel (b). Compared to the no-experience treatment, randomization probabilities were also stretched further away from 0.5 among the subjects in the full-experience treatment. Significant differences in the randomization probabilities between subjects in the full-experience treatment and the no-experience treatment were also observed between 2 euros to 4.5 euros. Asymmetric treatment effects on randomization probabilities could also be observed here for sure payments above 5 compared to those below 5, which we discuss further in the theoretical models in Appendix A.3.

The increase in decision confidence from subjects' experience with the lotteries did not affect the high correlation between self-reported decision confidence and randomization probabilities. The median correlation between self-reported decision confidence and randomization probabilities was broadly similar in the two treatments. More details can be found in Table B.1, B.2, and B.4 in Appendix B.

2.4.3 Order effects

An important concern of the within-subject design is order effects: subjects' earlier decisions may affect their subsequent decisions. We are especially concerned with the order effects arising from priming: when randomization probabilities were elicited after self-reported confidence measures, subjects could be primed to link randomization probabilities to decision confidence and consequently reported randomization probabilities that cohered

with self-reported decision confidence measures. Below we highlight the key findings about order effects and leave the details to Appendix B.1.

We find some order effects, suggesting that priming effects on randomized choices could be present. For example, subjects randomized strictly ($0 < \text{randomization probability} < 1$) in fewer choices when randomized choices were presented first (Order 2) compared to later (Order 1 and 3) in all treatments: averaging across treatments, 34% in Order 2 compared to 41% and 45% in Order 1 and 3 respectively (see Table B.6 and Figure B.5 in Appendix B.1 for details). Also, the cumulative distributions of the correlations in Order 2 tend to be lower (on the left of the other two orders), with a larger difference at the lower percentiles and in the complex lottery treatments.

Despite these differences, we find support for our hypotheses among the subjects in Order 2 where the aforementioned priming effects were absent, although the support is sometimes weaker than in the full sample, possibly due to the reduction of sample size. Figure B.6 in Appendix B.1 shows that across treatments, the median cross-measure correlations in Order 2 are high, consistent with H1. Subjects reported low decision around the switching choices and chose randomization probabilities around 0.5 in these choices, supporting H2 (see Table B.7). Table B.8 suggests that subjects in Order 2 had lower decision confidence in decisions involving the complex lottery compared to the simple lottery, consistent with H3. In Order 2, similar to the other two orders, decisions involving the complex lottery showed higher confidence in the full experience treatment than in the non-experience treatment, consistent with H4, although the treatment effect is not statistically significant in any individual order. This high level of consistency between randomization probabilities and self-reported confidence measures as well as their similar reactions to exogenous change of decision confidence suggest that these measures likely share common psychological foundations, even if not identical.

Finally, it is worth noting that the order effects discussed above do not necessarily suggest that randomization probabilities are a poorer proxy for decision confidence than the self-reported confidence measures. Confidence statements and probabilistic confidence are also noisy proxies of decision confidence, and it is not obvious what the “right” amount of strict randomization is. When we assess the value of decision confidence based on its correspondence with actual choices (when subjects report lower confidence in choosing an option, they should be less likely to choose that option), we find both randomization probabilities and probabilistic confidence corresponded to actual choices. Importantly, randomization probabilities exhibited a closer correspondence to binary choices than probabilistic confidence, especially in Order 2. See Appendix B.1 for more details.

2.4.4 Alternative interpretations of randomization probabilities

We have interpreted randomization behavior as a lack of decision confidence in the face of preference uncertainty. This is consistent with the findings from our post-experiment questionnaire as well as findings in Agranov and Ortoleva (2023) where many subjects

explicitly mentioned unsureness, complexity, difficulty, and hedging as reasons for randomization (see Appendix B.3 for details). However, subjects may have other reasons for randomization. While it is not possible to eliminate all alternative interpretations, we show that indifference, random errors, or utility difference alone cannot be the driving force behind subjects' randomization behavior and the treatment effects.

First, indifference is not the driving reason for randomization because the majority of the subjects randomized at least twice (see Table B.9 in Appendix B.2), while randomization from indifference should occur for at most one value of sure payments. Second, randomization was unlikely a result of random errors because subjects' randomization probabilities of choosing x decreased monotonically with the value of y , even though they faced a random sequence of y (see e.g., Panel (c) in Figure 5). Further, the treatment effects on randomization probabilities in the expected directions suggest that subjects' randomization was likely a deliberate choice. Third, if randomization probabilities were due to utility differences alone, randomization probabilities should increase for each value of sure payments when the utility over the lottery increases (e.g., moving from the simple lottery to the complex lottery or the no-experience treatment to the full-experience treatment), which increases the lower bound and the upper bound of randomization as well. Our results clearly reject these predictions. Randomization probabilities were compressed toward 0.5 rather than increased monotonically, and the two bounds often moved in opposite directions. We elaborate on this in Appendix B.2.

2.5 Conclusion

We propose that letting individuals assign randomization probabilities according to which they receive options can be an incentivized way to elicit decision confidence. In two experiments, we elicited randomization probabilities as well as two self-reported confidence measures and further manipulated decision confidence exogenously.

We find that most subjects randomized frequently, and their randomization probabilities and self-reported confidence measures were linked in ways that are consistent with the hypotheses derived from two theoretical analyses. While there were some order effects depending on whether randomization probabilities were elicited before or after the self-reported confidence, cross-measure correlations were high, and randomization probabilities corresponded closely to probabilistic confidence in absolute levels, with a high randomization probability assigned to an option associated with high self-reported probabilistic confidence. Our further examination suggests that alternative interpretations of randomization such as indifference, random errors, or differences in utility alone are unlikely to be the driving factors. Overall, our results suggest that decision confidence can be meaningfully and accurately inferred from randomization probabilities.

The WTA-WTP gap and Complexity in a Representative Sample

Abstract: We quantitatively examine the impact of complexity on preference revelation in a representative sample. We elicited each participant's willingness-to-accept and willingness-to-pay (WTA-WTP) gap for monetary lotteries, and measured the complexity of the price list elicitation mechanism and lottery valuation. We find that complexity measures jointly explain approximately 50% of the gap, and its explanatory power persists across diverse subgroups, with a stronger effect among participants with low cognitive ability. In contrast, loss aversion explains the gap only among participants experiencing low complexity. Our study demonstrates that the WTA-WTP gap can arise from distinct behavioral reasons with their impacts varying across subjects.

This chapter is currently under review at the *Economic Journal*.

3.1 Introduction

Revealing unobservable preferences from observable choices is critical in economic analyses. One significant challenge in the revealed preference approach is the role of task complexity, defined as the difficulty of making optimal decisions in a task. Without effective ways to mitigate such complexity, decision-making may be subject to systematic biases and random noise, potentially distorting the inference of underlying preferences.

Two forms of complexity have gained increasing attention. In the first form, decision makers may find some preference elicitation mechanisms complex and do not fully understand the mechanism (referred to as mechanism complexity hereafter) (Plott and Zeiler, 2005, 2007; Cason and Plott, 2014; Smitizsky et al., 2021). Consequently, they may rely on heuristics to make decisions (Kahneman et al., 1982). Or, they may confuse one mechanism with another, for instance, misperceiving the Becker–DeGroot–Marschak mechanism (Becker et al., 1964) – a form of the second-price auction – as the first-price auction (Cason and Plott, 2014), or the deferred acceptance mechanism for the Boston mechanism (Rees-Jones and Shorrer, 2023). To overcome mechanism complexity, some researchers have sought to develop simpler mechanisms (Li, 2017), while others have emphasized the need to train subjects to understand the mechanisms (Plott and Zeiler, 2005) or better describe the mechanisms (Gonczarowski et al., 2022).

In the second form of complexity, decision-makers may have difficulty forming a clear valuation for choice objects. We refer to this as valuation complexity.⁶ The idea that subjects may not clearly know the value of choice objects is longstanding. Von Neumann and Morgenstern (1944) comment on the completeness axiom that “it is very dubious, whether the idealization of reality which treats this postulate as a valid one, is appropriate or even convenient.” In response to valuation complexity, decision makers may behave cautiously by sticking to the default (Masatlioglu and Ok, 2005), postponing their decisions if possible (Danan and Ziegelmeyer, 2006), randomizing deliberately (Agranov and Ortoleva, 2017, 2023; Cettolin and Riedl, 2019; Halevy et al., 2023), using less costly but error-prone valuation procedures (Oprea, 2024; Arrieta and Nielsen, 2024), or opting for satisficing instead of maximizing (Salant and Spenkuch, 2022).

In this study, we quantitatively assess the impact of both forms of complexity in revealing preferences via the willingness-to-accept and willingness-to-pay gap (the WTA-WTP gap, henceforth) (Thaler, 1980; Knetsch, 1989; Kahneman et al., 1990). While the WTA-WTP gap holds a central position in the field of behavioral economics, it remains a subject of considerable debate (Ericson and Fuster, 2014). In addition to reference dependence and loss

⁶The key underpinning of our notation of valuation complexity is that participants may not always know for sure their preferred option. In this spirit, valuation complexity includes several notions in the literature, such as incomplete preference (Bewley, 2002; Masatlioglu and Ok, 2005; Eliaz and Ok, 2006; Ok et al., 2012; Nishimuray and Ok, 2018); preference imprecision (Dubourg et al., 1994; Butler and Loomes, 2007; Cubitt et al., 2015); and the recently developed notion of cognitive noise, according to which decision-makers’ mental representation of the decision situation can be noisy and lead to a range of behavioral anomalies (Khaw et al., 2021; Woodford, 2020; Frydman and Jin, 2021; Vieider, 2021; Enke and Graeber, 2023; Oprea, 2024).

aversion, the WTA-WTP gap may be due to many factors, and “seemingly small procedural details matter substantially” (Ericson and Fuster, 2014, p. 571). In particular, it has been proposed and shown that subjects’ misconception about the elicitation mechanism may be responsible for the WTA-WTP gap (Loomes et al., 2003; Plott and Zeiler, 2005; and Cason and Plott, 2014. See also Plott and Zeiler, 2007, 2011, and Isoni et al., 2011 for discussions). It has also been hypothesized that valuation complexity may contribute to the WTA-WTP gap (Dubourg et al., 1994; Bayrak and Hey, 2020; McGranaghan and Otto, 2022). Concretely, when decision-makers are unsure about the valuation of an object, they act cautiously by bidding a low price when buying it and asking for a high price when selling it, resulting in the WTA-WTP gap (Cerreia-Vioglio et al., 2024). Building on these studies, we conduct a systematic investigation of the WTA-WTP gap in a representative sample and show that the two forms of complexity along with loss aversion are critical factors in this ongoing debate.

We elicit the WTA and WTP of two monetary lotteries in a within-subject setting using the price list elicitation mechanism. To assess the complexity of the price list, we elicit the WTA and WTP of an object with a known preference: a degenerate lottery of receiving 2.5 euro for sure (denoted as 500 points in the experiment). If subjects are loss averse and fully understand the incentives of the price list, they would exhibit a WTA-WTP gap for lotteries but not for the degenerate lottery. On the other hand, if subjects find the price list complex and carry the heuristic of buy-low and sell-high in real life into the lab or misperceive the selected price in the list as the final trading price, they may have a WTA-WTP gap for both lotteries and the degenerate lottery. Furthermore, the WTA-WTP gap for lotteries would correlate significantly with that for the degenerate lottery.

We assess the valuation complexity of monetary lotteries by eliciting a range of prices instead of one precise price. Specifically, as in a price list, subjects face a series of choices between the (fixed) lottery and a (varying) price. Differently, in each choice, subjects have three options: the lottery, the sure payment, or the randomization option, where the computer flips a virtual coin to determine whether subjects receive the lottery or the sure payment. When subjects find the valuation of the lottery complex, they may have a range of prices for which they would prefer to choose the randomization option (Cettolin and Riedl, 2019; Agranov and Ortoleva, 2023; Halevy et al., 2023; Arts et al., 2024a), and the range of prices can serve as a (probably conservative) measure of valuation complexity. Based on the hypothesis that valuation complexity and caution contribute to the WTA-WTP gap (Cerreia-Vioglio et al., 2024), we test whether subjects with a wider range of prices are more likely to exhibit a larger WTA-WTP gap.

We examine these two forms of complexity in a large-scale experiment with 1856 participants from the LISS (Longitudinal Internet studies for the Social Sciences) panel, a representative sample of the Dutch population. The LISS panel has been widely used for surveys and experiments in both economic policies and academic research (see e.g., Dimmock et al., 2016; Baillon et al., 2017; Cherchye et al., 2017). From each subject, we elicited the WTA and WTP for two monetary lotteries, measures of the two forms of complexity,

and a measure of cognitive abilities using matrix reasoning questions from the International Cognitive Ability Resource (ICAR, Condon and Revelle, 2014).

We have five main observations. First, on average, subjects exhibit a WTA-WTP gap for both lotteries and the degenerate lottery: The proportion with a positive (negative) WTA-WTP gap is 58% (27%) for the lotteries and 37% (18%) for the degenerate lottery. More importantly, the WTA-WTP gap for the lotteries and that for the degenerate lottery are significantly correlated. Second, subjects have a substantial range of prices (about 20% of the certainty equivalent of the lottery) and the range is positively linked to the WTA-WTP gap of the lotteries. Third, in a set of regression analyses, we show that controlling for the WTA-WTP gap of the degenerate lottery reduces the WTA-WTP gap for lotteries by about 40%, and controlling for the ranges of prices reduces it by 15%. Controlling for both forms of complexity reduces it by about 50%. Fourth, we find the above three patterns are robust with respect to various specifications and in various demographic subgroups, such as age, gender, education, and income. Overall, these observations cannot be accounted for by loss aversion and suggest that the observed WTA-WTP gap for lotteries is partly due to complexity.

Finally, when we merge our data with previous data from Goossens and Knoef (2022), which includes a measure of behavioral loss aversion, we find that loss aversion does not correlate with the WTA-WTP gap in the general sample, consistent with findings based on representative samples in Chapman et al. (2024) and Fehr and Kübler (2022). However, the correlation is significantly positive among subjects who do not experience either form of complexity (Spearman's $\rho = 0.507$, $p < 0.01$). Furthermore, the role of loss aversion differs substantially across subsamples with different cognitive abilities. Among subjects who performed above the median on the matrix reasoning questions, the WTA-WTP gap is reduced by 53% when accounting for loss aversion and 35% when accounting for the two forms of complexity. In contrast, among subjects who performed below the median in the matrix reasoning questions, the WTA-WTP gap is actually increased by 11% when accounting for loss aversion and reduced by 83% when accounting for the two forms of complexity. These results suggest that subjects may exhibit similar WTA-WTP gaps but differ substantially in the underlying reasons.

Our paper combines the two streams of literature on complexity and makes several contributions. First, we quantify the potential role of complexity in driving behavioral anomalies. While it is well known that complexity significantly influences decision-making (see, e.g., Plott and Zeiler, 2007, for the impact of mechanism complexity on WTA-WTP gap), we demonstrate how substantial this effect can be — accounting for approximately 50% of the WTA-WTP gap observed in our study. Our demonstration is possible because our elicitation of WTA and WTP is within-subject, and the individual level WTA-WTP gap provides a natural benchmark for the severity of deviation from classical theory (Brown et al., 2017, 2021). By examining the reduction of the WTA-WTP gap after controlling for the two forms of complexity separately and jointly, we can show the relative importance of each form in explaining the observed gap, as well as the remaining room for other

factors, such as loss aversion. Such investigation would not have been feasible with the between-subject elicitation of WTA and WTP, as commonly seen in most previous studies.

Second, our study considers the WTA-WTP gap in a representative sample of the Dutch population. Responses from representative samples often play a vital role in policy design. Due to the inclusion of low education, low cognitive capacity and short attention span subjects, however, complexity may have a particularly significant role in those samples. Therefore, it is important to differentiate responses that reflect preferences from noise or biases. Snowberg and Yariv (2021) compared a wide range of choice behaviors of different subject pools, including a student population, a representative US sample, and subjects from Amazon Mechanical Turk, and showed that, whereas correlations between behaviors were similar across samples, non-student samples exhibited higher levels of noise. Adding to two recent studies on the WTA-WTP gap in representative samples (Chapman et al., 2024; Fehr and Kübler, 2022), our study suggests that participants with low cognitive ability in the general population are influenced more by complexity, and complexity-driven choices explain the lack of correlation between the WTA-WTP gap and loss aversion.

Third, we use a price list elicitation mechanism and assess its complexity. Previous studies have shown that BDM mechanism can be complex (Plott and Zeiler, 2005; Cason and Plott, 2014). While the price list elicitation mechanism is commonly believed to be simpler and thus widely used, its complexity is less well studied. We demonstrate that the mechanism complexity of the price list significantly influences choice behavior, and accounts for 45% of the WTA-WTP gap.

Finally, our findings can help reconcile some important results regarding the WTA-WTP gap that may otherwise seem surprising or contradictory. For example, the exchange between Plott and Zeiler (2005) and Isoni et al. (2011) showed that, after controlling for subjects' understanding of the elicitation mechanism, the WTA-WTP gap disappeared for mugs but persisted for monetary lotteries. While Plott and Zeiler (2011) did not explicitly identify the underlying mechanism based on their data, they proposed that lotteries may possess a distinct form of complexity. Our study suggests that this complexity could be related to the valuation complexity of lotteries (Oprea, 2024). Another example is the mixed evidence on the role of expectation-based reference points in the endowment effect (Ericson and Fuster, 2011; Sprenger, 2015; Cerulli-Harms et al., 2019). Our finding suggests that the instability of expectation-based reference effects may be due to complexity, in which expectations require simultaneous consideration of several states and values (Martínez-Marquina et al., 2019; Esponda and Vespa, 2023). A further example is the mixed effects of market experiences on exchange asymmetry, which is closely related to the WTA-WTP gap. Several studies showed that exchange asymmetry can be mitigated or eliminated by extensive market trading experience (List, 2003; Engelmann and Hollard, 2010; List, 2011; Anagol et al., 2018), while Apicella et al. (2014) observed that exchange asymmetry is present for hunter-gatherers with exposure to markets but not for those living in isolated regions. The use of the buy-low and sell-high heuristic when facing complexity shown in our study suggests that this may be because some experience with buying and selling is required to

develop this heuristic, but extensive experiences may reduce complexity and thus overcome its misuse. In a recent study, Fehr et al. (2022) showed that exchange asymmetry is less pronounced among the poor than the rich, and they attributed this observation to increased attention to decisions that involve higher stakes. Increased attention may help reduce the experienced complexity of the choice environment, thus reducing the exchange asymmetry of the poor.

To sum up, our results show that complexity has a substantial impact on the WTA-WTP gap, and this impact persists across diverse subgroups. In contrast, loss aversion explains the WTA-WTP gap only among those experiencing low complexity and those exhibiting high performance on matrix reasoning questions. These findings highlight the difficulty in revealing preferences from choices and emphasize the need for caution when interpreting choices as reflections of underlying preferences (Plott and Zeiler, 2005, 2007; Cason and Plott, 2014; Enke and Graeber, 2023; Oprea, 2024; Enke et al., 2025).⁷

The paper proceeds as follows. Section 3.2 explains the experimental design. Section 3.3 reports the experimental results, and Section 3.4 concludes.

3.2 Experimental Design

Each subject was presented with two monetary lotteries and reported their WTA and WTP for each lottery via price lists. We also measured the mechanism complexity of the price list and the valuation complexity of each lottery. All monetary outcomes were denoted by points in the experiment, with an exchange rate of 100 points = 0.50 euro. We explain each of these tasks in detail below.

3.2.1 Eliciting within-subject WTA and WTP

We elicited, within-subject, the WTA and WTP for two monetary lotteries. The first lottery (L1) offered 900 points or 100 points with equal probability. The second lottery (L2) offered 975 points with 20% chance, 850 points with 30% chance, 150 points with 30% chance, and 25 points with 20% chance.

Following Chapman et al. (2024), we elicited WTA and WTP with a price list with 20 rows (Holt and Laury, 2002). In each row, subjects chose between two options: a lottery and a price. The lottery was fixed in the entire list, while the price changed monotonically across rows. The prices were 0, 100, 200, 250, 300, 325, 350, 375, 400, 425, 450, 475, 500, 525, 550, 600, 700, 800, 900, and 1000 points. Note that the lottery dominated the price of 0 points and was dominated by the price of 1000 points. Similar to Chapman et al. (2024), we pre-selected the dominating option in these two rows. In the price list, we asked subjects to select the row in which they would switch from one option to the other. If a price list

⁷Training can help reduce complexity (see e.g., Study 3 in Chapman et al., 2024). However, extensive training is often impractical. Even when it is provided, its effectiveness can be challenging to evaluate, with some participants still struggling with complexity. Another approach is to exclude subjects who do not pass the understanding test; however, this is not feasible when studying a representative sample or the general population.

Question type	The choice in each row	
	Option 1	Option 2
WTA	Sell the lottery for y points.	Keep the lottery.
WTP	Buy the lottery for y points and keep the remaining $[1000-y]$ points.	Keep the 1000 points.
Lotteries		
L1	50%, 900 points; 50%, 100 points.	
L2	20%, 975 points; 30%, 850 points; 30%, 150 points; 20% 25 points.	
Prices y		
Both lotteries	0, 100, 200, 250, 300, 325, 350, 375, 400, 425, 450, 475, 500, 525, 550, 600, 700, 800, 900, 1000.	

Table 2: Summary of the elicitation of WTA and WTP and the lotteries in the price list. The options appearing on the left or the right of the price list as well as the (increasing or decreasing) sequence of prices in the price list were randomized on the individual subject level.

was randomly chosen for payment, subjects' payoff was determined by their choice in the randomly chosen row of the price list. To control for the effects of presentation on the elicitation of WTA and WTP, we randomized the option that appeared on the right or the left of the price list as well as the (increasing or decreasing) sequence of prices at the individual level.

To elicit WTA, we informed subjects that they were endowed with the lottery. In each row of the price list they chose between selling the lottery for a price or keeping the lottery. To elicit WTP, we endowed subjects with 1000 points. In each row of the price list they chose between buying the lottery for a price or keeping the 1000 points. We define WTA and WTP as the midpoint of the prices in the two switching rows, and the WTA-WTP gap is the difference between WTA and WTP. We randomized the eliciting order of WTA and WTP and separated them by other tasks. Using the language of buying and selling to elicit WTA and WTP is consistent with the literature (see, e.g., Isoni et al., 2011; Chapman et al., 2024). To check whether the framing of the WTA and WTP questions worked as expected, we also elicited the certainty equivalents of the two monetary lotteries in a neutral frame from a separate sample.

3.2.2 Measuring mechanism complexity

To measure the mechanism complexity of the price list, we elicited subjects' WTA and WTP for a degenerate lottery – receiving 500 points with 100% chance – with the same type of price list as those for the two lotteries. The prices in the list for the degenerate lottery were 450, 475, 499, 501, 525, and 550 points. We replaced the exact value of 500 points with 499 points and 501 points to avoid indifference between the price and the degenerate lottery of 500 points. Similar to the \$2 card used by Cason and Plott (2014), our degenerate lottery

has no payoff uncertainty and should thus induce a known preference. Should subjects understand the price list correctly, they would switch between 499 and 501 on both WTA and WTP questions. This would result in $WTA_{sure} = WTP_{sure} = 500$, and a WTA-WTP gap of zero. Deviations of the WTA and WTP from 500 points violate dominance, and we interpret these dominance violations as reflecting mechanism complexity.

3.2.3 Measuring valuation complexity

Price	Choices		
0	Lottery	Randomization	Price
100	Lottery	Randomization	Price
200	Lottery	Randomization	Price
250	Lottery	Randomization	Price
300	Lottery	Randomization	Price
325	Lottery	Randomization	Price
350	Lottery	Randomization	Price
375	Lottery	Randomization	Price
400	Lottery	Randomization	Price
425	Lottery	Randomization	Price
450	Lottery	Randomization	Price
475	Lottery	Randomization	Price
500	Lottery	Randomization	Price
525	Lottery	Randomization	Price
550	Lottery	Randomization	Price
600	Lottery	Randomization	Price
700	Lottery	Randomization	Price
800	Lottery	Randomization	Price
900	Lottery	Randomization	Price
1000	Lottery	Randomization	Price

Figure 7: An example of the R-range price list. The lower bound of the R-range is 437.5 points, the upper bound is 512.5 points, and the size of the R-range is $512.5 - 437.5 = 75$ points.

To measure valuation complexity, we modified the binary choices in the standard price list with a third option: the randomization option of receiving the lottery or the price according to a computerized coin flip. For each lottery, subjects faced an extended price list with 20 rows, referred to as the R-range price list. In each row of the R-range price list, subjects could choose the left option or the right option, which were either the lottery or the price as in the standard price list, or the new middle option of randomization. Similar to the elicitation of WTA and WTP, subjects needed to indicate the row of the R-range price list in which they would switch from the left option to the middle option of randomization, and the row in which they would switch from the middle option to the right option. Subjects could make these two switches in the same row and not select the randomization option at all. Figure 7 gives an example of the R-range price list. We measure valuation complexity by the range of prices in which subjects chose the randomization

option (the R-range). As shown in Figure 7, the lower bound of the R-range is the midpoint of the prices between which subjects switch from the lottery to the randomization option, and the upper bound of the R-range is the midpoint of the prices between which subjects switch from the randomization option to the price. The size of the R-range is the difference between the two bounds. Whereas choosing the randomization option more than once could be due to multiple reasons (Chew et al., 2022), it is often interpreted as evidence of incomplete/imprecise preferences and reveals valuation complexity of the objects (Cettolin and Riedl, 2019; Agronov and Ortoleva, 2023).⁸

The R-range price list is more complicated than the standard price list, and subjects may also perceive mechanism complexity. To measure the mechanism complexity of the R-range price list, we also elicited the R-range for the degenerate lottery of 500 points. As stipulated by stochastic dominance, subjects who correctly understood the elicitation mechanism of the R-range price list never chose the randomization option in this price list and would switch between 501 points and 499 points. Deviations from the above choices imply mechanism complexity. We used the R-range of the degenerate lottery to capture the complexity of the R-range elicitation mechanism.

3.2.4 Other measures

One potential source of mechanism and valuation complexity is limited cognitive ability. Given the same elicitation mechanism and choice objects, subjects with low cognitive ability are likely to be more prone to these two forms of complexity. To examine this hypothesis, we included six matrix reasoning questions from the International Cognitive Ability Resource (ICAR, Condon and Revelle, 2014) to assess subjects' cognitive ability. In these questions, subjects had to choose the image that best completed a 3 by 3 matrix (see Appendix B.7 for an example). We incentivized the elicitation of cognitive ability by making subjects' payment dependent on the number of correct answers they provided.

To better understand the forces underlying the R-range, we included a subset of questions from the desirability of control scale, which elicits how subjects perceive the importance or benefits of maintaining control (Burger and Cooper, 1979; Gebhardt and Brosschot, 2002). As Fudenberg et al. (2015) pointed out, a preference for randomization may arise when the decision-maker makes a trade off between the probability of errors and the cost of implementing the desired choice. In this sense, a decision-maker who prefers to maintain control may perceive the implementation cost to be small and be less willing to choose the randomization option. Appendix B.5.2 explains in more detail how desirability of control may be related to valuation complexity. The scale consisted of statements like "I enjoy making my own decisions" and "Others usually know what is best for me." Subjects reported

⁸ Arts et al. (2024a) validated randomization as a reliable measure of valuation complexity. They elicited from each subject three measures of valuation complexity, including self-reported confidence statements, probabilistic confidence, and randomization. Their results suggest that randomization correlated strongly with the other two measures of valuation complexity, and the correlations were robust to the exogenous manipulations of the payoff structure of the lottery and subjects' experiences with the lottery.

on a 7-point Likert scale how strongly they agreed or disagreed with these statements, ranging from completely disagree to completely agree. Appendix B.7 provides the complete list of statements.

The questions above also served to separate WTA questions, WTP questions, and R-range questions.

3.2.5 Sample and procedure

We conducted the experiment using the LISS (Longitudinal Internet studies for the Social Sciences) panel administered by CentERdata (Tilburg University, The Netherlands) in September 2021. The LISS panel is a representative sample of the Dutch population who participate in online surveys monthly.⁹

We collected data from a total of 1856 subjects. Demographic information about the subjects is reported in Table B.1 in Appendix B.1. For the main sample of 1236 subjects, we used a within-subject design in which subjects reported their WTA, WTP, and R-range for each lottery. To address the concern that subjects may be influenced by their earlier responses, we separated the elicitation of WTA, WTP, and R-range by inserting other questions. We also randomized the order in which subjects responded to the WTA, WTP, and R-range questions. The order of the decisions was either: 1) WTA → matrix reasoning → WTP → desirability of control → R-range; 2) WTP → matrix reasoning → WTA → desirability of control → R-range; 3) R-range → desirability of control → WTA → matrix reasoning → WTP; or 4) R-range → desirability of control → WTP → matrix reasoning → WTA. We further randomized the order of L1 and L2 in each task. For a separate sample of 620 subjects, we replaced the WTA and WTP questions with price lists in a neutral frame to elicit the certainty equivalents for the same two monetary lotteries.

After the online survey was completed, we randomly selected 10% of the subjects and paid them an additional amount based on one of their decisions in the experiment (WTA questions, WTP questions, R-range questions, or their performance on the six matrix reasoning questions). If the decision selected for payment was the randomization option, a random draw determined whether subjects would receive the lottery or the price. If they received the lottery, a further random draw determined the payment of the lottery. On average, the subjects who were selected for the additional payment received 2.97 euros on top of the flat fee of 3 euro for completing the survey. The experiment was approved by the Institutional Review Board of Radboud University.

3.3 Experimental Results

In reporting the experimental results, we use Wilcoxon signed-rank tests for within-subject statistical differences, Wilcoxon rank-sum tests for between-subject statistical differences,

⁹Households that could not otherwise participate were provided with a computer and Internet connection. The longitudinal survey is fielded in the panel every year, covering a large variety of domains including health, work, education, income, housing, time use, political views, values, and personality.

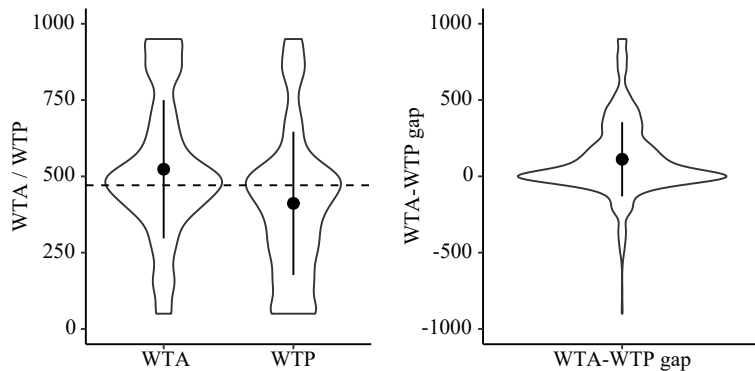


Figure 8: Violin plots of the average WTA, WTP, and WTA-WTP gap of the two lotteries. The dots in the violins denote the means, the lines represent the standard deviations. The dashed line is the average certainty equivalent of the lotteries from the neutral frame (470.84 points). See Table B.2 in Appendix B.1 for details on each lottery separately.

and Obviously-Related Instrumental Variables (ORIV; Gillen et al., 2019) to estimate correlations, unless stated otherwise.¹⁰

We find that on average subjects exhibited the standard WTA-WTP gap by reporting a higher WTA than WTP. Figure 8 reports violin plots of the average WTA, WTP, and WTA-WTP gap across the two lotteries. Across the two lotteries, the mean WTA is 523.47 and the mean WTP is 411.51, resulting in an average WTA-WTP gap of 111.96. The WTA-WTP gap is about 22% of the expected value of the lotteries, which is comparable to the finding reported by Chapman et al. (2024). In contrast, the mean certainty equivalent from the neutral frame is 470.84, which differs significantly from the WTA and WTP of the lotteries ($p < 0.01$). This suggests that the framing of the WTA and WTP questions in the experiment worked in the expected direction. On the individual level, 58.41% of subjects showed a positive WTA-WTP gap, 15.05% had a WTA-WTP gap of 0, and a non-negligible proportion (26.54%) of subjects exhibited a negative WTA-WTP gap. These values are broadly consistent with the proportions reported in previous studies. For example, in Chapman et al. (2024) these values are 60%, 10%, and 30% respectively.

In the following analyses, we present our results in three steps. First, we report measures of mechanism complexity and valuation complexity and examine their association with the WTA and WTP of the lotteries. Second, we correct the WTA-WTP gap for mechanism complexity and/or valuation complexity. Finally, we demonstrate the robustness of our results by considering different demographic groups, the inclusion of loss aversion, and a comparison across subsamples.

¹⁰We use the WTA and WTP of two lotteries for ORIV estimation of correlations. The choice behavior does not significantly differ between the two lotteries, as shown in the Appendix B.1.

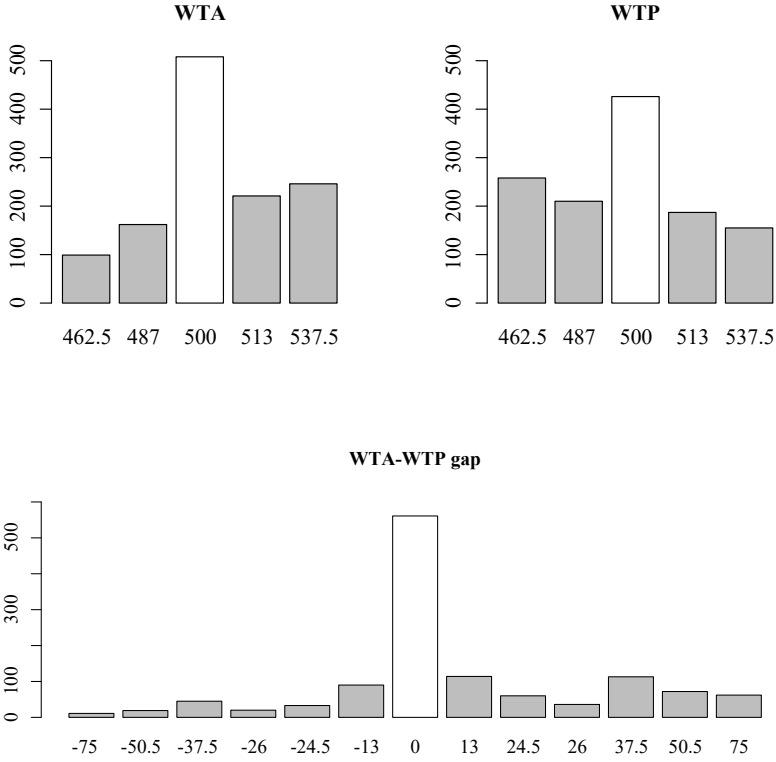


Figure 9: The frequencies of subjects at each level of the WTA, WTP, and WTA-WTP gap for the degenerate lottery. The white bars indicate the optimal WTA and WTP.

3.3.1 Mechanism complexity

We find that mechanism complexity, as measured by deviations of WTA_{sure} and WTP_{sure} from 500, is systematic. Figure 9 presents the distribution of WTA_{sure} , WTP_{sure} , and the $WTA-WTP$ gap_{sure}. We find that 64% of WTA_{sure} deviations were positive and 36% were negative. In contrast, 58% of WTP_{sure} deviations were negative and 42% were positive. This leads to more positive WTA-WTP gaps for the degenerate lottery than negative ones. Three logistic regressions (reported in Table B.3 in Appendix B.1) suggest that the likelihood of reporting $WTA_{sure} = WTP_{sure} = 0$ is significantly and positively related to subjects' performance on the matrix reasoning questions, education, self-reported understanding of the experimental tasks, and time spent on the tasks. These factors are commonly associated with subjects' ability or willingness to understand experimental tasks. Thus, deviations of the degenerate lottery's WTA and WTP from 500 are likely to arise from subjects' difficulty

with the complexity of the price list elicitation mechanism, which supports the use of these deviations as a proxy for mechanism complexity.

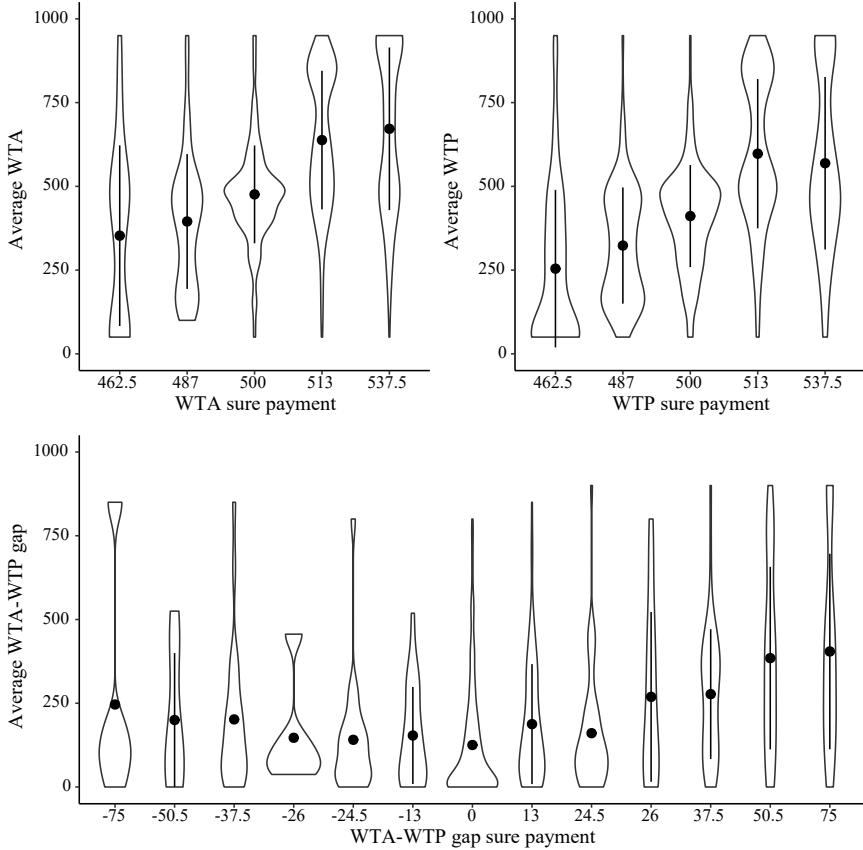


Figure 10: Violin plots of the average WTA, WTP, and WTA-WTP gap of the two lotteries at each level of the degenerate lottery's WTA, WTP, and WTA-WTP gap. The dots in the violins denote the means, the lines represent the standard deviations.

Systematic bias of WTA_{sure} and WTP_{sure} can also be seen when we relate them to the WTA and WTP of the two lotteries. There is a similar asymmetry in the proportions of positive and negative WTA-WTP gaps between the degenerate lottery (36.97% vs. 17.64%) and the lotteries (58.41% vs. 26.54%). Figure 10 depicts violin plots of the average WTA and WTP of the lotteries at each level of WTA_{sure} and WTP_{sure} , respectively. The data pattern suggests a systematic association of WTA and WTP between the degenerate lottery and the lotteries and further shows that the WTA of the lotteries associates more strongly with positive WTA_{sure} deviations than negative ones, while the WTP of the lotteries associates more strongly with negative WTP_{sure} deviations than positive ones. Consistent with Figure 10, we find significantly positive correlations between WTA_{sure} and the average WTA of

the lotteries ($\rho = 0.599, p < 0.01$), between WTP_{sure} and the average WTP of lotteries ($\rho = 0.612, p < 0.01$), and between the WTA-WTP gap of the degenerate lottery and the average WTA-WTP gap of the lotteries ($\rho = 0.578, p < 0.01$). Examining positive and negative deviations of the WTA and WTP for the degenerate lottery from 500 separately, we find that positive WTA deviations and negative WTP deviations, which are consistent with a positive WTA-WTP gap, correlate significantly with the WTA and WTP of the lotteries ($\rho = 0.107, p < 0.05$ for WTA, and $\rho = 0.228, p < 0.01$ for WTP), while correlations between negative WTA deviations or positive WTP deviations and those of the lotteries are weaker ($\rho = 0.123, p < 0.10$ for WTA and $\rho = -0.084, p > 0.10$ for WTP). These results suggest that positive WTA_{sure} deviations and negative WTP_{sure} deviations are systematic biases, while negative WTA_{sure} deviations and positive WTP_{sure} deviations are likely to arise from noise. We summarize our observation below.

Observation 1. *The WTA and WTP of the degenerate lottery deviate systematically from the known preference, and these deviations correlate strongly with the WTA and WTP of the lotteries.*

3.3.2 Valuation complexity

As we explained in the experimental design, we capture valuation complexity by the range of prices for which subjects chose the randomization option (the R-range). We find that the majority of subjects chose the randomization option at least once for either of the lotteries (67%), and many did so twice or more (56%). This results in a substantial R-range (mean R-range of 103.96 across the two lotteries). Table B.4 in Appendix B.1 provides more details on the upper bound, lower bound, and R-range for each lottery.

The R-range also exhibits some systematic patterns. First, we hypothesized that subjects who have a stronger desire for control are more willing to pick their desired option than the randomization option. Consistent with the hypothesis, we find a significant negative correlation between subjects' desirability of control and their average R-range across the two lotteries ($\rho = -0.103, p < 0.05$). We also find a statistically significantly higher level of desirability of control among subjects who do not have an R-range for either lottery than among those who chose the randomization option at least once in one of the two lotteries (5.28 vs. 5.16, $p < 0.01$).

Second, we find a positive and significant correlation between the R-range and the WTA-WTP gap ($\rho = 0.144, p < 0.01$). When we restrict to subjects who randomized at least once, we find an even stronger correlation ($\rho = 0.264, p < 0.01$). Consistent with the correlation tests, we find that subjects who chose the randomization option more than twice in the two lotteries have a significantly higher WTA-WTP gap than those who chose the randomization option only once or twice (122.07 vs 79.61, $p < 0.01$). This trend can also be seen in Figure 11. The picture for the group of subjects who never chose the randomization option is less clear. They have no R-range, but their WTA-WTP gap is significantly larger compared to those with low valuation complexity (122.04 vs. 78.85, $p < 0.05$) but similar

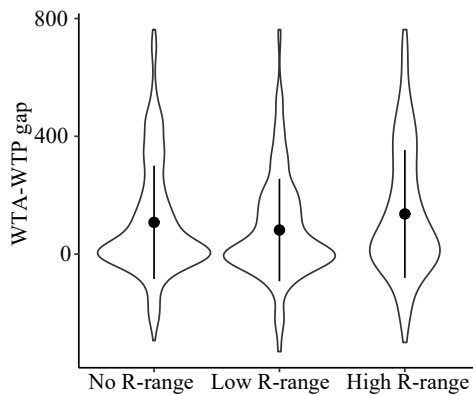


Figure 11: Violin plots of the average WTA-WTP gap at different sizes of the R-range in L1 and L2. Low and high R-range are computed by taking the median split of the R-range (100 points) for subjects with an R-range larger than zero. The dots in the violins denote the means, the vertical lines represent the standard deviations. For better readability, we excluded the top and bottom 2.5% of the WTA-WTP gap (-332.03 and 763.28 points).

to those with high valuation uncertainty (122.04 vs. 137.88, $p > 0.10$. See Table B.4 in Appendix B.1 for more details). One possible explanation for this is that some subjects did not understand the mechanism used to elicit the R-range. We will examine this point more closely in subsection 3.3.3 and Appendix B.3. A further possibility is that the WTA-WTP gap of the group with no R-range relates to factors other than complexity (e.g., loss aversion, to which we return in Subsection 3.3.5). We offer the following observation regarding valuation complexity.

Observation 2. *The randomization option was chosen frequently. Among subjects who selected the randomization option at least once, the R-range as a measure of valuation complexity is significantly and positively correlated with the WTA-WTP gap.*

To summarize, Table 3 provides the distribution of subjects with/out each form of complexity and their associated average WTA-WTP gap, education, and performance on the matrix reasoning questions. About 53% of subjects exhibited both forms of complexity, and only 9% did not exhibit either form of complexity. Further, subjects with mechanism complexity exhibited significantly higher WTA-WTP gaps than those without, and those with high valuation complexity had higher WTA-WTP gaps than those with low valuation complexity ($p < 0.01$ in both tests). Finally, subjects' experience of complexity is related to cognitive ability as measured by education and matrix reasoning performance: Compared with subjects with both forms of complexity, subjects who did not exhibit either form of complexity had more education and performed better on the matrix reasoning questions ($p < 0.01$ in both tests).

	Mechanism complexity			No mechanism complexity		
	Valuation complexity			Valuation complexity		
	High	Low	No	High	Low	No
Proportions	23.9%	28.6%	24.3%	6.9 %	7.4%	8.8%
WTA-WTP gap	158.99 (280.38)	91.74 (221.95)	133.23 (282.43)	71.47 (176.85)	31.05 (131.84)	91.23 (153.88)
Education	3.86 (1.48)	3.73 (1.47)	3.73 (1.50)	4.34 (1.29)	4.22 (1.49)	4.81 (1.27)
MR performance	3.28 (1.72)	3.02 (1.80)	2.87 (1.83)	4.20 (1.54)	4.22 (1.79)	4.72 (1.46)

Table 3: The distribution of the subjects with/out each form of complexity, and their associated average WTA-WTP gap, education, and performance in the matrix reasoning questions. The high and low valuation complexity groups are defined by a median split of the R-range (100 points for subjects with mechanism complexity, and 118.75 points for those without) among the subjects with a positive R-range. Education level is measured on a scale from 1 to 6. Standard deviations are in parentheses. MR performance is the number of correctly answered matrix reasoning questions from 0 to 5.

3.3.3 Controlling for mechanism and valuation complexity

The analysis presented above suggests that the majority of subjects exhibited mechanism complexity, which represents a systematic bias. Furthermore, this bias carries over to the WTA and WTP of the lotteries. There is also substantial valuation complexity among subjects, and this relates systematically to the WTA-WTP gap. Since these two forms of complexity do not represent subjects' preferences in the classical sense, we now examine how much the observed WTA-WTP gap can be accounted for by mechanism complexity and/or valuation complexity.

Correcting the WTA-WTP gap for mechanism complexity: We correct the WTA-WTP gap for mechanism complexity in three ways: based on a theoretical analysis (Cason and Plott, 2014), statistically for deviations ($500 - WTP_{sure}$ and $WTA_{sure} - 500$), and statistically for deviations while allowing different slopes for positive and negative deviations.

We first correct the WTA-WTP gap for mechanism complexity based on a theoretical analysis that extends from Cason and Plott (2014). In this analysis, we assume that subjects mistake the switching value in the price list for the final trading price, as in the first-price auction (see Appendix B.5.1 for more details). Given the reported WTP_X and WTA_X for the lottery X , $X = L1$ and $L2$, the theoretical analysis suggests that WTP_X^* and WTA_X^* , after correcting for mechanism complexity, are computed as follows:

$$\begin{aligned}
 WTP_X^* &= \frac{100}{WTP_{sure} - 400} WTP_X \\
 WTA_X^* &= \frac{100}{600 - WTA_{sure}} WTA_X - \frac{WTA_{sure} - 500}{600 - WTA_{sure}} 1000.
 \end{aligned}$$

Note that there is no correction ($WTP_X^* = WTP_X$ and $WTA_X^* = WTA_X$) when $WTP_{sure} = WTA_{sure} = 500$. After the correction, we observe a significantly lower WTA (497.86 vs. 523.47, $p < 0.01$), a higher, although not significantly, WTP (421.70 vs. 411.51, $p > 0.10$), and both are moving closer to the CE of 470.84 from the neutral frame. As a result, the WTA-WTP gap is significantly smaller, and it is reduced by one-third after the correction (76.16 vs. 111.96, $p < 0.01$). The results for the two lotteries separately, presented in Table B.5 in Appendix B.1, are virtually the same.

A limitation of the theoretical correction is that it assumes that mechanism complexity affects WTA and WTP through the specific misperception of the first-price auction, and that this misperception works to the same extent for L1 and L2. However, neither assumption may hold in the experiment. To relax these assumptions, we proceed to correct WTA and WTP statistically. In each statistical correction model, the dependent variable is the WTA or WTP of the lottery. The independent variables are the mechanism complexity-related variable(s) described above, denoted as MC_t , $t = a, p$ (a for WTA and p for WTP). Explicitly, we first run the following two OLS regressions for each lottery:

$$WTA_X = I_{a,X} + \beta_{a,X} MC_a + \epsilon_a; \quad WTP_X = I_{p,X} + \beta_{p,X} MC_p + \epsilon_p, \quad (1)$$

where I is the intercept, $X = L1$ or $L2$, or the average of the two lotteries, the vector of coefficient(s) is denoted by β , and the subscript a or p denote WTA or WTP. These regressions capture the idea that how subjects reported their WTA and WTP may have been systematically affected by mechanism complexity, and the term βMC captures the component of the WTA or WTP that does not derive from classical preferences. Similar regression forms have been extensively used in financial research to account for the influence of various risk factors on excess asset returns (see e.g., Fama and French, 1992). With the estimated $\hat{\beta}$, we can correct the WTA and WTP for mechanism complexity as follows:

$$WTA_X^* = WTA_X - \hat{\beta}_{a,X} MC_a; \quad WTP_X^* = WTP_X - \hat{\beta}_{p,X} MC_p.$$

The variable(s) MC differs across the correction method. In the second correction, $MC_a = (WTA_{sure} - 500)$ and $MC_p = (500 - WTP_{sure})$. Similar to the results of the theoretical correction, after the statistical correction we observe a significantly lower WTA (497.95 vs. 523.47, $p < 0.01$), a significantly higher WTP (428.13 vs. 411.51, $p < 0.05$), and both moving closer to the neutral frame. Consequently, the WTA-WTP gap is significantly smaller, reduced by about 38% after the correction (69.82 vs. 111.96, $p < 0.01$).

In the third correction, we further allow for different slopes for positive and negative deviations. As Figure 10 suggests, the WTA of the lotteries is more strongly associated with positive WTA deviations of the degenerate lottery than with negative ones, and the WTP of the lotteries is more strongly associated with negative WTP deviations of the degenerate lottery than with positive ones. To account for this, we also include an interaction term with deviations multiplied by a dummy variable that indicates negative deviations

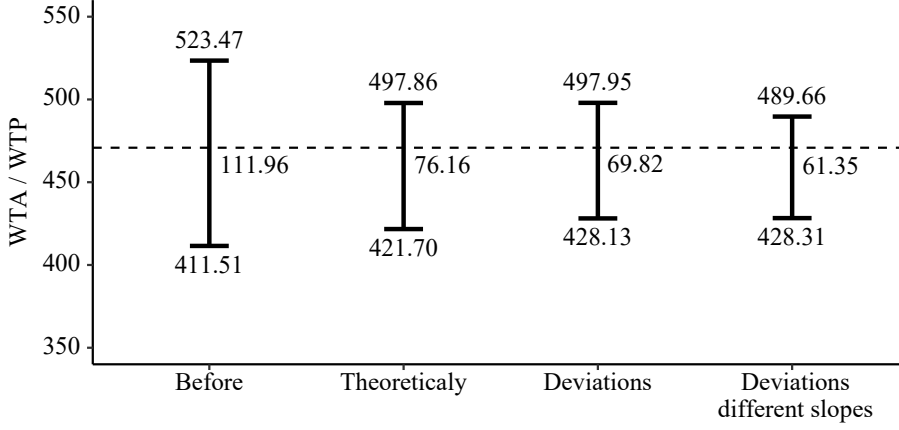


Figure 12: The WTA, WTP, and WTA-WTP gap for the lotteries before and after each of the three mechanism complexity correction methods. The dashed line is the average certainty equivalent of the lotteries from the neutral frame (470.84 points).

(D_{neg}). According to statistical model 1, $MC_a = (WTA_{sure} - 500, D_{neg}(WTA_{sure} - 500))$ and $MC_p = (500 - WTP_{sure}, D_{neg}(500 - WTP_{sure}))$. Consistent with the previous two corrections, in the third correction we observe a significantly lower WTA (489.66 vs. 523.47, $p < 0.01$), a significantly higher WTP (428.31 vs 411.51, $p < 0.01$), and a significantly smaller WTA-WTP gap, which is about 55% of the original level (61.35 vs. 111.96, $p < 0.01$).

Figure 12 provides an overview of the three corrections. As we can see, after using each correction method the WTA-WTP gap is reduced significantly, from about one-third to 45%. More details can be found in Table B.5 in Appendix B.1. In Appendix B.2, we perform an additional correction, where we use dummy variables for each deviation of the degenerate lottery's WTA or WTP from 500 to correct for the WTA, WTP, and WTA-WTP gap of the lotteries. The results are broadly similar, with a 40% reduction of the lotteries' WTA-WTP gap after the correction.

Controlling the WTA-WTP gap for valuation complexity: To control the WTA-WTP gap for valuation complexity, we follow the theoretical analysis in Appendix B.5.2 and control the WTA and WTP of the lottery separately for the corresponding R-range with the following regressions:

$$WTA_X = I_{a,X} + \beta_a R\text{-range}_X + \epsilon_a; \quad WTP_X = I_{p,X} + \beta_p R\text{-range}_X + \epsilon_p,$$

where I is the intercept, X is L1, L2, or the average of the two lotteries. With the estimated $\hat{\beta}_i, i = a, p$, the WTA^* and WTP^* , after controlling for valuation complexity, are calculated as follows:

$$WTA_X^* = WTA_X - \hat{\beta}_a R\text{-range}_X; \quad WTP_X^* = WTP_X - \hat{\beta}_p R\text{-range}_X.$$

By controlling for valuation complexity, we observe a change in the WTA, WTP, and WTA-WTP gap for the lotteries in the same direction as for the mechanism complexity corrections, although with a smaller magnitude. The WTA is significantly lower (514.92 vs. 523.47, $p < 0.01$), and the WTP is significantly higher (419.62 vs. 411.51, $p < 0.01$). Consequently, the WTA-WTP gap is reduced significantly by about 15% (95.30 vs. 111.96, $p < 0.01$).

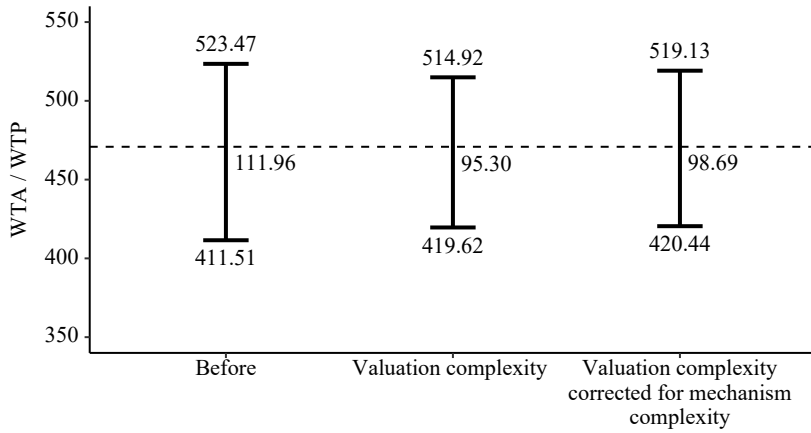


Figure 13: The WTA, WTP, and WTA-WTP gap for the lotteries before and after controlling for (mechanism complexity corrected) valuation complexity. The dashed line represents the average certainty equivalent of the lotteries in the neutral frame (470.84 points).

As discussed in section 3.2.3, there could be complexity with respect to the R-range price list, which could distort our measure of valuation complexity. Therefore, we perform an additional analysis by first correcting the upper bound and the lower bound of the R-range for mechanism complexity, and then using the mechanism complexity corrected R-range to control the WTA-WTP gap for valuation complexity. More details on these procedures can be found in Appendix B.3. Controlling for the mechanism complexity corrected R-range produces results comparable to those using the elicited R-range directly: The WTA decreases significantly (519.13 vs. 523.47, $p < 0.01$), and the WTP increases significantly (420.44 vs. 411.51, $p < 0.01$), with the average WTA-WTP gap reduced significantly by 12% (98.69 vs. 111.96, $p < 0.01$). Figure 13 provides an overview of the WTA, WTP, and WTA-WTP gap for the lotteries before and after controlling for (mechanism complexity

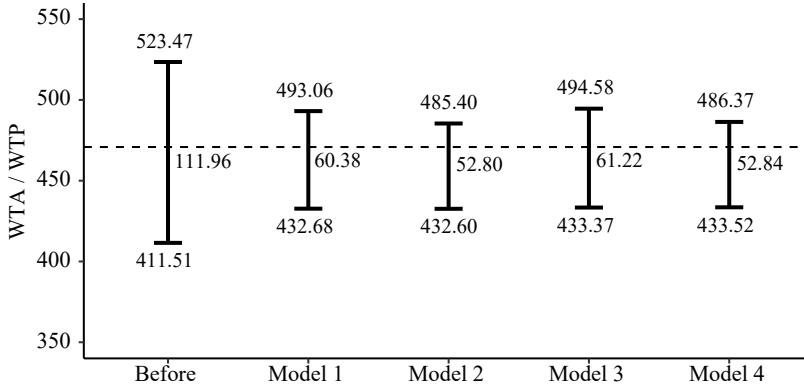


Figure 14: The WTA, WTP, and WTA-WTP gap for the lotteries before and after controlling for both mechanism complexity and valuation complexity. In Model 1 and 3 mechanism complexity is the deviation of the degenerate lottery's WTA or WTP from 500, and in Model 2 and 4 mechanism complexity additionally allows for different slopes for positive and negative deviations. In Model 1 and 2 valuation complexity is the elicited R-range, and in Model 3 and 4 valuation complexity is the R-range corrected for mechanism complexity. The dashed line represents the average certainty equivalent of the lotteries in the neutral frame (470.84 points).

corrected) valuation complexity. More details can be found in Tables B.11 and B.6 in Appendix B.1. Overall, our results suggest that valuation complexity plays an important role in the WTA-WTP gap.

Considering two forms of complexity simultaneously: We now examine the WTA, WTP, and resulting WTA-WTP gap for the lotteries accounting for mechanism complexity and valuation complexity simultaneously. We conduct four analyses. In Models 1 and 3, mechanism complexity is the deviation of the degenerate lottery's WTA or WTP from 500, and in Models 2 and 4, mechanism complexity also allows for different slopes for positive and negative deviations. In Models 1 and 2 valuation complexity is the elicited R-range, and in Models 3 and 4 valuation complexity is the R-range corrected for mechanism complexity.

As we can see from Figure 14, the WTA-WTP gap decreases substantially in all four models after accounting for both forms of complexity, ranging from 45% to 53%, with an average residual WTA-WTP gap around 50% of the original level.¹¹ More details can be found in Table B.7 in Appendix B.1. These results suggest that classical preference-based explanations of the WTA-WTP gap like loss aversion can not explain these results. For a

¹¹The combined explanation of the two forms of complexity is often smaller than the sum of the explanatory power of each individual form. Based on our earlier findings regarding the connection between subjects' matrix reasoning performance and their experienced complexity, this suggests that the two forms of complexity are not independent from each other, probably because they both depend on factors like cognitive ability. Indeed, the two forms of complexity are significantly correlated with each other, although the correlation coefficient is small (Spearman's $\rho = 0.1036$ between the R-range and the WTA-WTP gap of the degenerate lottery, $p < 0.01$).

more comprehensive understanding of the WTA-WTP gap, it is important to consider the two forms of complexity. We summarize our observation in the following statement.

Observation 3. *Accounting for mechanism complexity alone reduces the WTA-WTP gap by about 40%. Controlling for valuation complexity alone leads to a 15% reduction in the WTA-WTP gap. When considering both forms of complexity simultaneously, the WTA-WTP gap decreases by over 50%.*

3.3.4 Complexity versus preference-based explanations

The WTA-WTP gap is traditionally considered as reflecting non-standard preferences, and “[l]oss aversion has been the leading theory ” (Ericson and Fuster, 2014, p. 571). Below, we compare the explanatory potential of complexity to loss aversion.

We use the loss aversion measure from an independent study based on the LISS panel (Goossens and Knoef, 2022).¹² A unified notion of loss aversion would predict a significant and positive relationship between the WTA-WTP gap and loss aversion for risky bets involving gains and losses. Our sample has an overlap of 533 subjects with the sample of Goossens and Knoef (2022). The analyses below consider these subjects. The correlation between the WTA-WTP gap and the loss aversion coefficient is positive but not significant ($\rho = 0.067$, $p > 0.10$).¹³ This result is consistent with the findings of Chapman et al. (2024) and Fehr and Kübler (2022), who found no significant correlation between loss aversion and the endowment effect in representative samples from the United States and Germany, respectively. Remarkably, when categorizing subjects into groups with/out each form of complexity as in Table 3, we find a statistically significant correlation between loss aversion and the WTA-WTP gap in the group who did not experience mechanism or valuation complexity ($\rho = 0.507$, $p < 0.01$), while this correlation is close to zero and insignificant in the other three groups. Table B.8 further shows that the group that did not experience either form of complexity had a substantial WTA-WTP gap (122.78), comparable to the gap in the overall sample. This corroborates our earlier speculation that the substantial WTA-WTP gap in the group with no R-range could be partly due to loss aversion.

To examine quantitatively the role of loss aversion in the WTA-WTP gap, we follow the analyses in subsection 3.3.3 and use λ to control WTA, and $1/\lambda$ to control WTP separately. Table B.9 suggests that the WTA-WTP gap reduces significantly by about 25%. When we control loss aversion in addition to complexity, the inclusion of loss aversion significantly reduces the WTA-WTP gap further by about 10%.

¹²In this measure, subjects faced a price list with five rows. In each row subjects faced two lotteries *A* and *B*. The lottery *A* paid $-X$ and $+30$ with equal likelihood. The lottery *B* paid -2 and $+10$ with equal likelihood. Lottery *B* stayed the same across rows, while x in lottery *A* increased $(-22, -16, -11, -8, -6)$. Subjects were asked to indicate the row they would switch from *A* to *B*. The loss aversion coefficient is calculated as $\lambda = \frac{30-10}{-2-X}$. For example, $\lambda = 1$ when $X = -22$.

¹³We also checked the correlation between the WTA/WTP ratio or its square root and loss aversion, and the result is similar ($\rho = 0.002$, $p > 0.10$). Further, we do not find statistically significant correlations between loss aversion and mechanism complexity or valuation complexity (ρ close to zero, $p > 0.10$ for all tests).

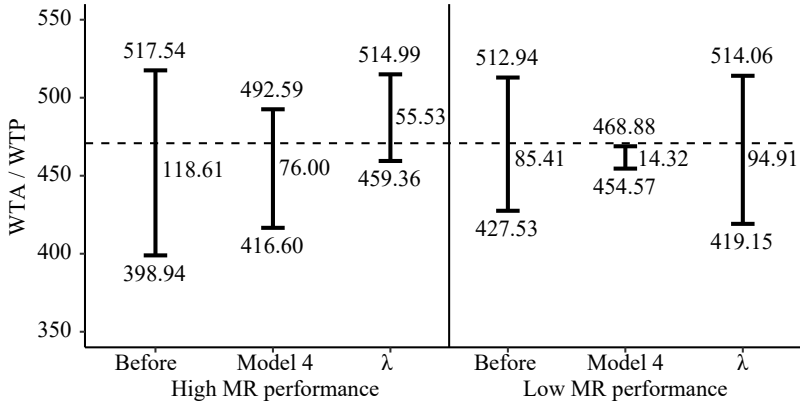


Figure 15: The WTA, WTP, and WTA-WTP gap for the lotteries in two subsamples based on subjects' performance on the matrix reasoning questions before and after controlling for mechanism complexity and the corrected valuation complexity as well as loss aversion (λ) in the overlapping subsample. The regression controlling for λ only considers loss aversion. Model 4, as defined in Figure 14, controls for both forms of complexity. The dashed line represents the average certainty equivalent of the lotteries in the neutral frame (470.84 points).

More generally, whatever form of non-standard preferences causes the WTA-WTP gap, they are more likely to show up in the choices of high cognitive ability subjects, as these subjects are likely to experience less complexity and make more informed decisions. Consequently, the explanatory potential of complexity for the WTA-WTP gap should be limited in this group of subjects. We thus split the sample into two subsamples based on subjects' performance (above or below the median) on the matrix reasoning questions. We then performed the same analysis in Subsection 3.3.3 for each subsample separately. Figure 15 illustrates the results. Accounting for the two forms of complexity significantly reduces the WTA-WTP gap in both subsamples, with the impact of complexity being greater in the subsample with low matrix reasoning performance: about 35% for those who performed equal to or above the median vs. 83% for those below the median on the matrix reasoning questions. This suggests that complexity plays a more important role for those with lower cognitive ability. As a comparison, when we use loss aversion instead of the two forms of complexity to account for the WTA-WTP gap, controlling for loss aversion alone reduces the gap by 53% for subjects with high matrix reasoning performance, but it even increases the gap by 11% among subjects with low matrix reasoning performance. More details are presented in Table B.15, Table B.16, and other related tables in Appendix B.4.¹⁴

¹⁴A perhaps surprising finding in Figure 15 (also tables B.15 and B.16) is that, after accounting for the two forms of complexity, the group with low matrix reasoning performance exhibited a lower WTA-WTP gap in comparison to the group with high matrix reasoning performance. However, this reduction should not be misinterpreted as evidence that these subjects would have naturally displayed a lower WTA-WTP gap in the absence of these two forms of complexity. If these subjects had a complete understanding of the elicitation mechanism and the valuation of the lotteries, they might have been able to make more informed decisions to reveal their loss

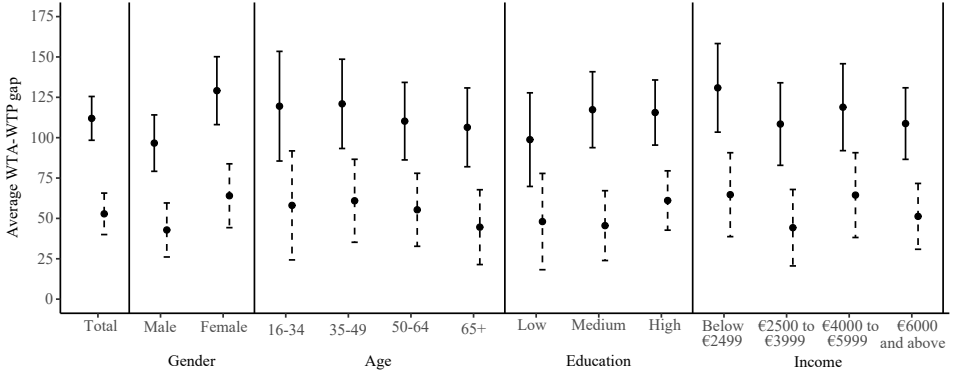


Figure 16: The WTA-WTP gap in various demographic groups. The solid lines are the elicited WTA-WTP gap. The dashed lines are after controlling for mechanism complexity and valuation complexity (Model 4 as defined in Figure 14). Error bars show the 95% confidence intervals.

Taking these results together, it suggests that not only is the WTA-WTP gap influenced by a variety of factors but also that these factors differ substantially among different subjects.

Observation 4. *Complexity plays a more important role in the subgroup with lower cognitive ability, and loss aversion plays a more important role in the subgroup with higher cognitive ability.*

3.3.5 Robustness checks

One concern is that the significant impact of mechanism complexity and valuation complexity on the WTA-WTP gap applies primarily to certain demographic groups, e.g., those with low education, and thus the insights cannot be generalized to all groups. Our representative sample allows us to examine this concern. We consider demographic groups based on gender (male, female), age (16-34, 35-49, 50-64, and 65 or above), education (based on the categorization of Statistics Netherlands, with low for primary or pre-vocational education, medium for pre-university or vocational education, and high for higher vocational or university education), and household monthly income (< 2499, 2500 – 3999, 2500 – 3999, and 6000 or above). Table B.1 in Appendix B.1 reports the percentages of subjects in each demographic group. Figure 16 displays means and 95% confidence intervals of the WTA-WTP gap in various demographic groups, both before and after controlling for mechanism complexity and valuation complexity. We make two main observations. First, the WTA-WTP gap exhibits some heterogeneity but exists in all demographic groups. Second, controlling the WTA-WTP gap for mechanism complexity and valuation complexity has a significant and

aversion-driven WTA-WTP gap like those with high matrix reasoning performance. Chew et al. (2018) measured subjects' comprehension of the experiment and showed that high-comprehension subjects exhibited ambiguity aversion while low-comprehension subjects appeared to be ambiguity neutral by choosing randomly.

substantial effect in all demographic groups. We summarize the results of this subsection as follows:

Observation 5. *The observed roles of mechanism and valuation complexity hold for different subgroups based on age, gender, education, and income.*

3.4 Concluding Remarks

Using a large representative sample of the Dutch population, we show that mechanism and valuation complexity contributed substantially to the WTA-WTP gap of the monetary lotteries elicited using a price list mechanism. More specifically, we find that mechanism complexity, as measured by deviations from a known preference, correlates significantly with the WTA and WTP of monetary lotteries. Meanwhile, valuation complexity, as measured by the R-range, relates systematically to the WTA-WTP gap of the monetary lotteries. A series of quantitative analyses demonstrate that mechanism complexity alone has the potential to account for about 40% of the observed WTA-WTP gap for the lotteries, valuation complexity alone for about 15%, and the two forms of complexity together for about 50%.

Our observations can be accommodated by linking complexity and heuristics. Complexity naturally arises in studies that involve experiments or surveys. When decision-makers face complexity, they often rely on heuristics to make decisions (Simon, 1955; Kahneman and Tversky, 1974; Gigerenzer et al., 1999; Ortoleva, 2010; Iyengar and Kamenica, 2010). Moreover, most heuristics tend to bias decisions systematically in certain directions.¹⁵ In our experiment, the buy-low and sell-high heuristic could reduce the WTP and increase the WTA, and thus result in the WTA-WTP gap of the lotteries and the degenerate lottery. In addition, recent studies have increasingly recognized that subjects may find choice objects difficult to value, and the complexity of valuing choice objects may be responsible for some behavioral biases that were previously thought to reflect underlying preferences (Oprea, 2024; Enke et al., 2025). When subjects are uncertain about the precise valuation of a choice object and have to decide, they may act cautiously, which can also be viewed as a heuristic (as with bad-deal aversion; e.g., Weaver and Frederick, 2012). Cerreia-Vioglio et al. (2024) demonstrated that valuation uncertainty and caution together can explain the endowment effect, loss aversion, and the certainty effect. Consistent with this idea, a part of the WTA-WTP gap in our setting arises because valuation uncertainty and caution reduce the WTP and increase the WTA, as demonstrated by the link between the R-range and the WTA-WTP gap.

Although the framework that links complexity and heuristics has gained increasing attention, it faces several significant challenges that are also present in the current study.

¹⁵For example, the heuristic of choosing in the middle could bias choices in the price list at the center of the list (Bosch-Domènech and Silvestre, 2013). The heuristic of maintaining the status quo unless offered a strictly better alternative could generate the endowment effect and the default bias (Masatlioglu and Ok, 2005; Sautua, 2017). The heuristic of anchor-and-adjustment could induce insufficient responses to information, resulting in anchoring effects and the over/underreaction of prices in financial markets (Hogarth and Einhorn, 1992).

First, there is no well-established approach to reveal and quantify the complexity that subjects perceive. Consistent with the method in Cettolin and Riedl (2019), Agranov and Ortoleva (2023), and Halevy et al. (2023), this study employed randomization to measure valuation complexity. However, there are alternative methods. For example, some studies have focused on subjects' decision confidence (Dubourg et al., 1994; Butler and Loomes, 2007; Cubitt et al., 2015; Enke et al., 2025) or cognitive uncertainty (Enke and Graeber, 2023). Others measure valuation complexity with respect to lotteries and temporal payments based on subjects' calculation errors in algebraic choices, which serve as no-risk-no-time mirrors of risky or temporal choices (Oprea, 2024; Enke et al., 2025; Enke and Shubatt, 2023). These elicitation methods reflect different conceptualizations of complexity, such as incomplete preferences, preference imprecision, preference uncertainty, cognitive noise, and cognitive uncertainty. While there is evidence that different measures of valuation complexity are significantly correlated with each other and thus may have similar behavioral contents (Arts et al., 2024a), it is desirable to have a unified framework to precisely delineate the conceptual disparities and behavioral consequences of these distinct methods.

It can also be difficult to measure mechanism complexity. A popular approach for eliciting mechanism complexity, which we adopted here, is to use objects with known preferences and measure complexity with deviations from these known preferences (Cason and Plott, 2014). However, the objects with known preferences differ from actual choice objects. This implies that we do not directly measure mechanism complexity for the actual choice objects. This difference is important when mechanism complexity could interact with valuation complexity, e.g., mechanism complexity may be more severe when choice objects become more complex (stochastic rather than deterministic, as in Martínez-Marquina et al., 2019). This implies that our estimate of the contribution of mechanism complexity to the WTA-WTP gap is likely conservative. The potential interaction between the two forms of complexity can also complicate the elicitation of valuation complexity, as it takes place in an elicitation mechanism. In our investigation, we find that our subjects faced substantial mechanism complexity when we elicited their R-ranges, although it appears that mechanism complexity did not systematically bias the impact of R-range on the WTA-WTP gap. Further, there is evidence that our mechanism for eliciting valuation complexity underestimates it, likely because a large group of subjects found the elicitation mechanism too complex and opted for simplistic choice rules (Oprea, 2020; Banovetz and Oprea, 2023). Research in this area has been rare, and future investigation is needed.

Second, it is difficult to identify the heuristics subjects use when facing complexity. There are many heuristics, and their application may depend on the decision situation and individual characteristics. In price list elicitation, subjects are likely to rely on several heuristics, including the central tendency (Bosch-Domènech and Silvestre, 2013), relying on a randomly generated anchor (Ariely et al., 2003), emphasizing the first choice in the choice list (Hermann and Musshoff, 2016), the buy-low and sell-high heuristic (Kahneman et al., 1982), or caution (Cerreia-Vioglio et al., 2024). We controlled for the central tendency and anchoring heuristics by using the same ranges of sure payments on the WTA and

WTP questions and randomizing the order (ascending or descending) of the sure payments in the price list. Additionally, the WTA-WTP gap of the degenerate lottery is consistent with the buy-low and sell-high heuristic but not with caution. While these results suggest that the buy-low and sell-high heuristic may be one important driving force behind the WTA-WTP gap of the lotteries, we leave it for future studies to understand the effects of multiple heuristics on the elicitation of WTA and WTP.

A further challenge relates to separating heuristics from noisy responses. For example, while the tendency to switch in the middle of the choice lists and to choose the middle of the budget lines is in line with heuristic rules, it can also be viewed as noisy responses (Choi et al., 2006; Enke and Graeber, 2023; Halevy and Mayraz, 2022). In our experiment, we find that many subjects had a negative WTA-WTP gap for the degenerate lottery. Further analysis indicates that these negative WTA-WTP gaps may result from noisy responses. However, without additional information, we are unable to conclusively differentiate noisy responses from heuristics. Consequently, our control of complexity on preference inference is likely to be too conservative.

To sum, our study adds to the emerging literature on the effect of complexity on choice behavior and points to the need to separate preferences from heuristics and noise. Earlier works in behavioral economics have provided substantial insights into the identification of heuristics in various situations (Simon, 1955; Kahneman and Tversky, 1974). Recent studies have provided more concrete evidence and formal analysis with respect to differentiating preferences from noise and biases (Khaw et al., 2021; Woodford, 2020; Frydman and Jin, 2021; Vieider, 2021; Cerreia-Vioglio et al., 2024; Oprea, 2024; Enke and Graeber, 2023; Enke et al., 2025). These studies highlight the importance of unraveling complexity as a fundamental aspect of cognitive processes toward a better understanding of decision-making.

Choice (in)consistency and household wealth

Abstract: We examine the relationship between choice inconsistency and household wealth by combining experimental data with administrative records. In a representative Dutch sample, we reveal choice inconsistency through an experiment and quantify its severity using the Money Pump Index (MPI), which reflects potential monetary losses from inconsistent choices. Linking MPI to household wealth from Statistics Netherlands, we find a significant association: a one standard deviation increase in MPI corresponds to a 14.8% reduction in household wealth. This relationship holds after accounting for socioeconomic factors, cognitive ability, and risk attitudes, and its magnitude doubles when using self-reported wealth data.

This chapter has been published as: Arts S, Ong Q, Qiu J, Vyrastekova J. Choice (in)consistency and household wealth. *Journal of Risk and Uncertainty*. 2026. doi:10.1007/s11166-026-09499-5

4.1 Introduction

In the last few decades, economic studies have documented various types of choice inconsistency in the field, surveys, and laboratory experiments (see e.g. Seidl, 2002; Rieskamp et al., 2006; Barberis, 2013; Keren and Wu, 2015, for some reviews). Associated with the discovery of these types of choice inconsistency, there are developments in non-standard decision theories that try to incorporate them (Machina, 1982; Schoemaker, 1982; Starmer, 2000; Barberá et al., 2013), discussions on how to perform welfare analysis when a well-behaved utility function cannot capture choices (Koszegi and Rabin, 2008; Bernheim and Rangel, 2009; Apesteguia and Ballester, 2015; Bouacida and Martin, 2021), and policy interventions that can improve individual and social welfare in the face of such inconsistent choices (Thaler and Sunstein, 2008; Benartzi et al., 2017; Mertens et al., 2022).

Despite the considerable interest in choice inconsistency, there remains a notable gap in investigating how choice inconsistency observed in various experimental settings relate to real-life outcomes. Understanding this relationship is important to fully appreciate the significance of earlier findings regarding the prevalence of choice inconsistency and theoretical developments. If choice consistency indeed plays a key role in economic interactions, as suggested by many economic studies, then violations of this consistency should result in suboptimal economic outcomes, with demonstrable impact on wealth accumulation. Conversely, if individuals who behave consistently have similar economic outcomes as those exhibiting inconsistent choices, the observation of choice inconsistency, along with the related theoretical advancements and policy interventions, would be less meaningful.

As inconsistent choices violate some normative conditions that are regarded fundamental criteria for rational decision-making, they are often seen as indicators of poor decision-making quality.¹⁶ Formally, choice inconsistency is related to real economic outcomes via the concept of economic selection in social settings. According to this idea, individuals who make inconsistent choices may be susceptible to exploitation when interacting with arbitrageurs who possess classical economic preferences, as illustrated in “Dutch books” (see e.g., Ramsey, 1931; Raiffa, 1968; De Finetti, 1974; Eells, 1982; Yaari, 1985; Green, 1987; Kelsey, 1995). Consequently, we should observe a clear association between choice inconsistency and the actual economic outcomes of the decision-makers.

There are opposing views on this association. First, there are concerns about the external validity of the choice inconsistency measured in experiments. Some argue that decision errors arise because experimental subjects face unfamiliar situations and mistakenly rely on heuristics or decision rules that they developed in different contexts (Tversky and

¹⁶For example, Allais-type choices (Blavatskyy et al., 2022) and preference for lottery mixing (Feldman and Rehbeck, 2022) violate the independence axiom. Postponing a decision at a cost (Danan and Ziegelmeyer, 2006), refusing to make a decision (Gerasimou, 2017; Costa-Gomes et al., 2022), or altering their choices when decisions are repeated (Agranov and Ortoleva, 2017) are consistent with the violation of the completeness axiom. Individuals’ choices could also violate transitivity (Tversky, 1969; Grether and Plott, 1979; Loomes et al., 1991; Kalenscher et al., 2010) or even first-degree stochastic dominance (Fishburn, 1988; Blavatskyy, 2009). See Subsection 4.3.5 for a more systematic discussion.

Kahneman, 1974; Gigerenzer and Selten, 2001; Kahneman, 2011), or because they are not sufficiently motivated to make serious decisions as the stakes in experimental settings are often small (Smith and Walker, 1993; Camerer and Hogarth, 1999; Holt and Laury, 2002; Levitt and List, 2007). If this is the case, subjects may not make the same errors in real-life situations as they do in experiments, and therefore the choice inconsistency observed in experiments may not translate to real economic consequences. Second, the empirical evidence on choice consistency may not be robust due to methodological challenges in data collection, modeling and analysis. These challenges can cause individuals who adhere to the expected utility theory (EUT) to appear as if they are violating it (Kelsey and Milne, 1997), transitive preferences to be misinterpreted as violations of transitivity (Regenwetter et al., 2011), and the utility maximization theory to be incorrectly rejected (Aguiar and Kashaev, 2020). Third, some strategies that are proposed to exploit individuals require additional assumptions that are not justified, for example, the application of “transitive thinking to an intransitive world” (Fishburn, 1991, p. 118), the assumption of consequentialism that is rejected by many non-expected utility models (Machina, 1989), or the assumption of monopoly power by the arbitrageur that does not hold in a competitive market (Laibson and Yariv, 2007). Finally, individuals who are aware of their choice inconsistency can sometimes avoid economic exploitation by adhering to certain decision rules. For example, individuals with intransitive revealed preferences resulting from incomplete underlying preferences can avoid economic exploitation by maintaining the status quo (Mandler, 2005) or consciously selecting a specific option among the non-comparable options instead of randomizing among them (Danan, 2010). Taken together, these studies suggest that the relationship between choice inconsistency and economic outcomes in real life is not straightforward and requires careful empirical investigation.

Compared to significant theoretical interests in the relationship between choice inconsistency and economic outcomes, empirical evidence remains limited. Investigating this relationship empirically faces two major challenges. First, it is difficult to gain access to subjects’ real-life economic data – such as household wealth, income, or financial activities – in a large and heterogeneous sample, in combination with experimental data. Most studies on choice consistency have relied on self-reported information, which are often constrained in detail and whose accuracy cannot be assured. Second, revealing choice consistency and quantifying its severity are not straightforward. When choice inconsistency is revealed via well-known choice anomalies (e.g., Allais paradox), the observed prevalence can be sensitive to the chosen experimental parameters (Lucia, 2023): a subject may exhibit choice anomalies in one set of parameters but not in another. In addition, when experimental tasks are overly complex, choice inconsistency may instead reflect confusion or the use of heuristics rather than true preferences, especially among subjects with lower education background (Miao et al., 2024; Arts et al., 2024b).

In this study, we address these challenges by identifying choice inconsistency via a simple task in a representative sample of the Dutch population and quantifying their severity. Crucially, we link our experimental data to the subjects’ household wealth obtained from the

administrative records of Statistics Netherlands (CBS: Centraal Bureau voor de Statistiek). We focus on household wealth because the accumulation of household wealth reflects the cumulative outcomes of numerous decisions, which may seem insignificant in isolation but can collectively lead to substantial differences among households. Our analyses suggest a statistically significant and economically important relationship between the choice inconsistency observed in the experiment and the household's wealth.

Specifically, we administered an experiment on subjects in the LISS (Longitudinal Internet studies for the Social Sciences) panel, which comprises a representative sample of the Dutch population. In the experiment, subjects faced a random sequence of decisions between a fixed lottery and a varying sure payment, with each decision presented on a separate screen. For each decision, subjects could choose the lottery, choose the sure payment, or randomize by letting the computer choose with equal probability either the lottery or the sure payment. Subjects' choices were considered inconsistent when they chose the dominated option, switch multiple times between the lottery and the sure payment, and/or randomized more than once across the series of choices (Bruner, 2011; Chew et al., 2022). We quantified the severity of choice inconsistency by calculating the monetary loss that subjects with inconsistent choices could incur when interacting with arbitrageurs with classical preferences, calling this measure the Money Pump Index (MPI, as in Echenique et al., 2011).¹⁷ Using administrative records from Statistics Netherlands, we related the MPI to household wealth.

We found that the vast majority (81%) of subjects made inconsistent choices, in which two-thirds of the participants switched multiple times between the lottery and the sure payment (66%), nearly two-thirds chose the randomization option more than once (64%), and around one in ten chose the dominated option (9%). The mean MPI is 28% of the maximum exploitation potential, with substantial heterogeneity among subjects (standard deviation: 25% of the maximum exploitation potential). Consistent with the existing literature (Echenique et al., 2011; Choi et al., 2014), younger, higher-educated, and higher-income subjects displayed greater levels of consistency in terms of MPI than older, lower-educated, and lower-income subjects. Further, subjects with low MPI exhibited higher cognitive ability. Most importantly, we found that MPI has a robust relationship with household wealth. Our baseline OLS regression model, controlling for socioeconomic characteristics, suggests that one standard deviation increase in MPI is associated with a 14.5% decrease in household wealth. This relationship remained robust in various subsamples, and stayed significant after controlling for cognitive ability, risk attitudes, and alternative measures of decision-making quality. Interestingly, when self-reported wealth is used instead of administrative wealth data, the relationship becomes even stronger, likely due to the interaction

¹⁷Generally, there are two approaches to measure the severity of choice inconsistency. The first is to count the number of violations (Houtman and Maks, 1985; Swofford and Whitney, 1987; Famulari, 1995; Bernheim and Rangel, 2009; Apesteguia and Ballester, 2015; Nishimura, 2018). The second is to measure the monetary loss due to the inconsistent choices (Afriat, 1973; Varian, 1990; Echenique et al., 2011; Dean and Martin, 2016). Our MPI measure is in line with the second approach.

between biases in self-reported wealth and MPI. Overall, our findings suggest that, despite the artificial setting and small incentives, subjects' choice inconsistency in the experiment provides important insights into their economic behavior in real life.

Our study builds on the limited but growing literature on the relationship between decision-making ability and economic outcomes. Previous research has shown that choice inconsistency, as measured by Allais-type behavior, correlates with loan default or late repayment (Herrmann et al., 2017), and less educated and unemployed individuals are more likely to exhibit such behavior (Huck and Müller, 2012). Measuring choice inconsistency with the critical cost efficiency index (CCEI; Afriat, 1972), Choi et al. (2014) found that households of higher CCEI had greater self-reported household wealth, and Carvalho et al. (2024) found that borrowers of lower CCEI were more likely to take high-cost loans. Using MPI calculated from budgetary choices, Echenique et al. (2011) found that the severity of violations of general axiom of revealed preferences is correlated to demographic variables such as age, education, and income. There were also attempts to measure choice consistency by estimating the decision error parameter in stochastic choice functions. These studies found that individuals with high decision error were less likely to save and more likely to take up informal credit (Jacobson and Petrie, 2009), and tended to make less economically efficient decisions (Castillo et al., 2018). Bodnaruk and Simonov (2016) measured choice inconsistency with loss aversion and showed that fund managers with higher aversion to losses chose portfolios with lower downside risk, increased their risk-taking more in response to poor past performance, and had lower overall performance. When examining the relationship between time discounting and wealth inequality, Epper et al. (2020) included a control for choice inconsistency (a binary variable indicating non-monotonic choices) and found no effects of choice inconsistency on wealth.

There is also a substantial body of literature examining other measures of decision-making ability, often grounded in cognitive abilities and financial or mathematics knowledge. Cognitive abilities are found to be negatively associated with financial literacy and the ability to manage money effectively (Finke et al., 2016), positively related to the willingness-to-ask and willingness-to-pay gap in valuing annuities (Brown et al., 2017), and positively related to the ability to form consistent inflation expectations as well as to adjust consumption plans to changes in inflation expectations (D'Acunto et al., 2023). Numeracy is also found to be related significantly to wealth (Estrada-Mejia et al., 2016) and financial literacy, explaining the gender gap in pension savings (Preston and Wright, 2023).

Our study contributes to the investigation of the implications of inconsistent choices on real-life economic outcomes in two ways. First, we complement previous studies that mainly relied on self-reported household wealth, which are prone to measurement errors and underreporting (Fagereng et al., 2020), with administrative records. While we find a statistically significant negative correlation between choice inconsistency and household wealth from both data sources, our findings caution that choice inconsistency may interact with the underreporting bias, and relying on self-reported wealth may lead to an overestimation of the correlation with choice inconsistency. Further, while there

are earlier studies that correlated choice inconsistency with economic outcomes from administrative records, they tended to focus on specific financial activities, such as purchases from a single retailer, payments on credit cards, or a particular form of savings (Jacobson and Petrie, 2009; Bodnaruk and Simonov, 2016; Herrmann et al., 2017). We complement these studies by showing that there is a robust relationship between choice inconsistency and general real-life outcomes. Second, our experimental approach has several advantages over earlier approaches. In terms of elicitation, our task is simple to understand and robust to the calibration of experimental parameters, unlike decision tasks in classical anomalies (e.g., Allais paradox) and budgetary choices. Further, our task not only naturally links choice inconsistency to financial outcomes by illustrating how individuals could be exploited, it also provides insights into some non-EUT models (Machina, 1989). As a quantitative measure, the MPI allows more detailed comparisons across subjects by quantifying the level of choice inconsistency for each subject, rather than grouping them based on broad categories of inconsistent choices.

Our paper proceeds as follows. Section 4.2 explains the experimental sample, design, and the administrative records. Section 4.3 reports results. In Section 4.4, we conclude our study and discuss promising future research directions.

4.2 Experimental sample, design, and data

Our study combines experimental data collected from a representative household panel with their administrative records provided by Statistics Netherlands. We first describe the sample. Then, we explain the three experimental tasks in which we elicit choice inconsistency, cognitive ability, and risk attitudes. Finally, we describe the household wealth information and other socioeconomic variables from the administrative records used in our regression analyses.

4.2.1 Experimental sample

The experiment involved 620 participants of the LISS (Longitudinal Internet studies for the Social Sciences) panel, a representative sample of the Dutch population managed by the non-profit research institute Centerdata (Tilburg University, the Netherlands).

Due to the lack of access to the administrative records at the time of the experiment, we invited only participants who had completed the LISS Core Study on “Economic situation: assets” in 2020 for our experiment. As younger participants were less likely to complete this module of the LISS Core Study, our sample has a lower proportion of younger subjects (ages of 16 to 34 and 35 to 49) and a higher proportion of older subjects (ages of 65 and above) compared to the full LISS sample. 620 LISS participants who fulfilled both criteria completed our experiment. The first two columns of Table 4 show that our experimental subjects tended to be older, more likely to be retired, and more likely to be living in households without children.

	LISS Panel participants	Experiment subjects	Final sample
Sample size	8629	620	572
Female	51.44	48.06	48.08
Age			
16-34	27.41	12.42	12.41
35-49	20.62	15.48	15.91
50-64	24.87	25.32	25.35
65 and above	27.11	46.77	46.33
Education level			
Low	27.21	24.52	25.17
Medium	34.44	30.16	29.90
High	37.90	45.16	44.76
Occupation			
Paid work	51.86	42.90	43.36
Housekeeping	7.05	7.26	7.34
Retired	21.62	38.87	37.94
Others	19.47	10.97	11.36
Household composition			
Partner	71.76	64.19	64.86
Child at home	43.62	22.74	22.90

Table 4: Composition of all LISS panel members aged 16 and above (column 1), LISS panel members who participated in our experiment (column 2), and LISS panel members who participated in our experiment with accessible administrative records (column 3). Comparing our sample to the complete panel, we see that older/retired individuals are over-represented in our sample. This is due to our restriction on participants with more complete financial information. In addition, our sample is somewhat higher educated, and more likely to live alone compared to the whole panel.

Of the 620 subjects who participated in our experiment, 572 gave permission to link the experimental data to their administrative records. Hence, our analyses involving data from the administrative records are based on these 572 participants. Table 4 shows that the demographic composition of our final sample is comparable to that of our experimental sample.

4.2.2 Experimental design and data

Experimental payoffs were denoted in points instead of euros, with an exchange rate of 1000 points for €5. The experiment consisted of three tasks.

Choice inconsistency: In the choice inconsistency task, subjects faced a random sequence of decisions involving a lottery and a sure payment. Each decision was made on a separate screen. Across decisions, the lottery was fixed, and the sure payment took one of the 20 values (0, 100, 200, 250, 300, 325, 350, 375, 400, 425, 450, 475, 500, 525, 550, 600, 700, 800, 900, 1000 points). In each decision, subjects could choose the lottery, choose the sure payment, or randomize between the lottery and the sure payment by letting the computer toss a virtual coin. Figure C.4 in Appendix C.5 provides an example of a decision screen.

Choice options		
Option 1	Option 2	Option 3
Choose lottery x	Let the computer decide	Choose sure payment y
Lotteries		
L1: 50%, 900 points; 50%, 100 points.		
L2: 20%, 975 points; 30%, 850 points; 30%, 150 points; 20% 25 points.		
Sure payments y		
0, 100, 200, 250, 300, 325, 350, 375, 400, 425, 450, 475, 500, 525, 550, 600, 700, 800, 900, 1000.		

Table 5: Summary of the choice inconsistency task, the lotteries, and sure payments. Subjects made one choice per decision screen. The order of the lotteries and the sure payments (within each lottery) was randomized.

We elicited decisions on two lotteries with different payoff structures. Lottery 1 (L1) gave subjects a 50% chance of gaining 900 points and a 50% chance of gaining 100 points. Lottery 2 (L2) has the same expected value but offers more outcomes: a 20% chance to gain 975 points, a 30% chance to gain 850 points, a 30% chance to gain 150 points, and a 20% chance to gain 25 points. The sure payment of 1000 points dominates the lottery and the sure payment of 0 points is dominated by the lottery. Subjects completed twenty decisions for one lottery before they made the decisions for the other lottery. The order of the lotteries was randomized. Table 5 provides the summary of this task.

Approximately 81% of our subjects made at least one inconsistent choice in decisions involving either lottery. About two-thirds of the subjects switched multiple times between the lottery (L1 or L2) and the list of sure payments. A similar proportion of subjects chose the randomization option more than once (64% for L1 and 65% for L2). Around 9% of the subjects chose a dominated option in decisions involving either lottery.

While categorizing subjects into types offers some useful insights (see Appendix C.2 for more details), it does not account for the severity of different patterns of inconsistent choices. To capture this severity, we calculate the expected loss that subjects could incur due to inconsistent choices. Following Echenique et al. (2011), we call this measure money pump index (MPI). Take Subject 1 in Figure 17 as an example. Since Subject 1 chooses the randomization option twice, there are four possible states after the resolution of the computerized coin toss, illustrated as State 1 to 4 in Figure 17. Consider first State 1, in which Subject 1 chooses the lottery over 475 points and is willing to exchange the same lottery for 400 points. Recognizing this inconsistency, an arbitrageur can offer the lottery to the subject for 475 points and then buy it back for 400 points. This trade cycle results in a sure economic loss of 75 points for the subject. Since this state occurs with probability $1/4$, the expected loss is 18.25 points. By performing similar computations for the other three states and aggregating to obtain the total expected loss, we obtain a MPI of 56.25 points for Subject 1.

Subject 1					
● The lottery			○ Let the computer decide	○ 375 points	
○ The lottery			○ Let the computer decide	● 400 points	
● The lottery			○ Let the computer decide	○ 425 points	
○ The lottery			● Let the computer decide	○ 450 points	
○ The lottery			● Let the computer decide	○ 475 points	
○ The lottery			○ Let the computer decide	● 500 points	
State 1			State 2		
● The lottery	○ Let the computer decide	○ 375 points	● The lottery	○ Let the computer decide	○ 375 points
○ The lottery	○ Let the computer decide	● 400 points	○ The lottery	○ Let the computer decide	● 400 points
● The lottery	○ Let the computer decide	○ 425 points	● The lottery	○ Let the computer decide	○ 425 points
● The lottery	○ Let the computer decide	○ 450 points	○ The lottery	○ Let the computer decide	● 450 points
● The lottery	○ Let the computer decide	○ 475 points	○ The lottery	○ Let the computer decide	● 475 points
○ The lottery	○ Let the computer decide	● 500 points	○ The lottery	○ Let the computer decide	● 500 points
State 3			State 4		
● The lottery	○ Let the computer decide	○ 375 points	● The lottery	○ Let the computer decide	○ 375 points
○ The lottery	○ Let the computer decide	● 400 points	○ The lottery	○ Let the computer decide	● 400 points
● The lottery	○ Let the computer decide	○ 425 points	● The lottery	○ Let the computer decide	○ 425 points
● The lottery	○ Let the computer decide	○ 450 points	○ The lottery	○ Let the computer decide	● 450 points
○ The lottery	○ Let the computer decide	● 475 points	● The lottery	○ Let the computer decide	○ 475 points
○ The lottery	○ Let the computer decide	● 500 points	○ The lottery	○ Let the computer decide	● 500 points

Figure 17: Example of how the MPI is computed for a subject who made multiple switches and chose the computerized coin flip. There are 4 possible states associated with the two computerized coin flip. The economic loss that the subject could incur is 75, 25, 50, and 75 points in State 1 to State 4, respectively. The MPI for this subject is the weighted average of the losses across the four states: $(75 + 25 + 50 + 75)/4 = 56.25$ points

We compute subjects' MPI for L1 and L2 and use the mean MPI across the two lotteries as a measure of choice inconsistency to mitigate measurement errors. We find that the mean MPI across the two lotteries is 287.27, with substantial heterogeneity in the sample (standard deviation = 255.48). Figure 18 presents the histogram of mean MPI. As L2 is more complex than L1 (Arts et al., 2024a), subjects had higher MPI in L2 than in L1 (Mean: 294.40 for L2 vs. 280.13 for L1), though the difference was only significant at the 10% level (Wilcoxon signed rank test).

Cognitive ability: In the cognitive ability task, subjects answered six matrix reasoning questions from the International Cognitive Ability Resource (ICAR, Condon and Revelle, 2014). In each question, subjects faced a 3×3 matrix with a missing image and had to select one out of six images to complete the matrix. See Figure C.6 in Appendix C.5 for

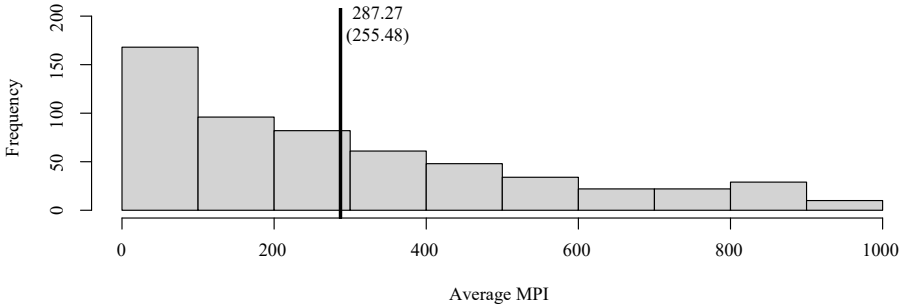


Figure 18: Distribution of the average MPI across the two lotteries. Mean and (SD) are shown at the vertical line.

an example. Subjects received 100 points for each correct answer. Cognitive ability was measured by the proportion of correctly answered matrix reasoning questions. On average, subjects correctly answered 3.36 out of 6 matrix reasoning questions (standard deviation: 1.75).

Risk attitudes: To elicit risk attitudes, we presented subjects with two price lists with twenty rows (Holt and Laury, 2002). Each price list contained a fixed lottery in one column and monotonically increasing or decreasing values of the sure payment in the other column. We used the same lotteries and sure payment values as those in the choice inconsistency task. Subjects indicated the row where they switched from preferring the lottery to preferring the sure payment for the first time. The price list was then automatically completed based on this choice. The certainty equivalent of the lottery was determined by taking the midpoint of the sure payment values in the two rows where subjects switched from preferring the lottery to preferring the sure payment. We use the average certainty equivalents of the two lotteries as a measure of risk attitudes.

Experimental procedure: Subjects completed the experimental tasks in the same order: the choice inconsistency task → the cognitive ability task → the risk attitude task. Subjects received a flat fee of €3 for participation. In addition, every subject had a 1 in 10 chance of receiving extra payment based on their decisions in one of the three tasks. If the choice inconsistency task or the risk elicitation task was randomly selected for the additional payment, one of the decisions in the task was chosen at random and subjects were paid based on that decision. If the cognitive ability task was randomly selected for the additional payment, subjects were paid based on the overall performance in that task. On average, the subjects selected for additional payment received 4.12 euros on top of the flat fee. Subjects took an average of 15 minutes to complete the experiment.

	Our sample	Population
Gross household wealth distribution (×1000)		
5th percentile	−13.39	−23.36
25th percentile	21.81	5.68
50th percentile	145.65	90.61
75th percentile	329.02	273.77
95th percentile	731.05	900.07
Gross household income distribution (×1000)		
5th percentile	19.92	19.49
25th percentile	41.40	41.61
50th percentile	63.86	77.20
75th percentile	101.57	118.84
95th percentile	171.00	204.84

Table 6: Distribution of gross annual household income and household wealth obtained from the administrative records of Statistics Netherlands in our sample (column 2) and in the total population (column 3).

4.2.3 Household wealth and socioeconomic variables

The experimental data are linked at the individual level to administrative data at Statistics Netherlands. We use administrative records as of January 1st 2020 for main analyses, and use administrative records from 2018 and 2019 for robustness checks. The administrative household wealth data provide details on a household's assets and debts. Assets include current and saving accounts, financial assets, business assets, and real estate. Debts include mortgages, study loans, and other liabilities. The administrative records also provide detailed information on income at the personal and the household level, e.g. income from labor, from capital, and from social security benefits, etc.

Table 6 shows the distribution of household wealth and household income in our sample (column 2) compared to the population (column 3). The median household wealth in our sample is 145,649 euros, higher than the population median of 90,613 euros, while the median household income in our sample is lower than that in the population. This is consistent with the higher proportion of older respondents in our sample.

Socioeconomic variables, such as gender, age, education, occupation, and household composition are obtained from the LISS Panel and used as controls in our analyses. We classify education levels according to Statistics Netherlands: primary or pre-vocational education as low education, pre-university or vocational education as medium education, and post-vocational or university education as high education. The primary occupation reported by each subject is categorized as paid work, housekeeping, retired, or others, following Choi et al. (2014).

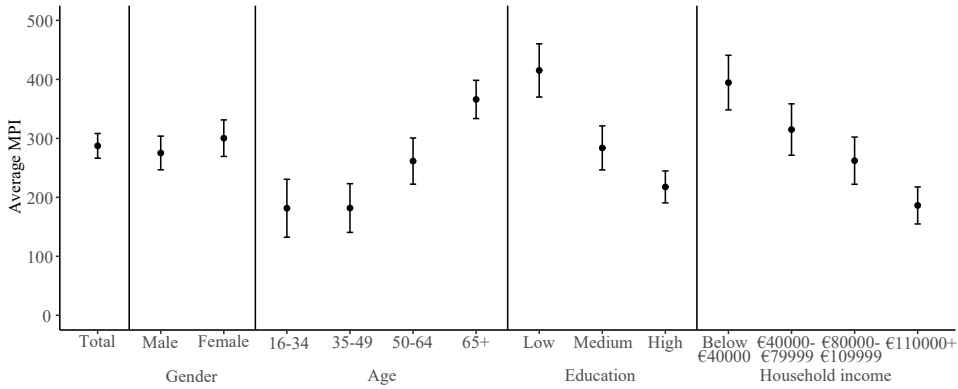


Figure 19: Mean MPI across the two lotteries in each demographic group. Error bars present 95 percent confidence intervals. The four categories of gross annual household income are relatively equally distributed: below 40000 euros, 40000-79999 euros, 80000-109999 euros, and 110000 euros and above. Figure C.1 in Appendix C.1 shows this for each lottery separately.

4.3 Results

To study the association between choice inconsistency and household wealth, we conduct our analyses in five steps. First, we show that MPI correlates in the expected direction with socioeconomic characteristics that are commonly associated with decision-making ability. Second, we employ a series of regression analyses to examine the association between household wealth and MPI while controlling for these socioeconomic characteristics. Third, we account for potential confounding factors such as cognitive ability and risk attitudes. Fourth, since most previous research uses self-reported household wealth, we compare our main results to those in the literature using self-reported household wealth from the LISS panel to further assess the robustness of our results. Finally, we explore the potential channels underlying the association between MPI and household wealth.

4.3.1 Relating MPI to demographic characteristics

We use the average MPI across the two lotteries for the main analyses. We provide separate analyses using the MPI from each lottery in Appendix C.1. MPI for each lotteries exhibits similar patterns across demographic characteristics, although the differences across demographic groups are slightly more pronounced when using the MPI from L2.

Figure 19 summarizes these results by displaying the mean and the 95 percent confidence intervals of the MPI across different socioeconomic groups. We find a significant positive correlation between MPI and age (Spearman's $\rho = 0.324$, $p < 0.01$) and significant negative correlations between MPI and education as well as household income (Spearman's $\rho = -0.275$ for education and -0.294 for household income; $p < 0.01$ for both). Column 1 of Table 7 further supports these results, showing statistically significant regression coefficients for age, education, and household income. Thus, younger, higher-educated,

		Average MPI	
Cognitive ability		−51.05***	(10.94)
Female		20.23	25.83
		(20.64)	(20.52)
Age	35 - 49	18.97	17.43
		(33.97)	(33.30)
	50 - 64	84.70**	72.14**
		(33.62)	(32.93)
65+	164.47***	149.61***	
	(41.95)	(45.68)	
Education	Medium	−69.74**	−52.51*
		(30.98)	(30.22)
	High	−124.98***	−96.66***
		(30.29)	(30.36)
Gross household income	40000 - 79999	−43.55	−40.79
		(30.43)	(31.10)
	80000 - 109999	−39.88	−34.71
		(32.95)	(32.35)
110000+	−98.22**	−87.49***	
	(33.16)	(32.23)	
Constant		318.71***	309.76***
		(41.72)	(41.72)
Controls		Yes	Yes
Adjusted R ²		0.153	0.187
Observations		571	571

Table 7: Results of the OLS regressions with the average MPI across lotteries as the dependent variable and socioeconomic variables (column 1) and additionally (standardized) cognitive ability (column 2) as explanatory variables. Other controls such as occupation (housekeeping, retired, others) and household composition (partner and children living at home) are in line with Choi et al. (2014). See Table C.1 in Appendix C.1 for the coefficients for the controls. The benchmark (omitted) category is male, low education, and household income below €40000. The significance *, **, and *** denotes $p < 0.10$, $p < 0.05$, and $p < 0.01$, respectively.

and higher-income subjects displayed greater levels of consistency in their choices than older, lower-educated, and lower-income subjects. Our findings that MPI is correlated with age, education, and household income align with the results from earlier studies that use decision-making quality measures such as CCEI in Choi et al. (2014) as well as the original MPI in Echenique et al. (2011) and Banks et al. (2019).

Decision-making ability is also commonly associated with cognitive ability. In our experiment, we measured cognitive ability with six matrix reasoning questions. Consistent with the literature (Burks et al., 2009), we find a significant and negative correlation between the number of correctly answered matrix reasoning questions and MPI (Spearman's $\rho = -0.294$, $p < 0.01$). In column 2 of Table 7, we add cognitive ability (rescaled to have the mean of 0 and standard deviation of 1) as an additional explanatory variable to the OLS model of column 1. The result shows a negative association between cognitive ability

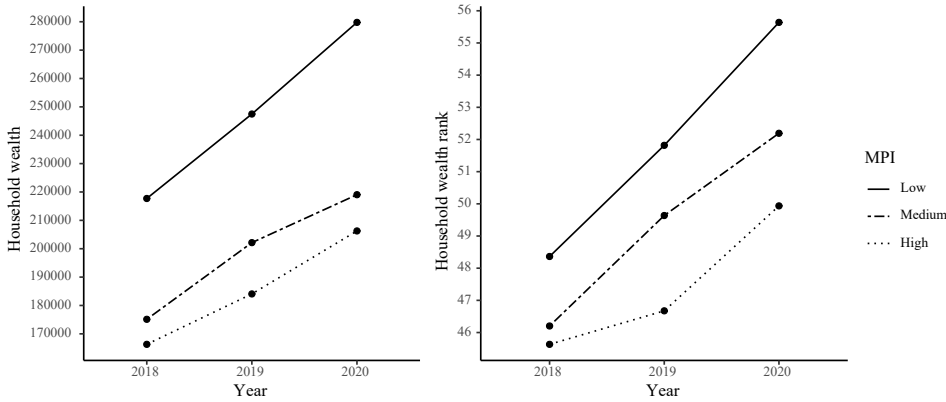


Figure 20: Average household wealth (left panel) and the rank percentile of the household wealth within the sample (right panel) for three levels of MPI in 2018, 2019, and 2020.

and MPI, with an additional standard deviation increase in the performance of the matrix reasoning questions corresponding to a decrease of about 50 points in MPI.

Overall, these findings suggest that MPI is a reliable measure of decision-making quality. It correlates with socioeconomic background variables previously linked to decision-making quality and aligns systematically with other decision-making quality proxies in the expected direction. Further, the substantial heterogeneity in MPI both within and across socioeconomic groups provides a strong basis to test the relationship between MPI and household wealth, to which we now turn.

4.3.2 The relationship between MPI and household wealth

Figure 20 provides a first illustration of the association between MPI and household wealth. We divided subjects into low, medium, and high MPI categories, using the 33rd percentile and 67th percentile as cutoffs. In the left panel, we show the average household wealth in each category in the year of the experiment (2020) as well as in 2018 and 2019.¹⁸ Household wealth is highest in the low MPI group, and lowest in the high MPI group across all three years. To ensure that the positive association between MPI and household wealth is not driven by outliers with extreme levels of household wealth, following Epper et al. (2020) we chart the average household wealth rank within our sample by three categories of MPI in the right panel. The findings are qualitatively similar.

Next, we examine the correlation between MPI and household wealth while accounting for socioeconomic differences among the subjects using OLS regressions. The dependent variable is the log of household wealth. We treat socioeconomic controls such as income, education, and family structure as predetermined, following the tradition in life-cycle

¹⁸We restrict the sample to subjects who retained the same household identifier across the years, while acknowledging that household composition may vary as members joined or left the household over time.

analysis (Browning and Crossley, 2001; Bernheim et al., 2001). We use contemporaneous household income as a proxy for the lifetime income trajectory (Choi et al., 2014). Although we treat socioeconomic controls as predetermined, we remain cautious about interpreting any correlation between MPI scores and household wealth as causal. The goal of this exercise is primarily to rule out the possibility that the observed correlation between MPI and household wealth merely reflects the socioeconomic characteristics of the subjects. As 9.7% of the subjects had negative household wealth, they were omitted from the regression sample.

Column 1 of Table 8 reports our baseline estimates. We find that higher MPI is significantly associated with lower household wealth. The coefficient of -0.157 for MPI indicates that one standard deviation increase in MPI is associated with a 14.5% decrease in household wealth. The complete regression results in Table C.2 in Appendix C.1 show that the associations between household wealth and socioeconomic variables are in the expected directions. In column 2, we further examine the relationship between MPI as a measure of decision-making ability and household wealth by focusing on subjects who are identified as the head of the household, as they are the primary decision-makers shaping the household's financial situation. The impact of MPI increases, with one standard deviation increase of MPI for household heads associating with a 22.2% decrease in household wealth. This suggests that the decision-making ability of the individual who is in charge of the financial decision-making in the household (Smith et al., 2010) is particularly important for the accumulation of the household wealth.

The remaining columns of Table 8 present the regression results for different subsamples. Columns 3 and 4 test the robustness of our results by excluding subjects who may drive our results. In column 3, we exclude subjects in the top and bottom 1 percent of the wealth distribution to ensure that our results are not driven by extreme outliers. The coefficient on the MPI is slightly smaller but remains statistically significant: a one standard deviation increase in MPI is now associated with a 13.5% decrease in household wealth. In column 4, we exclude subjects who made dominated choices as they could be inattentive and their MPI may not reflect the true underlying risk of exploitation. This exclusion leads to a larger coefficient on the MPI (-0.200), indicating that one standard deviation increase in MPI is associated with an 18.1% decrease in household wealth.

Columns 5 and 6 examine the sensitivity of our results to different restricted subsamples. In column 5, we exclude subjects younger than 35. The life cycle hypothesis posits that younger households have less incentive to hold wealth and may borrow to invest in their future. This implies that the correlation between decision-making ability and wealth may be stronger after excluding younger households. We find that the coefficient on the MPI after excluding younger subjects is only slightly larger than the baseline coefficient, suggesting that the correlation between decision-making ability and wealth may not be much weaker among younger households. Column 6 reports the results for the subsample that excludes retired subjects. Retired households tend to have more wealth as households tend to accumulate wealth until they retire (Modigliani and Brumberg, 1954; Modigliani, 1986),

	Log of household wealth					
	Total sample	Household head	Excluding:			
			Wealth outliers	Dominated choosers	Younger than 35	Retirees
	(1)	(2)	(3)	(4)	(5)	(6)
MPI	-0.157** (0.073)	-0.251*** (0.083)	-0.146** (0.073)	-0.200** (0.089)	-0.160** (0.075)	-0.182* (0.104)
Controls	Yes	Yes	Yes	Yes	Yes	
Adjusted R^2	0.385	0.399	0.380	0.374	0.344	0.456
Observations	511	374	505	443	468	305

Table 8: Results of the OLS regressions of the log of household wealth on the average (standardized) MPI across the two lotteries. Controls include gender (binary), a flexible age function, education, the log of household income (excluding income from capital), occupation, and household composition (partner and number of children living at home). Education and occupation categories are defined in Section 4.2.3. Complete regression results can be found in Table C.2 in Appendix C.1. The results for each lottery separately can be found in Table C.3 in Appendix C.1. The significance *, **, and *** denotes $p < 0.10$, $p < 0.05$, and $p < 0.01$, respectively.

and they also tend to have higher MPI due to their age. Hence, the inclusion of the retired households in our analysis may weaken the observed correlation between MPI and wealth. We find that the coefficient on the MPI is 16% larger than our baseline coefficient, although the coefficient is estimated somewhat less precisely (-0.182, significant only at the 10% level), possibly due to the substantial decline in the sample size.

4.3.3 Controlling for cognitive ability and risk attitudes

In this subsection, we examine the extent to which the observed association between MPI and household wealth is driven by unobserved characteristics and preferences, namely cognitive ability and risk attitudes. Table 9 reports the results. Past studies have found that cognitive ability significantly impacts wealth accumulation (Smith et al., 2010; Choi et al., 2014) and predicts behavioral inconsistency (Willadsen et al., 2024). We find that subjects with higher cognitive ability have lower MPI. This suggests the observed association between MPI and household wealth may be partially driven by cognitive ability. Column 2 of Table 9 adds subjects' performance in the matrix reasoning task to our baseline OLS specification. We find a significant and positive relationship between cognitive ability and household wealth, with one standard deviation increase in the performance of the matrix reasoning questions corresponding to a 13.7% increase in household wealth. While adding cognitive ability to the baseline specification slightly reduces the statistical significance of the MPI coefficient (from 5% to 10%), the size of the coefficient remains broadly similar (from -0.157 to -0.132).

Like cognitive ability, risk attitudes are commonly associated with household wealth. Standard financial theory posits that investors with a greater willingness to take financial risk accumulate wealth more rapidly due to the substantial premium of equity over bonds,

	Log of household wealth			
	(1)	(2)	(3)	(4)
MPI	-0.157** (0.073)	-0.132* (0.074)	-0.198** (0.078)	-0.174** (0.079)
Cognitive ability		0.128** (0.063)		0.132** (0.063)
Risk attitude			0.130** (0.062)	0.134** (0.062)
Controls	Yes	Yes	Yes	Yes
Adjusted R^2	0.385	0.388	0.389	0.393
Observations	511	511	511	511

Table 9: Results of the OLS regressions of the log of household wealth on the average MPI across the two lotteries. Columns 2 and 3 include in addition to the baseline specification cognitive ability and risk attitude, respectively, and column 4 includes both. MPI, cognitive ability, and risk attitude are all standardized. Controls are the same as in Table 8. Complete regression results can be found in Table C.4 in Appendix C.1. The results for each lottery separately can be found in Table C.5 in Appendix C.1. The significance *, **, and *** denotes $p < 0.10$, $p < 0.05$, and $p < 0.01$, respectively.

although the empirical evidence for this association is mixed (Bernheim et al., 2001). In addition, risk attitudes may affect wealth accumulation by influencing households' propensity to plan (Ameriks et al., 2003). There is also evidence that risk aversion may be related to choice inconsistency, with those who made inconsistent choices more likely to be risk-averse (Amador-Hidalgo et al., 2021; Stillman et al., 2020), and risk attitudes together with inconsistent choices can explain the use of financial instruments, such as participation in a savings group or taking up an informal loan (Jacobson and Petrie, 2009). Therefore, we examine the possibility that the correlation between MPI and wealth is driven by the relationship between risk attitudes and wealth.

We compute the average certainty equivalent for the lotteries for each individual and include it as a measure of individual risk attitude in Column 3 of Table 9. We find a significant and positive association between the average certainty equivalent for the lotteries and household wealth: one standard deviation increase in the certainty equivalent is associated with a 14% increase in household wealth. The inclusion of risk attitude as a control did not reduce the correlation between MPI and household wealth; instead, it increased the MPI coefficient from -0.157 to -0.198 . This suggests that risk attitude is not the omitted variable that is driving the relationship between MPI and wealth in our baseline specification. In column 4 of Table 9, we further included both cognitive ability and risk attitudes as controls. The correlation between MPI and wealth remains negative (-0.174 vs -0.157 in the baseline specification) and significant at 5% level.

4.3.4 Self-reported data vs administrative records

Previous studies linking choice inconsistency to real-life outcomes have relied on self-reported data (Smith et al., 2010; Choi et al., 2014). In our case, we have both wealth

	Log Household wealth				
	Self-reported data:			Administrative data	
	Definition 1	Definition 2	Definition 3	Total	Definition 3 sample
MPI	-0.321** (0.142)	-0.469*** (0.176)	-0.553*** (0.198)	-0.157*** (0.073)	-0.265** (0.103)
Controls	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.274	0.280	0.265	0.385	0.417
Observations	462	311	228	511	239

Table 10: Results of the OLS regressions of the log of household wealth on the average MPI across lotteries. MPI is standardized. Controls are the same as in Table 8. We run the regression for each of the three definitions of self-reported household wealth and for the restricted sample of administrative household wealth. The significance *, **, and *** denotes $p < 0.10$, $p < 0.05$, and $p < 0.01$, respectively.

information from administrative records and self-reported household wealth from the LISS Panel. To explore how the source of wealth data affects the estimated correlation between MPI and household wealth, we repeat our analyses using the self-reported wealth data.

Due to the prevalence of missing data in the self-reports, we created three definitions of household wealth that trade off between the reliability of the data and the number of observations. In Table 10, we show the results for each definition of household wealth. Column 1 shows the least restrictive definition, where we only exclude the top and bottom 1% of household wealth. In column 2, we additionally exclude subjects for whom one or more members of their household did not participate in the LISS core study on wealth. Finally, column 3 excludes all subjects for whom any household wealth data is missing, i.e. household members indicated that they had a certain type of assets but did not report the value of that asset. In all three definitions of household wealth, we find a negative and statistically significant correlation between MPI and self-reported household wealth. Thus, the relationship between MPI and household wealth also exists in self-reported data.

Interestingly, the coefficient of MPI estimated using self-reported household wealth is about twice the size of the coefficient estimated using administrative records. For example, the MPI coefficient in the sample of definition 3 is -0.553, while it is -0.265 when we use the administrative records from the same sample. This suggests that our main results using the administrative records are perhaps on the conservative side compared to previous studies. While it is difficult to pinpoint exactly where and how the differences between the administrative and self-reported data arise, our analyses suggest that subjects on average under-report their household wealth, and the underreporting bias interacts with MPI. The binned scatter plot in Figure 21 demonstrates this association by creating bins (or groups) for the average MPI across lotteries and calculating the mean log household wealth within each bin (see Starr and Goldfarb, 2020, for an overview of the method). As we can see, subjects with a higher MPI are more likely to under report their wealth, probably because they have a worse overview of their financial situation. The difference in the under-report

bias increases the association between MPI and the self-reported household wealth. See Appendix C.3 for more details.

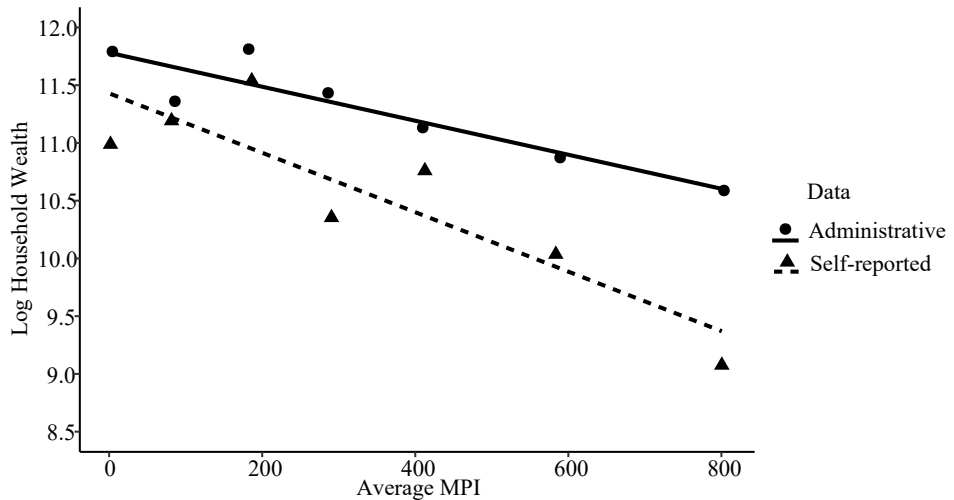


Figure 21: Binned scatter plot of the association between MPI and household wealth and corresponding best-fit regression line for self-reported (triangles, dashed) and administrative (dots, solid) measures of household wealth. The number of bins is selected by minimizing the integrated means squared error (Cattaneo et al., 2023).

4.3.5 Why does MPI relate to household wealth?

Our results so far robustly suggest a statistically significant and economically important association between MPI and household wealth. But why does choice inconsistency as measured by MPI relate to household wealth? Chew et al. (2022) suggested that choice inconsistency in settings like ours could arise from random preferences (Luce and Suppes, 1965; Eliashberg and Hauser, 1985; Gul and Pesendorfer, 2006; Apesteguia and Ballester, 2018), preference incompleteness/imprecision (Aumann, 1962; Cettolin and Riedl, 2019; Butler and Loomes, 2007), deliberate randomization to hedge uncertainty preferences (Fudenberg et al., 2015), or cognitive uncertainty (Enke and Graeber, 2023; Vieider, 2023).

One could argue that each of these factors is indicative of poor decision-making ability. Random preferences can be associated with instability of preferences, potentially leading to reversals of choices (Hey and Orme, 1994; Harless and Camerer, 1994). Incomplete, imprecise, or uncertain preferences are often attributed to an inability to differentiate between options, which could lead to default bias (Masatlioglu and Ok, 2005), procrastination in making decisions (Danan and Ziegelmeyer, 2006; Gerasimou, 2017; Costa-Gomes et al., 2022), and excessive randomization (Agranov and Ortoleva, 2017, 2023; Halevy et al., 2023). Enke and Graeber (2023) defined cognitive uncertainty as a cognitive limitation in making utility-maximizing decisions, which can induce subjects to use less costly but error-prone

valuation procedures (Oprea, 2024; Arrieta and Nielsen, 2024) or opt for satisficing instead of maximizing (Salant and Spenkuch, 2022). Subjects with higher MPI are more likely to be vulnerable to these decision-making weaknesses, which in turn can impact their household wealth.

A thorough examination of this hypothesis requires a different set of experiments and more comprehensive data, and it is thus beyond our scope. To provide some insights into this hypothesis, we focus on one particular prediction. Cerreia-Vioglio et al. (2024) suggested that individuals may behave cautiously when they do not know what to do (e.g., when preferences are uncertain, incomplete, or imprecise). In financial decision-making, one implication is that they will prefer to invest in safe assets that often deliver low returns. Consistent with this prediction, our analyses in Appendix C.4 suggest that subjects with higher MPI were more likely to hold their wealth in low-risk, low-return bank savings, while those with lower MPI were more likely to hold financially advantageous assets such as housing assets (Van Ewijk et al., 2007) and investment assets. Further, looking at income from labor and capital separately, we find that MPI relates significantly and negatively to income from capital, with one standard deviation increase in MPI corresponding to a decrease of 26.5% in household income from capital, whereas there is no significant association between MPI and income from labor. This is consistent with Bastani et al. (2023), who found that cognitive ability is better at predicting capital income than labor earnings. Appendix C.4 contains more details about each variable and the setup of the these regressions.

4.4 Conclusion

We have investigated the relationship between our experimental measure of choice inconsistency (MPI) and real-life household wealth obtained from the administrative records of Statistics Netherlands in a large and heterogeneous sample of the Dutch population. We find a robust association between MPI and household wealth, controlling for socioeconomic variables and other measures of decision-making ability. Comparing the results from administrative records and self-reported data, we find that the use of self-reported data overestimates the association between household wealth and choice inconsistency due to the interaction of biases in self-reported household wealth and MPI.

In exploring the relationship between MPI and household wealth, two critical issues require further investigation in future research. First, it is useful to clarify what MPI actually measures and how it differs from other measures such as cognitive abilities or numeracy. Previous studies suggest that cognitive abilities are related to choice consistency (Frederick, 2005; Burks et al., 2009; Benjamin et al., 2013; Gill and Prowse, 2016), stochastic choices (Andersson et al., 2016), and the formation of rational inflation expectations (D'Acunto et al., 2023). Our findings support this, showing a strong and significant correlation between MPI and cognitive abilities. However, we also find that MPI continues to be a significant predictor of household wealth after controlling for cognitive abilities and other socioeconomic factors.

This suggests that economic rationality and cognitive abilities do not measure the same kind of decision-making ability (Kim et al., 2018). To identify the similarities and differences between the two, it would be informative to decompose choice inconsistency into types based on the underlying reasons for violations (e.g., violations of the independence axiom or the completeness of preferences), and distinguish between inconsistency that closely linked to cognitive abilities and those that are independent of them. Such an investigation would give insights into the nature of choice inconsistency. An excellent example is Stango and Zinman (2022), who reduced 20 biases and standard preference measures to four behavioral common factors and showed that three of them are related to cognitive ability and outcomes in household finance, well-being, and health. Similarly, Carvalho and Silverman (2024) linked choice inconsistency, as measured by CCEI, to sophistication - the ability of knowing when to choose simple alternatives rather than solving complex problems.

A second crucial issue to explore is the underlying mechanisms that link MPI to household wealth. Due to limitations in our data, we were unable to fully address this question. Gaining a deeper understanding of these mechanisms would be highly valuable. Specifically, it would be useful to gather data on household economic activities that make them vulnerable to exploitation or financial inefficiencies (informal debts or credit card debt, as in Jacobson and Petrie, 2009; Meier and Sprenger, 2010) as well as those that are beneficial to wealth accumulation (e.g., the propensity to plan, as in Ameriks et al., 2003; van Rooij et al., 2012). We can then relate them to MPI. Such investigation would give direct insights into whether households with high MPI are more prone to engaging in economically detrimental activities but less likely to implement activities that promote wealth accumulation, clarifying the mechanisms underlying household wealth and MPI. Some progress has been made (Ameriks et al., 2003; Huck and Müller, 2012; Herrmann et al., 2017; Castillo et al., 2018; D'Acunto et al., 2023; Carvalho et al., 2024), and further research remains a promising direction.

Related to this issue, it would be useful to pinpoint the types of choice inconsistency that have the most significant economic impacts. There are interesting theoretical investigations that relate choice inconsistency to specific types of weakness in decision-making. For example, de Clippel and Rozen (2023) show that the violation of their rationality measure (referred to as First-order Conditions Departure Index in their paper) can be interpreted as mis-perception of prices, probabilities, or discount factor (Laibson, 1997; Barber and Odean, 2001; Dimmock et al., 2016, see also DellaVigna, 2009 for an review). Future studies can compare the association of different types of choice inconsistency with economic activities and identify the most significant channels. This knowledge could inform policy interventions that aim to reduce the negative financial consequences associated with choice inconsistency and improve overall household financial well-being.

Conclusion and discussion

5.1 General conclusion

This dissertation highlights the important role of complexity when it comes to understanding individual decision-making. In three chapters, each investigating a different facet of complexity in decision-making, we show that (1) randomization probabilities are an appropriate incentivized measure of decision confidence, which is a proxy for the decision complexity that individuals perceive, (2) that two forms of complexity contributes substantially to the WTA-WTP gap, and (3) that inconsistent choices, resulting from challenges in dealing with decision complexity, can have negative real-life economic consequences.

Chapter two introduces and validates an incentivized behavioral measure of decision confidence, drawing on individuals' tendency to randomize between choice options. We motivate the measure based on models of preference uncertainty. Our experimental results suggest that there is a strong and positive correlation between randomization probabilities and two alternative self-reported measures of decision confidence (with median Spearman correlations ranging from 0.86 to 0.89), and that the randomization probabilities align closely with self-reported probabilistic confidence in absolute levels. Furthermore, exogenous manipulations of decision confidence lead to corresponding shifts in randomization behavior and self-reported decision confidence. Specifically, increasing the complexity of a lottery reduced self-reported confidence and prompted individuals to randomize over a wider range of sure payments, with probabilities closer to 0.5. Conversely, providing subjects with more experience with a lottery enhanced their self-reported confidence, leading to a narrower range of randomization and probabilities further from 0.5. Together, we demonstrate that randomization probabilities are a robust and quantitative proxy for decision confidence.

Chapter three delves into the significant role of complexity in explaining the willingness-to-accept (WTA) and willingness-to-pay (WTP) gap. The research quantifies two distinct forms of complexity: "mechanism complexity" related to the elicitation method, and "valuation complexity" pertaining to individuals' uncertainty in valuing choice objects. Using a representative Dutch sample, the study reveals that both forms of complexity systematically influence the WTA-WTP gap. Accounting for mechanism complexity alone reduced the observed WTA-WTP gap for monetary lotteries by approximately 40%, while valuation complexity accounted for about 15%. When considered jointly, these two forms of complexity explained roughly 50% of the WTA-WTP gap. Notably, the impact of complexity on the WTA-WTP gap was particularly pronounced among participants with lower cognitive abilities, suggesting that complexity-driven heuristics play a crucial role, especially for this demographic. These findings underscore that the WTA-WTP gap is not a phenomenon with a single, simple explanation, challenging the singular focus on loss aversion. Instead, it high-

lights that complexity could be a critical determinant of seemingly “irrational” economic behaviors.

Chapter four bridges the gap between experimentally observed choice inconsistency and its real-life economic consequences, particularly for household wealth. We obtained an experimental measure of the Money Pump Index (MPI), which quantifies the potential monetary losses resulting from inconsistent choices in a large, representative sample. A key contribution is the linking of this experimental MPI data directly to administrative household wealth records from Statistics Netherlands (CBS). We find a strong and statistically significant inverse relationship: a one standard deviation increase in MPI was associated with a 14.8% reduction in household wealth. This relationship remained robust even after controlling for various socioeconomic factors, cognitive ability, and risk attitudes. Further analysis revealed that individuals with higher MPI tended to hold more of their wealth in low-risk, low-return bank savings and exhibited significantly lower income from capital. Overall, this chapter demonstrates the external validity of the experimentally measured decision-making quality, showing that choice inconsistencies observed in an experimental setting are not merely artifacts but are meaningfully linked to real-world financial outcomes and wealth accumulation.

5.2 Implications

5.2.1 Methodological implications

There are several methodological implications that can be derived from the results of this dissertation. First, regarding the experimental measures of decision confidence and complexity, this work makes a significant conceptual and methodological contributions. Chapter two, specifically, introduces and validates an incentivized behavioral measure of decision confidence based on randomization probabilities. This measure moves beyond conventional non-incentivized self-reports, providing a robust and quantitative alternative that is simple to elicit. Building on this, Chapter three elicits measures for the two distinct yet interconnected forms of decision complexity, mechanism complexity and valuation complexity. The study develops specific methods for quantifying these two forms of complexity and applies them in a large representative sample: mechanism complexity is assessed through deviations from choices involving objects with known preferences, while valuation complexity is quantified by the observed range of prices where individuals choose randomization. This allows for a more precise analysis of how different forms of complexity influence economic decisions. Finally, in Chapter four we build on existing work by using a simple, easily implementable measure of choice inconsistency that allows us to not only categorize participants into groups, but to quantify the level of inconsistency in their decision-making.

Second, this dissertation contributes to the external validity of experimental findings in behavioral economics. Both Chapter three and Chapter four underscore the importance

of utilizing representative samples. The use of such samples is crucial for ensuring that experimental observations are generalizable and relevant for policy design, moving beyond findings from potentially unrepresentative student populations. A particularly strong methodological contribution in Chapter four is the integration of experimental data with administrative records from Statistics Netherlands (CBS). This approach minimizes reliance on potentially biased or noisy self-reported data, leading to more reliable conclusions.

Finally, the results of this dissertation illustrate the importance of accounting for decision complexity when designing economic experiments. Failing to do so can introduce biases and noise, thereby distorting the inference of underlying preferences. Aside from clear instructions and training aimed to limit mechanism/valuation complexity, the experimental design should include measures to quantify and control for the complexity perceived by participants. This is because participants have limitations in cognitive capacity and attention. Despite careful explanation and extensive training, some participants may nevertheless perceive substantial complexity.

5.2.2 Theoretical implications

Theoretically, this dissertation provides links between different important strands of literature. In Chapter two we link decision confidence, cognitive uncertainty, preference uncertainty/imprecision, incomplete preferences, preference for randomization, and stochastic choice. The systematic relationship between randomization probabilities and the other concepts suggests that there may exist a common psychological factor underlying them. Similarly, in Chapter three we combine the literature on the two sources of complexity, mechanisms complexity and valuation complexity, to separately and jointly show their contribution to the WTA-WTP gap.

A further implication of our results on the role of decision complexity in the WTA-WTP gap is that it challenges the prevailing theoretical explanation of this phenomenon and provides evidence for a more nuanced perspective where the WTA-WTP gap is seen as a multifaceted phenomenon. While in Chapter three we focus on the WTA-WTP gap, the possibility that this also holds for other observed behavioral anomalies is not farfetched. Previous research has suggested that this may be the case for preference reversals (Butler and Loomes, 2007), stochastic choices (Agranov and Ortoleva, 2017), insensitivity to variation in probabilities (Enke and Shubatt, 2023), anomalies in intertemporal choices (Enke et al., 2025), small-stakes risk aversion (Khaw et al., 2021), and more (Butler and Loomes, 2011). This suggests that these phenomena are similarly affected by individuals' difficulty to process and respond to the inherent complexity of decision tasks and objects.

A final implication concerns the validity of theoretical findings more broadly. The evidence from Chapter four shows that experimentally measured inconsistency is systematically related to real-life economic outcomes. This link suggests that experimental constructs, such as decision complexity, are not confined to the laboratory, but capture processes with tangible consequences outside of it. Theoretically, this strengthens the

case for incorporating such constructs into models of decision-making: if experimental measures predict real outcomes, then the theories built around them are likely externally relevant and normatively significant.

5.2.3 Policy implications

The findings on the role of complexity in decision-making have direct implications for policymaking and consumer choices. Many policies rely on eliciting or observing public preferences. However, if mechanism complexity and valuation complexity systematically distort these preferences, it becomes difficult for policymakers to obtain an unbiased picture of what individuals truly want. Similarly, outcomes from market research or other consumer testing may be shaped not only by preferences, but also by the complexity of the elicitation method and the difficulty of valuating options. This highlights the need for not only incentive-compatible elicitation methods but also for approaches that explicitly account for decision complexity. This is particularly important in domains where inherently complex decision contexts coincide with high stakes, such as policy referendums, insurance and pension choices, investment products, or recruitment procedures. Failing to recognize the role of complexity risks designing overly complex policies that people cannot understand, or making policy recommendations that are based on distorted preferences and therefore lack both effectiveness and legitimacy.

A second implication follows from the link between choice inconsistency and household wealth observed in Chapter four and the differential impact of complexity on decision making among the different subgroups in Chapter three. Decision complexity does not affect all individuals equally: those with limited cognitive or financial resources are more vulnerable to avoidable losses, dominated offers, and exploitation. As a result, the inability to deal with complexity can reinforce or exacerbate existing inequalities in wealth and financial security. Policymakers should therefore target complexity reduction and error-proofing in critical decision domains. Examples include simplifying and standardizing language use, restricting dominated or overly complex offers, or providing safe default options that individuals can fall back on when confidence is low. Beyond these interventions, behavioral measures of confidence and inconsistency could serve as diagnostic tools to detect when markets or policy instruments are perceived as complex by individuals. They can signal to policymakers when intervention might be necessary.

5.3 Limitations and future research

Like any empirical work, the findings presented in this dissertation should be interpreted in light of a number of limitations. A first limitation concerns the measurement of complexity. While we propose and apply measures that capture important dimensions of decision complexity, it remains difficult to obtain exact estimates of the complexity experienced by individuals. Experimental designs rely on elicitation mechanisms, and as shown in Chapter three, these mechanisms themselves can be the subject of complexity and this can in turn

interact with the valuation process, creating an additional layer of complexity. This raises the possibility that observed outcomes reflect not only the complexity of the task or object but also interactions between different forms of complexity. Future research may further develop methods to disentangle these effects, for instance by combining behavioral data with physiological or neurological measures of cognitive load.

A second limitation relates to the underlying mechanisms through which complexity affects decision making and outcomes. While the results clearly demonstrate that complexity matters, the precise channels remain less well understood. For example, do individuals in complex environments rely more on heuristics, experience greater cognitive fatigue, or simply disengage from the task? Understanding these pathways is essential, not only for refining theoretical models of decision making but also for designing interventions that can effectively mitigate the adverse effects of complexity. Future studies could explore these mechanisms by integrating experimental tasks with more detailed real-life (administrative) data, eye-tracking, process-tracing methods, or other qualitative data.

A third direction for future research lies in extending the analysis to more behavioral anomalies. The WTA–WTP gap provides a clear example of how complexity can shape well-documented deviations from standard theory, but there are good reasons to expect similar dynamics in other domains. As discussed above, previous research has already theorized and/or investigated the role of concepts related to decision complexity for a broad range of behavioral anomalies. Future research can elaborate on this by testing the extent to which these anomalies persist once complexity is measured and accounted for.

Finally, a more general limitation concerns the scope and external validity of experimental methods. While this dissertation aims to account for this as much as possible by employing representative samples and linking experimental indicators to administrative outcomes, the laboratory and survey context still differ from the environments in which many high-stakes decisions are made. Replication across settings, populations, and decision domains will be essential to test the robustness of the conclusions and to better understand how complexity operates outside controlled experiments. In particular, field experiments that combines experimental controls with natural settings and high stakes would be highly valuable. Such field work would not only strengthen the empirical foundation for policy recommendations but also provide deeper insight into the generalizability of the theoretical links between the different concepts researched in this dissertation.

Bibliography

- Abdellaoui, M., Li, C., Wakker, P., and Wu, G. (2020). A defense of prospect theory in Bernheim and Sprenger's experiment. Working paper, Erasmus University Rotterdam.
- Afriat, S. N. (1972). Efficiency estimation of production functions. *International economic review*, 13(3):568–598.
- Afriat, S. N. (1973). On a system of inequalities in demand analysis: An extension of the classical method. *International Economic Review*, 14(2):460–72.
- Agranov, M. and Ortoleva, P. (2017). Stochastic choice and preferences for randomization. *Journal of Political Economy*, 125(1):40–68.
- Agranov, M. and Ortoleva, P. (2023). Ranges of preferences and randomization. *Review of Economics and Statistics*, Forthcoming.
- Aguiar, V. H. and Kashaev, N. (2020). Stochastic Revealed Preferences with Measurement Error. *Review of Economic Studies*, 88(4):2042–2093.
- Amador-Hidalgo, L., Branas-Garza, P., Espín, A. M., García-Muñoz, T., and Hernández-Román, A. (2021). Cognitive abilities and risk-taking: Errors, not preferences. *European Economic Review*, 134:103694.
- Ameriks, J., Caplin, A., and Leahy, J. (2003). Wealth accumulation and the propensity to plan. *Quarterly Journal of Economics*, 118(3):1007–1047.
- Anagol, S., Balasubramaniam, V., and Ramadorai, T. (2018). Endowment Effects in the Field: Evidence from India's IPO Lotteries. *Review of Economic Studies*, 85(4):1971–2004.
- Andersson, O., Holm, H. J., Tyrann, J.-R., and Wengström, E. (2016). Risk aversion relates to cognitive ability: Preferences or noise? *Journal of the European Economic Association*, 14(5):1129–1154.
- Apesteagua, J. and Ballester, M. A. (2015). A measure of rationality and welfare. *Journal of Political Economy*, 123(6):1278–1310.
- Apesteagua, J. and Ballester, M. A. (2018). Monotone stochastic choice models: The case of risk and time preferences. *Journal of Political Economy*, 126(1):74–106.
- Apicella, C. L., Azevedo, E. M., Christakis, N. A., and Fowler, J. H. (2014). Evolutionary origins of the endowment effect: Evidence from hunter-gatherers. *American Economic Review*, 104(6):1793–1805.
- Ariely, D., Loewenstein, G., and Prelec, D. (2003). "Coherent Arbitrariness": Stable Demand Curves Without Stable Preferences*. *Quarterly Journal of Economics*, 118(1):73–106.
- Arrieta, G. and Nielsen, K. (2024). Procedural decision-making in the face of complexity. Working paper, California Institute of Technology.
- Arts, S., Ong, Q., and Qiu, J. (2020). Measuring subjective decision confidence. MPRA Paper 106811, University Library of Munich, Germany.
- Arts, S., Ong, Q., and Qiu, J. (2024a). Measuring decision confidence. *Experimental Economics*, 27(3):582–603.
- Arts, S., Qiu, J., and Zhong, S. (2024b). The complexity of revealing preferences in a representative sample. Working paper, National University of Singapore.
- Aumann, R. J. (1962). Utility theory without the completeness axiom. *Econometrica*, 30(3):pp. 445–462.
- Aydogan, I. and Gao, Y. (2020). Experience and rationality under risk: Re-examining the impact of sampling experience. *Experimental Economics*, 23(4):1100–1128.
- Baillon, A., Bleichrodt, H., Huang, Z., and van Loon, R. P. (2017). Measuring ambiguity attitude: (Extended) multiplier preferences for the American and the Dutch population. *Journal of Risk and Uncertainty*, 54(3):269–281.
- Banks, J., Carvalho, L. S., and Perez-Arce, F. (2019). Education, decision making, and economic rationality. *The Review of Economics and Statistics*, 101(3):428–441.
- Banovetz, J. and Oprea, R. (2023). Complexity and Procedural Choice. *American Economic Journal: Microeconomics*, 15(2):384–413.
- Barber, B. M. and Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *Quarterly Journal of Economics*, 116(1):261–292.
- Barberá, S., Hammond, P., and Seidl, C., editors (2013). *Handbook of Utility Theory*. Springer, Berlin, Germany.
- Barberis, N. C. (2013). Thirty years of prospect theory in economics: A review and assessment. *Journal of Economic Perspectives*, 27(1):173–96.
- Bastani, S., Karlsson, K., and Waldenström, D. (2023). The different returns to cognitive ability in the labor and capital markets. Technical Report 2023:8, IFAU - Institute for Evaluation of Labour Market and Education Policy.
- Bayrak, O. K. and Hey, J. D. (2020). Decisions under risk: Dispersion and skewness. *Journal of Risk and Uncertainty*, 61(1):1–24.

- Becker, G., DeGroot, M., and Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science*, 9:226–232.
- Benartzi, S., Beshears, J., Milkman, K. L., Sunstein, C. R., Thaler, R. H., Shankar, M., Tucker-Ray, W., Congdon, W. J., and Galing, S. (2017). Should governments invest more in nudging? *Psychological Science*, 28(8):1041–1055.
- Benjamin, D. J., Brown, S. A., and Shapiro, J. M. (2013). Who is behavioral? cognitive ability and anomalous preferences. *Journal of the European Economic Association*, 11(6):1231–1255.
- Bernheim, B. D. and Rangel, A. (2009). Beyond revealed preference: Choice-theoretic foundations for behavioral welfare economics. *Quarterly Journal of Economics*, 124(1):51–104.
- Bernheim, B. D., Skinner, J., and Weinberg, S. (2001). What accounts for the variation in retirement wealth among u.s. households? *American Economic Review*, 91(4):832–857.
- Bewley, T. F. (2002). Knightian Decision Theory: Part I. *Decisions in Economics and Finance*, 25:79–110.
- Blavatskyy, P., Ortmann, A., and Panchenko, V. (2022). On the experimental robustness of the allais paradox. *American Economic Journal: Microeconomics*, 14(1):143–63.
- Blavatskyy, P. R. (2009). Preference reversals and probabilistic decisions. *Journal of Risk and Uncertainty*, 39(3):237.
- Bodnaruk, A. and Simonov, A. (2016). Loss-Averse Preferences, Performance, and Career Success of Institutional Investors. *Review of Financial Studies*, 29(11):3140–3176.
- Bosch-Domènech, A. and Silvestre, J. (2013). Measuring risk aversion with lists: a new bias. *Theory and Decision*, 75(4):465–496.
- Bouacida, E. and Martin, D. (2021). Predictive power in behavioral welfare economics. *Journal of the European Economic Association*, 19(3):1556–1591.
- Brown, J. R., Kapteyn, A., Luttmer, E. F. P., and Mitchell, O. S. (2017). Cognitive Constraints on Valuing Annuities. *Journal of the European Economic Association*, 15(2):429–462.
- Brown, J. R., Kapteyn, A., Luttmer, E. F. P., Mitchell, O. S., and Samek, A. (2021). Behavioral Impediments to Valuing Annuities: Complexity and Choice Bracketing. *Review of Economics and Statistics*, 103(3):533–546.
- Browning, M. and Crossley, T. F. (2001). The life-cycle model of consumption and saving. *Journal of Economic Perspectives*, 15(3):3–22.
- Bruner, D. M. (2011). Multiple switching behaviour in multiple price lists. *Applied Economics Letters*, 18(5):417–420.
- Burger, J. M. and Cooper, H. M. (1979). The desirability of control. *Motivation and emotion*, 3(4):381–393.
- Burks, S. V., Carpenter, J. P., Goette, L., and Rustichini, A. (2009). Cognitive skills affect economic preferences, strategic behavior, and job attachment. *Proceedings of the National Academy of Sciences*, 106(19):7745–7750.
- Butler, D., Isoni, A., Loomes, G., and Tsutsui, K. (2014). Beyond choice: Investigating the sensitivity and validity of measures of strength of preference. *Experimental Economics*, 17:537–563.
- Butler, D. and Loomes, G. (2011). Imprecision as an account of violations of independence and betweenness. *Journal of Economic Behavior & Organization*, 80(3):511–522.
- Butler, D. J. and Loomes, G. C. (2007). Imprecision as an account of the preference reversal phenomenon. *American Economic Review*, 97(1):277–297.
- Camerer, C. F. and Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty*, 19(1-3):7–42.
- Carvalho, L., Olafsson, A., and Silverman, D. (2024). Misfortune and mistake: The financial conditions and decision-making ability of high-cost loan borrowers. *Journal of Political Economy*, 132(9):000–000.
- Carvalho, L. and Silverman, D. (2024). Complexity and sophistication. *Journal of Political Economy: Microeconomics*, 2(1):46–76.
- Cason, T. and Plott, C. (2014). Misconceptions and game form recognition: Challenges to theories of revealed preference and framing. *Journal of Political Economy*, 122(6):1235 – 1270.
- Castillo, M., Petrie, R., Cotla, C. R., and Torero, M. (2018). Risk preferences and decision quality of the poor. Technical report, Working Paper.
- Cattaneo, M. D., Crump, R. K., Farrell, M. H., and Feng, Y. (2023). *Binsreg: Binscatter Estimation and Inference*. R package version 1.0.
- Cerreia-Vioglio, S., Dillenberger, D., and Ortoleva, P. (2015). Cautious expected utility and the certainty effect. *Econometrica*, 83(2):693–728.
- Cerreia-Vioglio, S., Dillenberger, D., and Ortoleva, P. (2024). Caution and Reference Effects. *Econometrica*, forthcoming.
- Cerreia-Vioglio, S., Dillenberger, D., Ortoleva, P., and Riella, G. (2019). Deliberately stochastic. *American Economic Review*, 109(7):2425–2445.
- Cerulli-Harms, A., Goette, L., and Sprenger, C. (2019). Randomizing endowments: An experimental study of rational expectations and reference-dependent preferences. *American Economic Journal: Microeconomics*, 11(1):185–207.
- Cettolin, E. and Riedl, A. (2019). Revealed preferences under uncertainty: Incomplete preferences and preferences for randomization. *Journal of Economic Theory*, 181:547 – 585.

- Chapman, J., Dean, M., Ortoleva, P., Snowberg, E., and Camerer, C. (2024). On the relation between willingness to accept and willingness to pay. Technical report.
- Cherchye, L., Demuynck, T., De Rock, B., and Vermeulen, F. (2017). Household consumption when the marriage is stable. *American Economic Review*, 107(6):1507–34.
- Chew, S. H. (1983). A generalization of the quasilinear mean with applications to the measurement of income inequality and decision theory resolving the Allais paradox. *Econometrica*, 51(4):1065–1092.
- Chew, S. H. (1989). Axiomatic utility theories with the betweenness property. *Annals of operations Research*, 19(1):273–298.
- Chew, S. H., Miao, B., Shen, Q., and Zhong, S. (2022). Multiple-switching behavior in choice-list elicitation of risk preference. *Journal of Economic Theory*, 204:105510.
- Chew, S. H., Miao, B., and Zhong, S. (2017). Partial ambiguity. *Econometrica*, 85(4):1239–1260.
- Chew, S. H., Ratchford, M., and Sagi, J. S. (2018). You need to recognise ambiguity to avoid it. *Economic Journal*, 128(614):2480–2506.
- Choi, S., Fisman, R., Gale, D., and Kariv, S. (2006). Substantive and procedural rationality in decisions under uncertainty. *ESRC Centre for Economic Learning and Social Evolution*.
- Choi, S., Kariv, S., Müller, W., and Silverman, D. (2014). Who is (more) rational? *American Economic Review*, 104(6):1518–50.
- Cohen, M., Jaffray, J.-Y., and Said, T. (1987). Experimental comparison of individual behavior under risk and under uncertainty for gains and for losses. *Organizational Behavior and Human Decision Processes*, 39(1):1–22.
- Condon, D. M. and Revelle, W. (2014). The international cognitive ability resource: Development and initial validation of a public-domain measure. *Intelligence*, 43:52–64.
- Costa-Gomes, M. A., Cueva, C., Gerasimou, G., and Tejiscak, M. (2022). Choice, deferral, and consistency. *Quantitative Economics*, 13(3):1297–1318.
- Cubitt, R., Kopsacheilis, O., and Starmer, C. (2022). An inquiry into the nature and causes of the Description - Experience gap. *Journal of Risk and Uncertainty*, 65(2):105–137.
- Cubitt, R., Navarro-Martinez, D., and Starmer, C. (2015). On preference imprecision. *Journal of Risk and Uncertainty*, 50:1–34.
- Cubitt, R., van de Kuilen, G., and Mukerji, S. (2020). Discriminating between models of ambiguity attitude: A qualitative test. *Journal of the European Economic Association*, 18(2):708–749.
- Danan, E. (2010). Randomization vs. selection: How to choose in the absence of preference? *Management Science*, 56(3):503–518.
- Danan, E. and Ziegelmeyer, A. (2006). Are preferences complete? an experimental measurement of indecisiveness under risk. Papers on strategic interaction, Max Planck Institute of Economics, Strategic Interaction Group.
- de Clippel, G. and Rozen, K. (2023). Relaxed Optimization: How Close Is a Consumer to Satisfying First-Order Conditions? *Review of Economics and Statistics*, 105(4):883–898.
- De Finetti, B. (1974). *Theory of Probability*. Wiley, New York.
- Dean, M. and Martin, D. (2016). Measuring Rationality with the Minimum Cost of Revealed Preference Violations. *Review of Economics and Statistics*, 98(3):524–534.
- Dekel, E. (1986). An axiomatic characterization of preferences under uncertainty: Weakening the independence axiom. *Journal of Economic Theory*, 40(2):304–318.
- DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic Literature*, 47(2):315–372.
- Dimmock, S., Kouwenberg, R., and Wakker, P. (2016). Ambiguity attitudes in a large representative sample. *Management Science*, 62(5):1363–1380.
- Dubourg, W. R., Jones-Lee, M. W., and Loomes, G. (1994). Imprecise preferences and the WTP-WTA disparity. *Journal of Risk and Uncertainty*, 9:115–33.
- Dubourg, W. R., Jones-Lee, M. W., and Loomes, G. (1997). Imprecise preferences and survey design in contingent valuation. *Economica*, 64(256):681–702.
- Dubra, J., Maccheroni, F., and Ok, E. A. (2004). Expected utility theory without the completeness axiom. *Journal of Economic Theory*, 115(1):118–133.
- Dwenger, N., Kübler, D., and Weizsäcker, G. (2014). Flipping a coin: Theory and evidence. CESifo Working Paper 4740, Center for Economic Studies and ifo Institute (CESifo), Munich.
- Dwenger, N., Kübler, D., and Weizsäcker, G. (2018). Flipping a coin: Evidence from university applications. *Journal of Public Economics*, 167:240–250.
- D’Acunतो, F., Hoang, D., Paloviita, M., and Weber, M. (2023). IQ, expectations, and choice. *Review of Economic Studies*, 90(5):2292–2325.
- Echenique, F., Lee, S., and Shum, M. (2011). The money pump as a measure of revealed preference violations. *Journal of Political Economy*, 119(6):1201–1223.
- Ells, E. (1982). *Rational Decision and Causality*. Cambridge University Press.

- Eliashberg, J. and Hauser, J. (1985). A measurement error approach for modeling consumer risk preference. *Management Science*, 31(1):1–25.
- Eliaz, K. and Ok, E. A. (2006). Indifference or indecisiveness? Choice-theoretic foundations of incomplete preferences. *Games and Economic Behavior*, 56(1):61–86.
- Engelmann, D. and Hollard, G. (2010). Reconsidering the effect of market experience on the “endowment effect”. *Econometrica*, 78(6):2005–2019.
- Enke, B. and Graeber, T. (2023). Cognitive uncertainty. *Quarterly Journal of Economics*, Forthcoming.
- Enke, B., Graeber, T., and Oprea, R. (2025). Complexity and time. *Journal of the European Economic Association*, page jvaf009.
- Enke, B. and Shubbatt, C. (2023). Quantifying lottery choice complexity. Working Paper w31677, National Bureau of Economic Research.
- Epper, T., Fehr, E., Fehr-Duda, H., Kreiner, C. T., Lassen, D. D., Leth-Petersen, S., and Rasmussen, G. N. (2020). Time discounting and wealth inequality. *American Economic Review*, 110(4):1177–1205.
- Ericson, K. and Fuster, A. (2011). Expectations as endowments: Evidence on reference-dependent preferences from exchange and valuation experiments. *Quarterly Journal of Economics*, 126(4):1879–1907.
- Ericson, K. and Fuster, A. (2014). The endowment effect. *Annual Review of Economics*, 6(1):555–579.
- Esponda, I. and Vespa, E. (2023). Contingent Thinking and the Sure-Thing Principle: Revisiting Classic Anomalies in the Laboratory. *Review of Economic Studies*.
- Estrada-Mejia, C., de Vries, M., and Zeelenberg, M. (2016). Numeracy and wealth. *Journal of Economic Psychology*, 54:53–63.
- Fagereng, A., Guiso, L., Malacrino, D., and Pistaferri, L. (2020). Heterogeneity and persistence in returns to wealth. *Econometrica*, 88(1):115–170.
- Fama, E. and French, K. (1992). The cross-section of expected stock returns. *Journal of Finance*, 47(2):427–465.
- Famulari, M. (1995). A household-based, nonparametric test of demand theory. *Review of Economics and Statistics*, 77(2):372–383.
- Fehr, D., Fink, G., and Jack, B. K. (2022). Poor and Rational: Decision-Making under Scarcity. *Journal of Political Economy*, 130(11):2862–2897.
- Fehr, D. and Kübler, D. (2022). The endowment effect in the general population. *CESifo Working Paper*, (10150).
- Feldman, P. and Rehbeck, J. (2022). Revealing a preference for mixtures: An experimental study of risk. *Quantitative Economics*, 13(2):761–786.
- Finke, M. S., Howe, J. S., and Huston, S. J. (2016). Old age and the decline in financial literacy. *Management Science*, 63(1):213–230.
- Fishburn, P. C. (1988). *Nonlinear Preference and Utility Theory*. Wheat-sheaf Books, Brighton, UK.
- Fishburn, P. C. (1991). Nontransitive preferences in decision theory. *Journal of Risk and Uncertainty*, 4(2):113–134.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4):25–42.
- Frydman, C. and Jin, L. J. (2021). Efficient coding and risky choice. *Quarterly Journal of Economics*, 137(1):161–213.
- Fudenberg, D., Iijima, R., and Strzalecki, T. (2014). Stochastic choice and revealed perturbed utility. Working paper.
- Fudenberg, D., Iijima, R., and Strzalecki, T. (2015). Stochastic choice and revealed perturbed utility. *Econometrica*, 83(6):2371–2409.
- Fudenberg, D. and Puri, I. (2021). Evaluating and extending theories of choice under risk. Working paper.
- Gebhardt, W. A. and 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.
- Gerasimou, G. (2017). Indecisiveness, undesirability and overload revealed through rational choice deferral. *Economic Journal*, 128:2450–2479.
- Gigerenzer, G. and Selten, R. (2001). *Bounded Rationality: The Adaptive Toolbox*. MIT Press.
- Gigerenzer, G., Todd, P. M., and the ABC Research Group (1999). *Simple Heuristics That Make Us Smart*. Oxford University Press.
- Gilboa, I. and Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18(2):141–153.
- Gill, D. and Prowse, V. (2016). Cognitive ability, character skills, and learning to play equilibrium: A level-k analysis. *Journal of Political Economy*, 124(6):1619–1676.
- Gillen, B., Snowberg, E., and Yariv, L. (2019). Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study. *Journal of Political Economy*, 127(4):1826–1863.
- Gonczarowski, Y. A., Heffetz, O., and Thomas, C. (2022). Strategyproofness-Exposing Mechanism Descriptions. Papers 2209.13148, arXiv.org.
- Goossens, J. T. and Knoef, M. (2022). Covid-19 crisis: Do extreme events affect preferences and trading behavior? Technical report, Tilburg University.

- Green, J. R. (1987). Making book against oneself: The independence axiom and nonlinear utility theory. *Quarterly Journal of Economics*, 102(4):785–796.
- Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1(1):114–125.
- Grether, D. M. and Plott, C. R. (1979). Economic theory of choice and the preference reversal phenomenon. *American Economic Review*, 69(5):623–638.
- Gul, F. and Pesendorfer, W. (2006). Random expected utility. *Econometrica*, 74(1):121–146.
- Halevy, Y. (2007). Ellsberg revisited: An experimental study. *Econometrica*, 75(2):503–536.
- Halevy, Y. and Mayraz, G. (2022). Identifying Rule-Based Rationality. *Review of Economics and Statistics*, pages 1–44.
- Halevy, Y., Walker-Jones, D., and Zrill, L. (2023). Difficult decisions. Working papers, University of Toronto, Department of Economics.
- Harless, D. W. and Camerer, C. F. (1994). The Predictive Utility of Generalized Expected Utility Theories. *Econometrica*, 62(6):1251–89.
- Herrmann, D. and Musshoff, O. (2016). Measuring time preferences: Comparing methods and evaluating the magnitude effect. *Journal of Behavioral and Experimental Economics*, 65:16–26.
- Herrmann, T., Hübner, O., Menkhoff, L., and Schmidt, U. (2017). Allais for the poor: Relations to ability, information processing, and risk attitudes. *Journal of Risk and Uncertainty*, 54(2):129–156.
- Hertwig, R., Barron, G., Weber, E. U., and Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, 15(8):534–539. PMID: 15270998.
- Hey, J. and Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica*, 62(6):1291–1326.
- Hogarth, R. M. and Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, 24(1):1–55.
- Holt, C. A. and Laury, S. K. (2002). Risk aversion and incentive effects. *American economic review*, 92(5):1644–1655.
- Houtman, M. and Maks, J. (1985). Determining all maximal data subsets consistent with revealed preference. *Kwantitatieve methoden*, 19(1):89–104.
- Huck, S. and Müller, W. (2012). Allais for all: Revisiting the paradox in a large representative sample. *Journal of Risk and Uncertainty*, 44(3):261–293.
- Isoni, A., Loomes, G., and Sugden, R. (2011). The willingness to pay–willingness to accept gap, the "endowment effect," subject misconceptions, and experimental procedures for eliciting valuations: Comment. *American Economic Review*, 101(2):991–1011.
- Iyengar, S. S. and Kamenica, E. (2010). Choice proliferation, simplicity seeking, and asset allocation. *Journal of Public Economics*, 94(7–8):530–539.
- Jacobson, S. and Petrie, R. (2009). Learning from mistakes: What do inconsistent choices over risk tell us? *Journal of Risk and Uncertainty*, 38:143–158.
- Johnson, E. J. and Goldstein, D. (2003). Do defaults save lives? *Science*, 302(5649):1338–1339.
- Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux, New York.
- Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1990). Experimental Tests of the Endowment Effect and the Coase Theorem. *Journal of Political Economy*, 98(6):1325–48.
- Kahneman, D., Slovic, P., and Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press.
- Kahneman, D. and Tversky, A. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131.
- Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–292.
- Kalenscher, T., Tobler, P., Huijbers, W., Daselaar, S., and Pennartz, C. (2010). Neural signatures of intransitive preferences. *Frontiers in Human Neuroscience*, 4.
- Kelsey, D. (1995). Dutch books arguments and learning in a nonexpected utility framework. *International Economic Review*, 36(1):187–206.
- Kelsey, D. and Milne, F. (1997). Induced preferences, dynamic consistency and dutch books. *Economica*, 64(255):471–481.
- Keren, G. and Wu, G., editors (2015). *The Wiley Blackwell Handbook of Judgment and Decision Making*. John Wiley & Sons, Ltd., Chichester, UK.
- Khaw, M. W., Li, Z., and Woodford, M. (2021). Cognitive imprecision and small stakes risk aversion. *The review of economic studies*, 88(4):1979–2013.
- Kim, H. B., Choi, S., Kim, B., and Pop-Eleches, C. (2018). The role of education interventions in improving economic rationality. *Science*, 362(6410):83–86.

- Klibanoff, P., Marinacci, M., and Mukerji, S. (2005). A smooth model of decision making under ambiguity. *Econometrica*, 73(6):1849–1892.
- Knetsch, J. (1989). The endowment effect and evidence of nonreversible indifference curves. *American Economic Review*, 79(5):1277–84.
- Koszegi, B. and Rabin, M. (2008). Choices, situations, and happiness. *Journal of Public Economics*, 92(8-9):1821–1832.
- Laibson, D. (1997). Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics*, 112(2):443–478.
- Laibson, D. and Yariv, L. (2007). Safety in markets: An impossibility theorem for dutch books. Working papers, Princeton University. Economics Department.
- Levitt, S. D. (2021). Heads or tails: The impact of a coin toss on major life decisions and subsequent happiness. *The Review of Economic Studies*, 88(1):378–405.
- Levitt, S. D. and List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*, 21(2):153–174.
- Li, S. (2017). Obviously strategy-proof mechanisms. *American Economic Review*, 107(11):3257–3287.
- List, J. A. (2003). Does Market Experience Eliminate Market Anomalies? *Quarterly Journal of Economics*, 118(1):41–71.
- List, J. A. (2011). Does market experience eliminate market anomalies? the case of exogenous market experience. *American Economic Review*, 101(3):313–17.
- Loomes, G. and Pogrebn, G. (2017). Do preference reversals disappear when we allow for probabilistic choice? *Management Science*, 63(1):166–184.
- Loomes, G., Starmer, C., and Sugden, R. (1991). Observing violations of transitivity by experimental methods. *Econometrica*, 59(2):425–439.
- Loomes, G., Starmer, C., and Sugden, R. (2003). Do anomalies disappear in repeated markets? *Economic Journal*, 113(486):C153–C166.
- Loomes, G. and Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *Economic Journal*, 92(368):805–824.
- Loomes, G. and Sugden, R. (1995). Incorporating a stochastic element into decision theories. *European Economic Review*, 39(3-4):641–648.
- Luce, R. and Suppes, P. (1965). *Preference, utility, and subjective probability*, chapter 3. in *Handbook of Mathematical Psychology*, Wiley, New York.
- Lucia, A. (2023). What drives violations of the independence axiom? the role of decision confidence. Working paper, California Institute of Technology (Caltech).
- Machina, M. (1989). Dynamic consistency and non-expected utility models of choice under uncertainty. *Journal of Economic Literature*, 27(4):1622–68.
- Machina, M. J. (1982). Expected utility analysis without the independence axiom. *Econometrica*, 50(2):277–323.
- Machina, M. J. (1985). Stochastic choice functions generated from deterministic preferences over lotteries. *Economic Journal*, 95(379):575–94.
- Mandler, M. (2005). Incomplete preferences and rational intransitivity of choice. *Games and Economic Behavior*, 50(2):255–277.
- Martínez-Marquina, A., Niederle, M., and Vespa, E. (2019). Failures in contingent reasoning: The role of uncertainty. *American Economic Review*, 109(10):3437–3474.
- Masatlioglu, Y. and Ok, E. (2005). Rational choice with status quo bias. *Journal of Economic Theory*, 121(1):1–29.
- McGranaghan, C. and Otto, S. (2022). Choice uncertainty and the endowment effect. *Journal of Risk and Uncertain*, 65:83–104.
- Meier, S. and Sprenger, C. (2010). Present-biased preferences and credit card borrowing. *American Economic Journal: Applied Economics*, 2(1):193–210.
- Mertens, S., Herberz, M., Hahnel, U. J. J., and Brosch, T. (2022). The effectiveness of nudging: A meta-analysis of choice architecture interventions across behavioral domains. *Proceedings of the National Academy of Sciences*, 119(1):e2107346118.
- Miao, B., Yang, S., and Zhong, S. (2024). Narrowly rational. Working paper, National University of Singapore.
- Miao, B. and Zhong, S. (2018). Probabilistic social preference: how Machina’s mom randomizes her choice. *Economic Theory*, 65(1):1–24.
- Modigliani, F. (1986). Life cycle, individual thrift, and the wealth of nations. *American Economic Review*, 76(3):297–313.
- Modigliani, F. and Brumberg, R. (1954). Utility analysis and the consumption function: An interpretation of cross-section data. In Kurihara, K. K., editor, *Post-Keynesian economics*, pages 388–436. Rutgers University Press, New Brunswick, NJ.
- Mosteller, F. and Noguee, P. (1951). An experimental measurement of utility. *Journal of Political Economy*, 59(5).
- Myagkov, M. and Plott, C. R. (1997). Exchange economies and loss exposure: Experiments exploring prospect theory and competitive equilibria in market environments. *American Economic Review*, 87(5):801–828.

- Nishimura, H. (2018). The transitive core: Inference of welfare from nontransitive preference relations. *Theoretical Economics*, 13(2).
- Nishimuray, H. and Ok, E. (2018). Preference structures. Working paper, New York University.
- Ok, E. A., Ortoleva, P., and Riella, G. (2012). Incomplete Preferences Under Uncertainty: Indecisiveness in Beliefs versus Tastes. *Econometrica*, 80(4):1791–1808.
- Ong, Q. and Qiu, J. (2023). Paying for randomization and indecisiveness. *Journal of Risk and Uncertainty*, 67:45–72.
- Oprea, R. (2024). Decisions under risk are decisions under complexity. *American Economic Review*, 114(12):3789–3811.
- Oprea, R. D. (2020). What makes a rule complex. *American Economic Review*, 110(12):3913–3951.
- Ortoleva, P. (2010). Status quo bias, multiple priors and uncertainty aversion. *Games and Economic Behavior*, 69(2):411–424.
- Plott, C. R. and Zeiler, K. (2005). The willingness to pay - willingness to accept gap, the "endowment effect," subject misconceptions, and experimental procedures for eliciting valuations. *American Economic Review*, 95(3):530–545.
- Plott, C. R. and Zeiler, K. (2007). Exchange asymmetries incorrectly interpreted as evidence of endowment effect theory and prospect theory? *American Economic Review*, 97(4):1449–1466.
- Plott, C. R. and Zeiler, K. (2011). The willingness to pay–willingness to accept gap, the "endowment effect," subject misconceptions, and experimental procedures for eliciting valuations: Reply. *American Economic Review*, 101(2):1012–28.
- Preston, A. and Wright, R. E. (2023). Gender, financial literacy and pension savings. *Economic Record*, 99(324):58–83.
- Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior & Organization*, 3(4):323–343.
- Rabin, M. and Bazerman, M. (2019). Fretting about modest risks is a mistake. *California Management Review*, 61(3):34–48.
- Raiffa, H. (1968). *Decision Analysis: Introductory Lectures on Choices under Uncertainty*. Addison-Wesley, Reading, Massachusetts.
- Ramsey, F. P. (1931). Truth and probability. In Braithwaite, R. B., editor, *The Foundations of Mathematics and Other Logical Essays*, pages 156–198. Kegan Paul, Trench, Trubner & Co. and Harcourt, Brace and Company, London and New York. Originally published in 1926.
- Rees-Jones, A. and Shorrer, R. (2023). Behavioral economics in education market design: A forward-looking review. *Journal of Political Economy Microeconomics*, 1(3):557–613.
- Regenwetter, M., Dana, J., and Davis-Stober, C. P. (2011). Transitivity of preferences. *Psychological Review*, 118(1):42–56.
- Rieskamp, J., Busemeyer, J. R., and Mellers, B. A. (2006). Extending the bounds of rationality: Evidence and theories of preferential choice. *Journal of Economic Literature*, 44(3):631–661.
- Rubinstein, A. (2002). Irrational diversification in multiple decision problems. *European Economic Review*, 46(8):1369 – 1378.
- Salant, Y. and Spenkuch, J. L. (2022). Complexity and Satisficing: Theory with Evidence from Chess. NBER Working Papers 30002, National Bureau of Economic Research, Inc.
- Samuelson, W. and Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, 1(1):7–59.
- Sautua, S. I. (2017). Does uncertainty cause inertia in decision making? an experimental study of the role of regret aversion and indecisiveness. *Journal of Economic Behavior and Organization*, 136:1 – 14.
- Schoemaker, P. J. H. (1982). The expected utility model: Its variants, purposes, evidence and limitations. *Journal of Economic Literature*, 20(2):529–563.
- Seidl, C. (2002). Preference reversal. *Journal of Economic Surveys*, 16(5):621–655.
- Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, 69(1):99–118.
- Smith, J. P., McArdle, J. J., and Willis, R. (2010). Financial Decision Making and Cognition in a Family Context. *Economic Journal*, 120(548):F363–F380.
- Smith, V. L. and Walker, J. M. (1993). Monetary rewards and decision cost in experimental economics. *Economic Inquiry*, 31(2):245–261.
- Smitzsky, G., Liu, W., and Gneezy, U. (2021). The endowment effect: Loss aversion or a buy-sell discrepancy? *Journal of Experimental Psychology: General*, pages 1–11.
- Snowberg, E. and Yariv, L. (2021). Testing the waters: Behavior across participant pools. *American Economic Review*, 111(2):687–719.
- Sprenger, C. (2015). An endowment effect for risk: Experimental tests of stochastic reference points. *Journal of Political Economy*, 123(6):1456–1499.
- Stango, V. and Zinman, J. (2022). We Are All Behavioural, More, or Less: A Taxonomy of Consumer Decision-Making. *Review of Economic Studies*, 90(3):1470–1498.

- Starmer, C. (2000). Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature*, 38(2):332–382.
- Starr, E. and Goldfarb, B. (2020). Binned scatterplots: A simple tool to make research easier and better. *Strategic Management Journal*, 41(12):2261–2274.
- Stillman, P. E., Krajbich, I., and Ferguson, M. J. (2020). Using dynamic monitoring of choices to predict and understand risk preferences. *Proceedings of the National Academy of Sciences*, 117(50):31738–31747.
- Swofford, J. L. and Whitney, G. A. (1987). Nonparametric test of utility maximization and weak separability for consumption, leisure, and money. *Review of Economics and Statistics*, 69(3):458–464.
- Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of economic behavior & organization*, 1(1):39–60.
- Thaler, R. H. and Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press, New Haven, CT.
- Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, 76(1):31–48.
- Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131.
- Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5:297–323.
- van de Kuilen, G. (2009). Subjective probability weighting and the discovered preference hypothesis. *Theory and Decision*, 67:1–22.
- van de Kuilen, G. and Wakker, P. (2006). Learning in the Allais paradox. *Journal of Risk and Uncertainty*, 33:155–164.
- Van Ewijk, C., Jacobs, B., and De Mooij, R. (2007). Welfare effects of fiscal subsidies on home ownership in the Netherlands. *De Economist*, 155(3):323.
- van Rooij, M. C., Lusardi, A., and Alessie, R. J. (2012). Financial Literacy, Retirement Planning and Household Wealth. *Economic Journal*, 122(560):449–478.
- Vanberg, C. (2008). Why do people keep their promises? An experimental test of two explanations. *Econometrica*, 76(6):1467–1480.
- Varian, H. R. (1990). Goodness-of-fit in optimizing models. *Journal of Econometrics*, 46(1-2):125–140.
- Vieider, F. M. (2021). Noisy coding of time and reward discounting. Working Paper 21/1036, Ghent University, Faculty of Economics and Business Administration.
- Vieider, F. M. (2023). Noisy coding of time and reward discounting. Working Papers of Faculty of Economics and Business Administration, Ghent University, Belgium 21/1036, Ghent University, Faculty of Economics and Business Administration.
- Von Neumann, J. and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton University Press.
- Weaver, R. and Frederick, S. (2012). A reference price theory of the endowment effect. *Journal of Marketing Research*, 49:696 – 707.
- Willadsen, H., Zaccagni, S., Piovesan, M., and Wengström, E. (2024). Measures of cognitive ability and choice inconsistency. *Journal of Economic Behavior & Organization*, 220:495–506.
- Woodford, M. (2020). Modeling imprecision in perception, valuation, and choice. *Annual Review of Economics*, 12(1):579–601.
- Wulff, D. U., Mergenthaler-Canseco, M., and Hertwig, R. (2018). A meta-analytic review of two modes of learning and the description-experience gap. *Psychological Bulletin*, 144(2):140–176.
- Yaari, M. E. (1985). On the role of “Dutch Books” in the theory of choice under risk. In Jacobs, D., Kalai, E., Kamien, M., and Schwartz, N., editors, *Frontiers of Research in Economic Theory: The Nancy L. Schwartz Memorial Lectures, 1983–1997*, Econometric Society Monographs, pages 33–46. Cambridge University Press, Cambridge.
- Zizzo, D. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, 13(1):75–98.

Appendices to chapter 2

A.1 Theoretical analysis

A.1.1 The theoretical analysis based on Cerreia-Vioglio et al. (2015) and Klibanoff et al. (2005)

To accommodate the potential that a decision-maker might not be fully confident about her choices, we assume an individual has multiple utility functions that we call multiple selves, with each self representing one particular way to trade off conflicting objectives in choices. Such a modelling technique has been used in models of incomplete preferences (see e.g., Bewley, 2002; Dubra et al., 2004; Cerreia-Vioglio et al., 2015).

Specifically, let u_τ denote the utility function of the self τ , and $\mathcal{T}$ denote the set of selves. Let π denote the subjective probability distribution over $\mathcal{T}$, which, similar to the modelling technique of Loomes and Sugden (1982), represents “the individual’s degree of belief or confidence in the occurrence of the corresponding states” (Loomes and Sugden, 1982, p. 807). This belief could come from introspection or experiences with similar options. Given a utility function u_τ , we follow the standard assumption that the self behaves according to EUT. Let $U_\tau(l)$ denote the expected utility of an option $l \in L$.¹⁹ We further assume that the individual dislikes disagreement among selves. This is because, to arrive at a choice when there are multiple selves with different preferences is, in essence, similar to situations where a group of people with different opinions tries to reach a consensus. The more strongly group members disagree with each other, the harder it is for the group to make compromises and agree on a single opinion. Hence, aversion to disagreement among selves can be interpreted as the cost of forcing different selves to reach a consensus. With the above assumptions, we can write the individual’s preference over an option l as:

$$V(l) = \int_{\mathcal{T}} \phi[U_\tau(l)] d\pi, \quad (2)$$

where concave $\phi(\cdot)$ implies an aversion to disagreement - deviations from the mean expected utility - among different selves. Similar to the connection between the concavity of the utility function and risk aversion, the concavity of $\phi(\cdot)$ implies that the individual places more weight on the selves who have lower value for l . Such a cautious attitude is consistent with Levitt (2021) who showed that subjects who have difficulties making a decision are often excessively cautious in the sense of preferring to maintain the status quo.

Equation 2 extends directly from Klibanoff et al. (2005) and Cerreia-Vioglio et al. (2015). It can be seen as a smooth version of the cautious expected utility model (Cerreia-Vioglio et al.,

¹⁹The function $U(\cdot)$ could be made more general to allow for non-EUT preferences to incorporate unsureness about how strongly to weight the extra factor, such as probability weighting or loss aversion, in a non-EU model.

2015). It is also a parallel of the smooth ambiguity model of Klibanoff et al. (2005). Indeed, in the smooth ambiguity model, an individual is unsure about the probability distribution of the states of nature, and she has a subjective belief over these probability distributions. Likewise, in this model, an individual is unsure about her utility function, and she has a subjective belief over her multiple selves. Note that this does not mean this model only applies to decision-making under risk. If there is preference uncertainty under risk (or even under certainty, e.g., over options about experience goods) because individuals have difficulties evaluating options, this uncertainty is also likely to be present in more complex situations of decision-making under ambiguity. In this sense, this model complements the smooth model of ambiguity and general models about uncertainty in beliefs. Ultimately, the lack of decision confidence arises from the difficulties in evaluating options, which may be due to uncertainty in both beliefs and preferences. A general model accommodating both sources of uncertainty could be written as:

$$V(a) = \int_{\mathcal{M}} \int_{\mathcal{T}} \phi [U_{\tau, \mu}(a)] d\pi d\mu,$$

where a represents an act, and μ is a subjective probability distribution over M , the set of probability distributions of the states of nature.

We are now ready to establish the link between decision confidence and the randomization probability in the randomized choices. Specifically, recall that in our mechanism, the individual chooses a randomization probability $\lambda \in [0, 1]$ and builds a lottery $(\lambda, x; (1-\lambda), y)$: She receives x with probability λ and y with probability $1 - \lambda$. Since for any given self τ , the individual's preference over the lottery $(\lambda, x; (1 - \lambda), y)$ satisfies EUT, we have $U_{\tau}[\lambda x + (1 - \lambda)y] = \lambda U_{\tau}(x) + (1 - \lambda)U_{\tau}(y)$. The individual's decision is then to maximize her utility by choosing the optimal randomization probability $0 \leq \lambda \leq 1$:

$$\text{Max}_{\lambda} V[\lambda x + (1 - \lambda)y] = \int_{\mathcal{T}} \phi [\lambda U_{\tau}(x) + (1 - \lambda)U_{\tau}(y)] d\pi.$$

In the experiment, y is a sure payment. Sure monetary payments are probably the easiest options to evaluate, hence we assume the individual is always confident about her evaluation of a sure payment: $U_{\tau}(y) = u(y)$, $\forall \tau \in \mathcal{T}$. Applying the Taylor expansion to the above equation at y , we can derive the optimal λ as:²⁰

$$\lambda^* \approx \frac{1}{-\frac{\phi''[u(y)]}{\phi'[u(y)]}} \times \frac{E_{\pi} [U_{\tau}(x)] - u(y)}{\sigma_x^2} \quad (3)$$

where $\sigma_x^2 = E_{\pi} [U_{\tau}(x) - E_{\pi}(U_{\tau}(x))]^2$ is the standard deviation of the valuation of the lottery across multiple selves and approximates how strongly different selves disagree with each

²⁰More precisely, since $0 \leq \lambda \leq 1$, $\lambda^* \approx \min \left\{ \max \left\{ 0, \frac{1}{-\frac{\phi''[u(y)]}{\phi'[u(y)]}} \times \frac{\Delta u}{\sigma_x^2} \right\}, 1 \right\}$. The detailed derivation can be found below.

other. Similar to decision-making under risk, $-\frac{\phi''(u(y))}{\phi'(u(y))}$ can be interpreted as a metric of attitudes towards disagreement among selves. Thus, the randomization probability aggregates the three important determinants of decision confidence: preference uncertainty, the utility difference between the two options, and her attitude toward preference uncertainty. It is in this sense we argue that the randomization probability captures decision confidence.

Deriving the hypotheses

To see how the individual may randomize for sure payments that yield similar utility as the lottery, notice that the certainty equivalent of the lottery is

$$\begin{aligned} u(CE_x) &= \int \phi[U_\tau(l)] d\pi \\ &\approx E_\pi \left\{ E_\pi[U_\tau(x)] + \phi'(E_\pi[U_\tau(x)]) [U_\tau - E_\pi[U_\tau(x)]] \right. \\ &\quad \left. + \frac{\phi''(E_\pi[U_\tau(x)])}{2} [U_\tau - E_\pi[U_\tau(x)]]^2 \right\} \\ &= E_\pi[U_\tau(x)] + \frac{\phi''(E_\pi[U_\tau(x)])}{2} \sigma_x^2. \end{aligned}$$

The optimal randomization probability at the sure payment which is equal to the certainty equivalent of the lottery ($u(y) = u(CE_x) = E_\pi[U_\tau(x)] + \frac{\phi''(E_\pi[U_\tau(x)])}{2} \sigma_x^2$) is

$$\begin{aligned} \lambda^* &\approx \frac{1}{-\frac{\phi''[u(CE_x)]}{\phi'[u(y)]}} \times \frac{E_\pi[U_\tau(x)] - u(CE_x)}{\sigma_x^2} \\ &= \frac{1}{2} \times \frac{\phi'[u(CE_x)] \phi''(E_\pi[U_\tau(x)])}{\phi''[u(CE_x)]}. \end{aligned}$$

When $\phi'[u(CE_x)]$ is close to one and the function $\phi(\cdot)$ is smoothly concave, which is likely to hold for options with moderate payoffs, the randomization probability is around 0.5. This implies that the individual would choose randomization probabilities close to 0.5 when two options yield similar utilities. Furthermore, the smallest sure payment that the individual chooses $\lambda^* < 1$ (the lower bound), and the largest sure payment that the individual chooses $\lambda^* > 0$ (the upper bound) are defined by $u(\underline{y}_x) = E_\pi[U_\tau(x)] - \frac{-\phi''[u(y)]}{\phi'[u(y)]} \sigma_x^2$, and $u(\bar{y}_x) = E_\pi[U_\tau(x)]$. The range of sure payments that the individual randomizes strictly is

$$u(\bar{y}_x) - u(\underline{y}_x) = \frac{-\phi''[u(y)]}{\phi'[u(y)]} \sigma_x^2,$$

which varies with preference uncertainty (σ_x^2).

Relating these results to our experiment, we expect subjects to have more preference uncertainty about a complex lottery than a simple lottery, as the individual may find it

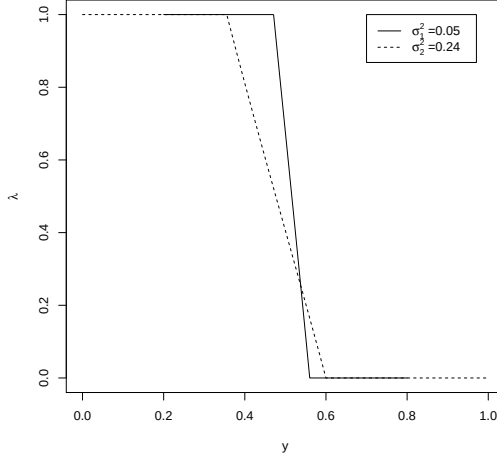


Figure A.1: The relationship between the randomization probability λ and the sure payment y . The figure is produced by assuming $\phi(U_\tau) = 1 - e^{-U_\tau}$, $\pi(u_1) = 0.6$, $\pi(u_2) = 0.4$, $U_1(x_1) = 0.8$ and $U_2(x_1) = 0.2$, $U_1(x_2) = 1.0$ and $U_2(x_2) = 0$, and $U_1(y) = U_2(y) = y$.

harder to evaluate a complex lottery. She considers relevant a larger set of utility functions and the subjective belief π becomes flatter. This translates into larger preference uncertainty (δ_x increases). Experience with a lottery, on the other hand, reduces preference uncertainty about the lottery because the individual attains clearer preferences about the lottery when she gains more experience (the set of utility functions becomes smaller and δ_x decreases). These lead to the hypotheses in the main text.

As a concrete illustration, consider the following numerical example: the individual has two selves $\tau = 1, 2$, and $\pi(u_1) = 0.6$, $\pi(u_2) = 0.4$. The individual's preference over the lottery x_1 is such that $U_1(x_1) = 0.8$ and $U_2(x_1) = 0.2$. Her preference over the lottery x_2 is such that $U_1(x_2) = 1.0$ and $U_2(x_2) = 0$. Thus, the individual perceives more preference uncertainty about the lottery x_2 than the lottery x_1 ($\sigma_{x_1} = 0.05 < \sigma_{x_2} = 0.24$). Option y is a sure payment, and $u_1(y) = u_2(y) = y$. The function $\phi(U_\tau) = 1 - e^{-U_\tau}$. Simple calculation shows that $\lambda_{x_1} = -\frac{1}{0.8-0.2} \ln(\frac{0.4}{0.6} \times \frac{y}{1-y})$ and $\lambda_{x_2} = -\frac{1}{1-0} \ln(\frac{0.4}{0.6} \times \frac{y}{1-y})$, subject to $0 \leq \lambda \leq 1$. Figure A.1 shows the relationship between the optimal λ and sure payment y . The Figure shows that the randomization probability decreases with y , and approaches to 0.5 for y that yields similar decision utility as the lottery ($y = 0.515$ for x_1 and $y = 0.476$ for x_2). Furthermore, the individual randomizes over a wider range of y for x_2 which she perceives higher preference uncertainty compared to x_1 .

Derivation of the optimal λ^*

Taking the first order derivative of the optimization equation gives:²¹

$$\begin{aligned} \frac{dV[\lambda x + (1-\lambda)y]}{d\lambda} &= \int_{\mathcal{T}} \phi' [\lambda U_{\tau}(x) + (1-\lambda)u(y)] \\ &\quad \times [U_{\tau}(x) - u(y)] d\pi = 0. \end{aligned}$$

Note that $U_{\tau}(x)$ is a random variable governed by the subjective probability distribution π . Let $X = U_{\tau}(x)$, and $\Delta_{\tau} = X - u(y)$. With these notations, we have

$$\phi' [\lambda U_{\tau}(x) + (1-\lambda)u(y)] = \phi' [u(y) + \lambda\Delta_{\tau}].$$

We are most interested in scenarios where the individual finds it difficult to choose between x and y , i.e., when the two options are close and Δ_{τ} is small relative to X and $u(y)$. When this is the case, we can use the Taylor expansion at y and obtain

$$\begin{aligned} \phi' [u(y) + \lambda\Delta_{\tau}] &= \phi'(u(y)) + \phi''(u(y))\lambda\Delta_{\tau} + O(\lambda\Delta_{\tau}) \\ &\approx \phi'(u(y)) + \phi''(u(y))\lambda\Delta_{\tau}, \end{aligned}$$

where $O(\lambda\Delta_{\tau})$ is the sum of the terms that have $\lambda\Delta_{\tau}$ with a power of two or higher. The above first order condition can be written as

$$\begin{aligned} \frac{dV[\lambda x + (1-\lambda)y]}{d\lambda} &= \int_{\mathcal{T}} \phi' [u(y) + \lambda\Delta_{\tau}] \Delta_{\tau} d\pi, \\ &\approx \int_{\mathcal{T}} [\phi'(u(y)) + \phi''(u(y))\lambda\Delta_{\tau}] \Delta_{\tau} d\pi \\ &= E_{\pi} [\phi'(u(y))\Delta_{\tau}] + \lambda E_{\pi} [\phi''(u(y))\Delta_{\tau}^2] = 0, \end{aligned}$$

where $E_{\pi}(\cdot)$ is the expectation operator with respect to the distribution π . Solving for λ , we have:

$$\lambda^* \approx \min \left\{ \max \left\{ 0, \frac{1}{-\frac{\phi''[u(y)]}{\phi'[u(y)]}} \times \frac{\Delta_u}{\sigma_x^2 - \Delta_u^2} \right\}, 1 \right\} \approx \min \left\{ \max \left\{ 0, \frac{1}{-\frac{\phi''[u(y)]}{\phi'[u(y)]}} \times \frac{\Delta_u}{\sigma_x^2} \right\}, 1 \right\}$$

where $\Delta_u = E_{\pi} [U_{\tau}(x)] - u(y)$ is the (expected) utility difference of x and y , $\sigma_x^2 = E_{\pi} [U_{\tau}(x) - E_{\pi}(U_{\tau}(x))]^2$ is the standard deviation of $U_{\tau}(x)$.

²¹The second-order derivative is $\frac{d^2V[\lambda x + (1-\lambda)y]}{d\lambda^2} = \int_{\mathcal{T}} \phi'' [\lambda U_{\tau}(x) + (1-\lambda)u(y)] \times [U_{\tau}(x) - u(y)]^2 d\pi$. Since $\phi(\cdot)$ is concave, $\phi''(\cdot)$ is negative. We are interested in situations where options x and y are not the same, i.e., $U_{\tau}(x) \neq u(y)$ for some $\tau \in \mathcal{T}$. Together we have $\phi'' [\lambda U_{\tau}(x) + (1-\lambda)u(y)] \times [U_{\tau}(x) - u(y)]^2 \leq 0$, and the inequality is strict for some $\tau \in \mathcal{T}$. Consequently, $\frac{d^2V[\lambda x + (1-\lambda)y]}{d\lambda^2} = \int_{\mathcal{T}} \phi'' [\lambda U_{\tau}(x) + (1-\lambda)u(y)] \times [U_{\tau}(x) - u(y)]^2 d\pi < 0$. This ensures we are indeed seeking for the maximum.

A.1.2 The theoretical analysis based on Fudenberg et al. (2015)

Below, we perform a theoretical analysis of our experiment based on Fudenberg et al. (2015) to demonstrate the links between randomization probabilities and decision confidence.²² Fudenberg et al.'s (2015) original representation concerns final outcomes. To apply their model to our experiments with lotteries, we write the individual's preference over randomizing between lottery x and sure payment y as:²³

$$V(\lambda, x; 1 - \lambda, y) = \lambda U(x) - c(\lambda) + (1 - \lambda)u(y) - c(1 - \lambda),$$

where $U(x)$ is the expected utility of the lottery x and $c(\lambda)$ is a weak cost function with finite steepness (the first order derivative of the cost function at the limit of 0 is not infinite). Using the weak cost function allows the model to accommodate zero choice probability that is present in our experiment. The cost function captures the implementation costs of making the desired choice, such as time and cognitive resources. In the Fudenberg et al.'s (2015) main representation, the cost function is independent of the option and the choice set. In an earlier version of their paper (Fudenberg et al., 2014), they proposed two extensions (item-invariant and menu-invariant APU) in which the cost function may depend on the preference uncertainty over options or the choice problem. We consider these two extensions to examine the effects of our treatments (increasing the complexity of the lottery or increasing subjects' experience with the lottery) on the cost function.

When $c(\lambda)$ is strictly convex, there exists an optimal randomization probability λ^* which maximizes the individual's utility, as defined by the equation $c'(\lambda^*) - c'(1 - \lambda^*) = U(x) - u(y)$, where $c'(\lambda^*) - c'(1 - \lambda^*)$ measures the convexity of the cost function $c''(\cdot)$. While the exact value of the optimal randomization probability depends on the cost function, some observations are in order. First, the optimal randomization probability approaches 0.5 when $U(x)$ is close to $u(y)$. Second, for the same utility difference between the two options, the individual chooses a randomization closer to 0.5 when the cost function is more convex. More generally, as Proposition 3 in Fudenberg et al. (2015) demonstrates, the individual becomes less selective and randomizes more when $c''(\cdot)$ increases. Third, simple calculations show that the largest sure payment that the individual chooses $\lambda^* = 0.9$ (the lower bound) is $u(\underline{y}) = U(x) - \Delta$, and the smallest sure payment she chooses $\lambda^* = 0.1$ (the upper bound) is $u(\bar{y}) = U(x) + \Delta$, where $\Delta = c'(0.9) - c'(0.1) > 0$.²⁴ Thus, the individual randomizes over

²²Cerreia-Vioglio et al. (2019) predict preference for randomization when the individual faces non-degenerated lotteries. However, when one of the two options is a sure payment, as in our experiment, the individual has no preference for randomization. This follows directly from the axiom of Weak Stochastic Certainty Effect.

²³Cerreia-Vioglio et al. (2019, Footnote 22, p.2437) proposed an alternative approach in which the individual integrates the lottery and the sure payment into a compound lottery, applies the reduction of the compound lottery, and implements the cost function to each outcome. We illustrate their approach and point out the differences between the two below. In particular, that approach predicts that the optimal randomization probability for the pair of the lottery and the sure payment that the individual is indifferent with depends on the number of outcomes in the lottery.

²⁴The values of 0.1 and 0.9 were chosen to accommodate experimental data.

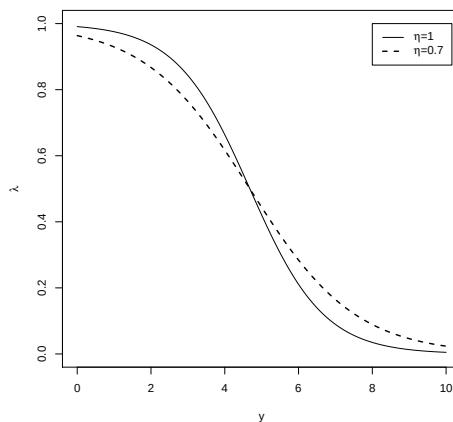


Figure A.2: The relationship between the optimal randomization probability λ^* and the sure payments y . The figure is produced according to the logit/logistic choice rule $\lambda^* = \frac{e^{U(x)/\eta}}{e^{U(x)/\eta} + e^{u(y)/\eta}}$. The parameter η captures the preference uncertainty over lottery x , with a larger η implying more convexity in the cost function and thus more preference uncertainty.

a larger range of sure payments when the cost function is more convex ($u(\bar{y}) - u(\underline{y}) = 2\Delta$). According to Fudenberg et al. (2015), the cost function may depend, among other things, on the individual's perceived preference uncertainty over the options and her attitude towards uncertainty. Using this interpretation of the cost function, the three properties of randomization probabilities correspond to the three properties of decision confidence we outlined in the main body of the paper. It is in this sense that we say randomization probabilities measure decision confidence.

If we are willing to make more specific assumptions about the cost function, we can obtain a direct solution of the optimal randomization probability. For example, when the cost function takes the form of $c(\lambda) = \eta \lambda \log(\lambda)$, we can derive the familiar logit/logistic choice rule:

$$\lambda^* = \frac{e^{U(x)/\eta}}{e^{U(x)/\eta} + e^{u(y)/\eta}}. \quad (4)$$

As shown by Holman and Marley, the parameter η can be linked to the variance of the i.i.d. Gumbel preference shocks in a random utility representation (Luce and Suppes, 1965, p.338). In the context of our study, η can be interpreted as the individual's preference uncertainty about lottery x . Figure A.2 depicts the relationship between the optimal randomization probability λ^* and the sure payments y . As we can see, randomization probabilities decrease with the value of y and approach 0.5 when the two options have similar utilities. Furthermore, when η increases, the cost function becomes more convex

and the individual's randomization probabilities become more compressed (the dashed line) and closer to 0.5.

Individuals may perceive more preference uncertainty over the complex lottery than over the simple lottery ($\Delta_c > \Delta_s$, where c denotes the complex lottery and s denotes the simple lottery), and experience with the lottery may reduce preference uncertainty about the lottery ($\Delta_e < \Delta_n$, where e denotes experience and n denotes no experience). In light of our analysis above, we expect that subjects' randomization probabilities are closer to 0.5 and that they randomize strictly over a wider range of sure payments when they make decisions about the complex lottery than when they make decisions about the simple lottery. In addition, compared to the no-experience treatment, randomization probabilities of subjects in the experience treatments are stretched away from 0.5, and subjects randomize strictly over a smaller range of sure payments. Figure A.3 demonstrates the effects.

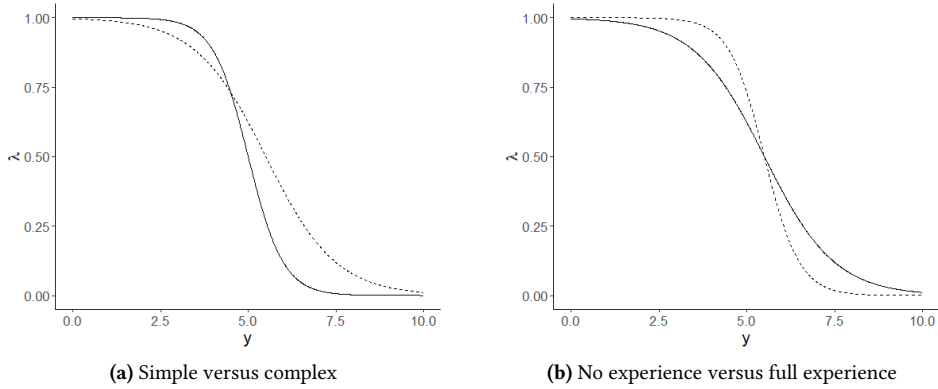


Figure A.3: The effects of complexity and experience on the lower bound, the upper bound, and the size of randomization range.

Cerreia-Vioglio et al. (2019)'s approach

Cerreia-Vioglio et al. (2019, Footnote 22, p.2437) proposed an alternative approach to apply Fudenberg et al.'s (2015) model to lotteries. We illustrate their approach with the following example. Consider an individual who faces a choice between a sure payment y and a lottery $x = 9_{0.5}1$ which pays 9 or 1 with equal likelihood. Cerreia-Vioglio et al. (2019) treat the randomized choice as a compound lottery. With the reduction of the compound lottery, the randomization of $(\lambda, x; 1 - \lambda, y)$ becomes $9_{0.5\lambda}1_{0.5\lambda}y$, and the individual' preference over $9_{0.5\lambda}1_{0.5\lambda}y$ is

$$\begin{aligned} V(\lambda, x; 1 - \lambda, y) &= 0.5\lambda u(9) - c(0.5\lambda) + 0.5\lambda u(1) - c(0.5\lambda) \\ &\quad + (1 - \lambda)u(y) - c(1 - \lambda) \\ &= \lambda U(x) - 2c(0.5\lambda) + (1 - \lambda)u(y) - c(1 - \lambda). \end{aligned}$$

This formulation predicts an optimal randomization probability of $2/3$ when the expected utility of the lottery is close to the utility of the sure payment ($c'(0.5\lambda) - c'(1 - \lambda) = U(x) - u(y) = 0 \Rightarrow \lambda = 2/3$). The intuition is that the above formulation rewards the individual for randomizing over *more outcomes*, and thus the individual assigns a higher randomization probability to lotteries with more outcomes. It can be shown that, when the lottery x has four outcomes which are equally likely, the optimal randomization probability is $\lambda = 4/5$ when $U(x) = u(y)$. These predictions are different from those obtained based on Fudenberg et al. (2015)'s approach.

A.1.3 The asymmetric treatment effects on the lower and upper bound of randomization range

We illustrate the asymmetric treatment effects on the lower bound and the upper bound of randomization range in this section. Recall that $\underline{y}$ denotes the largest sure payment that the individual chooses $\lambda^* < 1$ (the lower bound) and $\bar{y}$ denotes the smallest sure payment she chooses $\lambda^* > 0$ (the upper bound).

In the model extended from Cerreia-Vioglio et al. (2015) and Klibanoff et al. (2005),

$$\begin{aligned} u(\bar{y}_x) &= E_\pi [U_\tau(x)], \\ \underline{y} &= E_\pi [U_\tau(x)] - \frac{-\phi'' [u(y)]}{\phi' [u(y)]} \sigma_x^2. \end{aligned}$$

The changes in the upper and lower bounds depend on both $E_\pi [U_\tau(x)]$ and σ_x^2 . We observe that subjects on average valued the complex lottery higher than the simple lottery (mean CE of 4.68 for the simple lottery versus 4.98 for the complex lottery in Experiment 2, $p < 0.01$). Since the complex lottery has a larger σ_x^2 and the average valuation of the lottery is $E_\pi [U_\tau(x)] - \frac{-\phi''(E_\pi[U_\tau(x)])}{2} \sigma_x^2$, this implies an increase in $E_\pi [U_\tau(x)]$ for the complex lottery. The increase in $E_\pi [U_\tau(x)]$ increases both the upper bound and the lower bound, while the increase in σ_x^2 decreases only the lower bound. Together, they imply that the treatment effect on the upper bound could be larger than on the lower bound. Similarly, we observe an increase, albeit small, in the valuation of the complex lottery in the full-experience treatment (mean CE of 4.98 in the no-experience treatment versus 5.07 in the full-experience treatment in Experiment 2, $p > 0.10$). The increase in $E_\pi [U_\tau(x)]$ increases both the upper and lower bounds, and the decrease in σ_x^2 increases the lower bound further. Consequently, the treatment effect could be stronger on the lower bound than on the upper bound.

The analysis based on Fudenberg et al. (2015) follows similarly. In Fudenberg et al. (2015), $\bar{y} = EU(x) + \Delta$, $\underline{y} = EU(x) - \Delta$. The changes in the upper and lower bounds depend on both $EU(x)$ and Δ . Since the average valuation of the lottery is $EU(x)$, the higher average valuation of the complex lottery implies higher $EU(x)$ of the complex lottery compared to the simple lottery. Higher $EU(x)$ and Δ imply a stronger treatment effect on the upper bound than on the lower bound. Likewise, an increase in experience level is associated with an increase in $EU(x)$ and a decrease in Δ , which jointly imply a stronger treatment effect on the lower bound than the upper bound.

A.2 Additional figures and tables

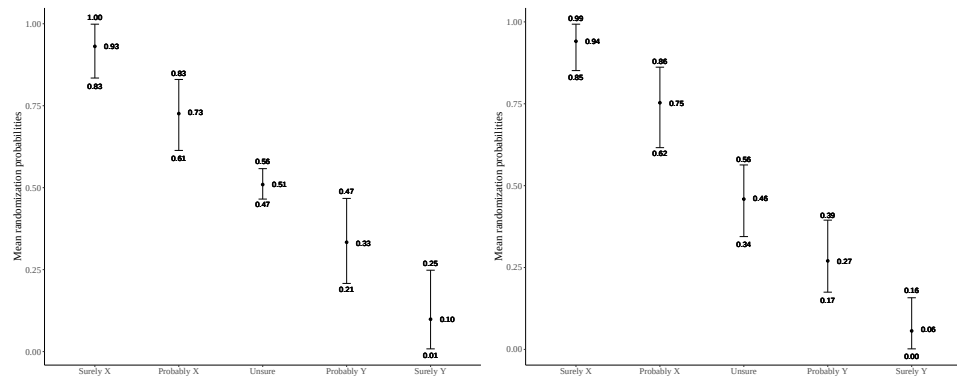


Figure A.4: The mean randomization probabilities at each confidence statement. The bars show the average minimum and maximum values. The values show the aggregate values for the baseline treatment – simple lottery, no-experience – in Experiment 1 (left) and Experiment 2 (right). The mean, minimum, and maximum values for the separate treatments in each of the experiments can be found in Table A.2. These values are broadly consistent with the cutoff probabilistic confidence levels of each confidence statement reported in Vanberg (2008, Footnote 10, p.1472: the probabilistic confidence level of 0.85 as the cutoff between surely and probably, 0.68 as the cutoff between probably and unsure, and 0.50 as the mean value for unsure). The minimum randomization probabilities were 0.83 and 0.85 for “Surely x ” and 0.61 and 0.62 for “Probably x ,” and the mean randomization probabilities were 0.51 and 0.46 for “Unsure” in Experiments 1 and 2 respectively.

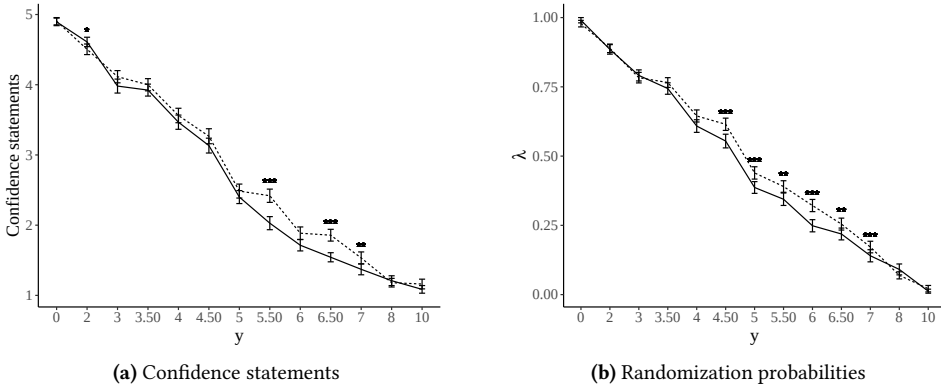


Figure A.5: The mean self-reported decision confidence and randomization probabilities for each value of y obtained from decisions about the simple lottery (solid line) and decisions about the complex lottery (dashed line) in Experiment 1. Wilcoxon signed-rank tests were performed to test the difference between the simple lottery and the complex lottery for each value of y : * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

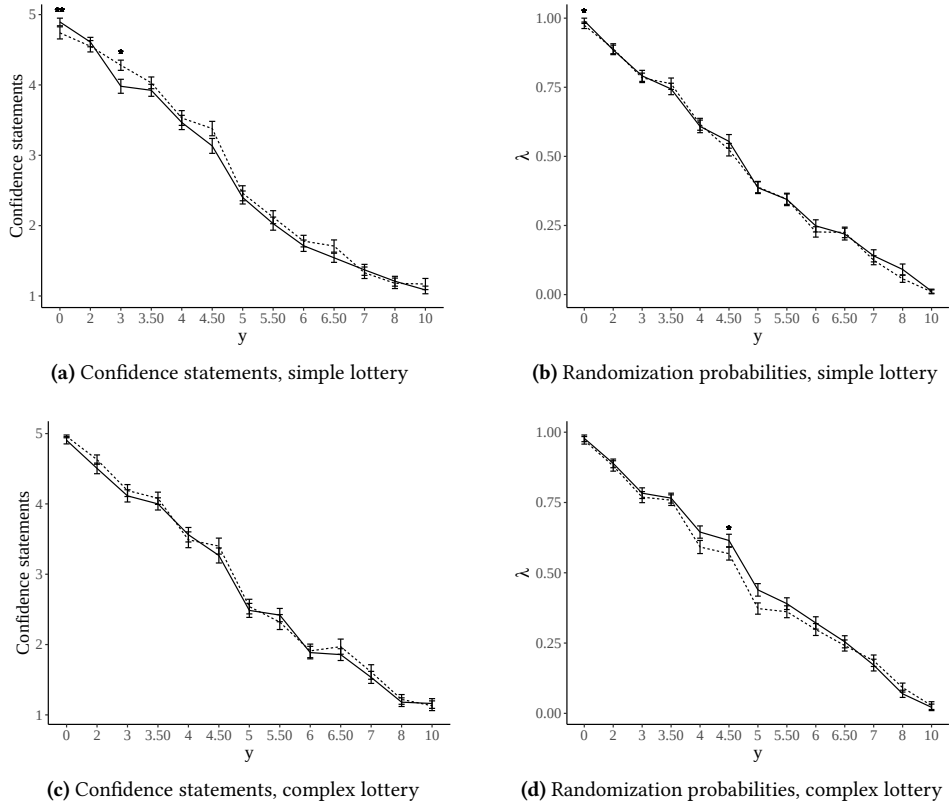


Figure A.6: The mean self-reported decision confidence and randomization probabilities for each value of y in the no-experience treatment (solid line) and the partial-experience treatment (dashed line) in Experiment 1. Wilcoxon rank-sum tests were performed to test the difference between the partial-experience treatment and no-experience treatment for each value of y : * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

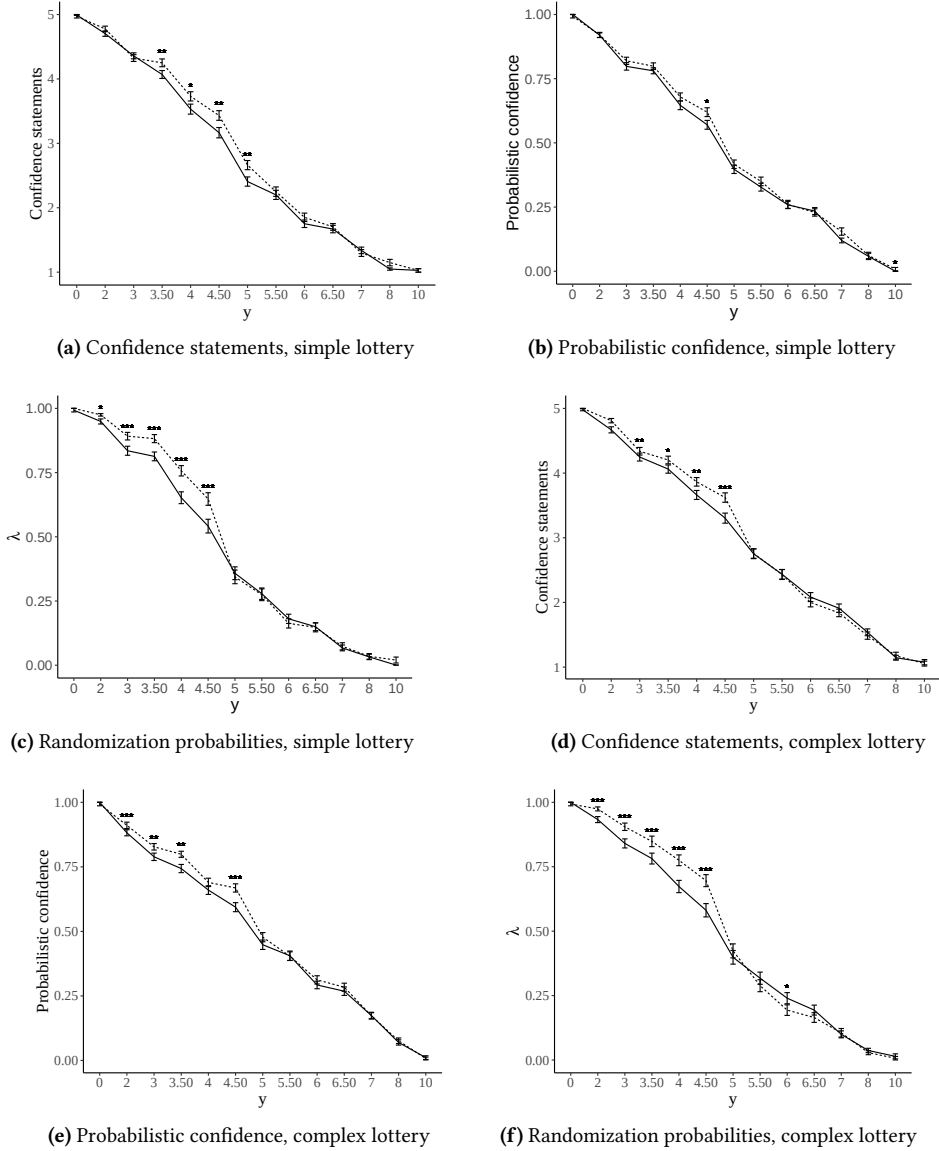


Figure A.7: The mean self-reported decision confidence and randomization probabilities for each value of y in the no-experience treatment (solid line) and the full-experience treatment (dashed line) in Experiment 2. Wilcoxon rank-sum tests were performed to test the difference between full-experience treatment and no-experience treatment for each value of y : * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Lottery	Treatment		Correlation between randomization probabilities and confidence statements		
			prob. confidence		
			Experiment 1	Experiment 2	Experiment 2
Simple	No experience	10th percentile	0.60	0.71	0.73
		25th percentile	0.78	0.82	0.82
		median	0.91	0.89	0.90
		75th percentile	0.95	0.94	0.96
		90th percentile	0.97	0.96	0.98
	Experience	10th percentile	0.60	0.78	0.77
		25th percentile	0.85	0.85	0.85
		median	0.93	0.90	0.91
		75th percentile	0.96	0.95	0.96
		90th percentile	0.97	0.97	0.99
Complex	No experience	10th percentile	0.69	0.67	0.64
		25th percentile	0.83	0.77	0.83
		median	0.90	0.88	0.89
		75th percentile	0.94	0.93	0.95
		90th percentile	0.97	0.96	0.97
	Experience	10th percentile	0.62	0.69	0.77
		25th percentile	0.81	0.80	0.84
		median	0.88	0.90	0.90
		75th percentile	0.94	0.94	0.95
		90th percentile	0.96	0.97	0.97

Table A.1: Nonparametric Spearman correlation between randomization probabilities and the two self-reported confidence measures at the 10th percentile, 25th percentile, median, 75th percentile, and 90th percentile in the two experiments for each lottery and experience treatment group.

Treatment	Lottery	Surely x	Probably x	Unsure	Probably y	Surely y		
Experiment 1								
No-experience	Simple	Mean	0.93 (0.011)	0.73 (0.016)	0.51 (0.023)	0.33 (0.018)	0.10 (0.010)	
		Min	0.83 (0.027)	0.61 (0.022)	0.47 (0.027)	0.21 (0.019)	0.01 (0.005)	
		Max	1 (0.001)	0.83 (0.015)	0.56 (0.025)	0.47 (0.026)	0.25 (0.025)	
	Complex	Mean	0.92 (0.012)	0.72 (0.015)	0.56 (0.015)	0.35 (0.017)	0.09 (0.011)	
		Min	0.82 (0.024)	0.62 (0.019)	0.49 (0.023)	0.24 (0.020)	0.01 (0.005)	
		Max	0.99 (0.006)	0.82 (0.016)	0.63 (0.020)	0.46 (0.018)	0.22 (0.026)	
	Partial-experience	Simple	Mean	0.90 (0.018)	0.68 (0.017)	0.52 (0.021)	0.33 (0.018)	0.10 (0.012)
			Min	0.79 (0.0129)	0.57 (0.021)	0.45 (0.022)	0.24 (0.021)	0.01 (0.008)
			Max	0.98 (0.013)	0.79 (0.019)	0.59 (0.026)	0.42 (0.021)	0.24 (0.027)
Complex		Mean	0.89 (0.015)	0.69 (0.016)	0.51 (0.023)	0.31 (0.016)	0.11 (0.013)	
		Min	0.77 (0.029)	0.55 (0.023)	0.44 (0.027)	0.21 (0.018)	0.01 (0.008)	
		Max	0.98 (0.009)	0.82 (0.016)	0.58 (0.028)	0.42 (0.022)	0.22 (0.024)	
Experiment 2								
No-experience	Simple	Mean	0.94 (0.010)	0.75 (0.018)	0.46 (0.023)	0.27 (0.020)	0.06 (0.008)	
		Min	0.85 (0.022)	0.62 (0.027)	0.34 (0.026)	0.17 (0.020)	0 (0.001)	
		Max	0.99 (0.007)	0.86 (0.017)	0.56 (0.029)	0.39 (0.026)	0.16 (0.022)	
	Complex	Mean	0.95 (0.008)	0.73 (0.019)	0.46 (0.025)	0.22 (0.017)	0.05 (0.009)	
		Min	0.88 (0.018)	0.59 (0.028)	0.34 (0.027)	0.13 (0.015)	0 (0)	
		Max	1 (0)	0.86 (0.017)	0.58 (0.029)	0.34 (0.026)	0.13 (0.021)	
Full-experience	Simple	Mean	0.95 (0.008)	0.79 (0.017)	0.51 (0.026)	0.22 (0.019)	0.06 (0.010)	
		Min	0.87 (0.022)	0.68 (0.026)	0.39 (0.031)	0.12 (0.018)	0.01 (0.007)	
		Max	1 (0.002)	0.89 (0.015)	0.62 (0.031)	0.34 (0.026)	0.16 (0.023)	
	Complex	Mean	0.95 (0.009)	0.78 (0.018)	0.50 (0.025)	0.24 (0.019)	0.51 (0.010)	
		Min	0.87 (0.022)	0.65 (0.027)	0.39 (0.030)	0.15 (0.019)	0.01 (0.007)	
		Max	1 (0.001)	0.89 (0.015)	0.62 (0.028)	0.36 (0.026)	0.13 (0.021)	

Table A.2: The mean, minimum, and maximum randomization probabilities that correspond to each confidence statement for all treatments in the two experiments. The values in parentheses are the standard errors of the mean.

	Lottery	Randomization probabilities	Confidence statements	Probabilistic confidence
Experiment 1				
Lower bound	Simple	2.99	2.95	
	Complex	2.84*	2.94	
Upper bound	Simple	6.61	6.30	
	Complex	6.90***	6.56**	
Range size	Simple	3.63	3.36	
	Complex	4.06***	3.62	
Experiment 2				
Lower bound	Simple	3.16	3.03	2.63
	Complex	3.19	2.99	2.59
Upper bound	Simple	6.18	6.19	7.00
	Complex	6.38***	6.5***	7.21***
Range size	Simple	3.03	3.15	4.37
	Complex	3.19*	3.58***	4.63***

Table A.3: Comparisons of the lower bound, the upper bound, and the range size between the simple lottery and complex lottery in the no-experience treatment in the two experiments. The lower bound, the upper bound, and the range sizes are defined by randomization probabilities ($0.10 \leq \lambda \leq 0.90$), confidence statements (“Probably x ”, “Unsure”, “Probably y ”) and probabilistic confidence (between “90% x , 10% y ” and “10% x , 90% y ”). Wilcoxon signed-rank tests were performed to test the difference between the simple lottery and the complex lottery for each measure: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	Self-reported probabilistic confidence									
	100% x 0% y	90% x 10% y	80% x 20% y	70% x 30% y	60% x 40% y	40% x 60% y	30% x 70% y	20% x 80% y	10% x 90% y	0% x 100% y
Simple lottery, no-experience treatment										
Rand. prob.	0.98 (0.008)	0.86 (0.021)	0.82 (0.022)	0.73 (0.022)	0.58 (0.027)	0.36 (0.024)	0.23 (0.023)	0.18 (0.022)	0.09 (0.018)	0.02 (0.004)
Complex lottery, no-experience treatment										
Rand. prob.	0.97 (0.007)	0.85 (0.024)	0.76 (0.026)	0.73 (0.024)	0.59 (0.025)	0.33 (0.024)	0.21 (0.021)	0.18 (0.024)	0.08 (0.016)	0.04 (0.012)
Simple lottery, full-experience treatment										
Rand. prob.	0.97 (0.007)	0.92 (0.016)	0.85 (0.024)	0.76 (0.025)	0.63 (0.026)	0.37 (0.025)	0.23 (0.024)	0.16 (0.020)	0.07 (0.015)	0.03 (0.009)
Complex lottery, full-experience treatment										
Rand. prob.	0.98 (0.007)	0.92 (0.017)	0.81 (0.025)	0.71 (0.028)	0.62 (0.027)	0.35 (0.023)	0.20 (0.024)	0.16 (0.027)	0.06 (0.015)	0.02 (0.008)

Table A.4: The mean randomization probabilities at each self-reported probabilistic confidence level in Experiment 2 for each lottery and experience treatment group. The standard errors of the mean are reported in the parentheses. We compute the mean randomization probability at each level of probabilistic confidence for each subject before taking its mean across subjects.

Lottery		Experience	Randomization probabilities	Confidence statements	Probabilistic confidence
Experiment 1					
Simple	Lower bound	No	2.99	2.95	
		Partial	2.75	2.80	
	Upper bound	No	6.61	6.30	
		Partial	6.60	6.25	
	Range size	No	3.63	3.36	
		Partial	3.85	3.45	
Complex	Lower bound	No	2.84	2.94	
		Partial	2.74	3.03	
	Upper bound	No	6.90	6.56	
		Partial	6.90	6.52	
	Range size	No	4.06	3.62	
		Partial	4.16	3.49	
Experiment 2					
Simple	Lower bound	No	3.16	3.03	2.63
		Full	3.44**	3.26*	2.71
	Upper bound	No	6.18	6.19	7.00
		Full	6.18	6.30	7.10
	Range size	No	3.03	3.15	4.37
		Full	2.74	3.04	4.38
Complex	Lower bound	No	3.19	2.99	2.59
		Full	3.60***	3.29**	2.73*
	Upper bound	No	6.38	6.57	7.21
		Full	6.27	6.45	7.31
	Range size	No	3.19	3.58	4.63
		Full	2.67**	3.16**	4.58

Table A.5: Comparisons of the lower bound, the upper bound, and the range size between the no-experience treatment and experience treatments by lottery type in the two experiments. The lower bound, the upper bound, and the range sizes are defined by randomization probabilities ($0.10 \leq \lambda \leq 0.90$), confidence statements (“Probably x ”, “Unsure”, “Probably y ”) and probabilistic confidence (between “90% x , 10% y ” and “10% x , 90% y ”). Wilcoxon rank-sum tests were performed to test the difference between experience treatment and no-experience treatment for each measure: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.2.1 Order effects in experiment 2

We selected three different orders and randomly assigned subjects to each order: Order 1) binary choices and confidence statements $\rightarrow$ randomized choices $\rightarrow$ probabilistic confidence, Order 2) randomized choices $\rightarrow$ binary choices and confidence statements $\rightarrow$ probabilistic confidence, Order 3) probabilistic confidence $\rightarrow$ binary choices and confidence statements $\rightarrow$ randomized choices. Order 1 is similar to the task order in Experiment 1, allowing us to assess the robustness of the findings in Experiment 1. Order 2 removes the potential priming effects of the self-reported confidence measures on randomized choices. Order 3 preserves the potential of the priming effects, but allows us to look at the probabilistic confidence measure when it is elicited first.

Given our proposal to capture decision confidence with randomization probabilities, the question that is most relevant to us is whether the subjects randomized differently when they made randomization choices prior to and after they completed the self-reported confidence measures.

	Simple		Complex	
	No experience	Experience	No experience	Experience
Order 1	0.456	0.368	0.452	0.373
Order 2	0.345	0.331	0.369	0.328
Order 3	0.471	0.412	0.498	0.400
Wilcoxon rank-sum tests:				
Order 1 vs Order 2	$p < 0.10$	$p = 0.625$	$p = 0.217$	$p = 0.541$
Order 2 vs Order 3	$p < 0.05$	$p = 0.192$	$p < 0.05$	$p = 0.194$

Table (a)

	$y = 0$ to 3.5	$y = 4$ to 6	$y = 6.5$ to 10
Order 1	0.303	0.622	0.267
Order 2	0.272	0.514	0.197
Order 3	0.339	0.650	0.296
Wilcoxon rank-sum tests:			
Orders 2 vs 1	$p = 0.258$	$p < 0.05$	$p < 0.05$
Orders 2 vs 3	$p < 0.05$	$p < 0.01$	$p < 0.01$

Table (b)

Table A.6: Panel (a) reports the proportion of strict randomization choices ($0 < \lambda < 1$) across treatments in each order. Panel (b) the average proportions of strict randomization at different ranges of sure payments aggregated across treatments and lotteries in each order.

We find that subjects were less likely to randomize strictly ($0 < \lambda < 1$) when randomization probabilities were elicited before self-reported decision measures (Order 2). Table A.6(a) compares the proportion of decisions in which subjects randomized strictly when the randomized choices were made first versus when they were made later across orders and treatments. On average, subjects in Order 2 randomized strictly in fewer choices compared to subjects in the other two orders. This difference is statistically significant in three out of eight comparisons, but it is not significant when we aggregate across all treatments. Table

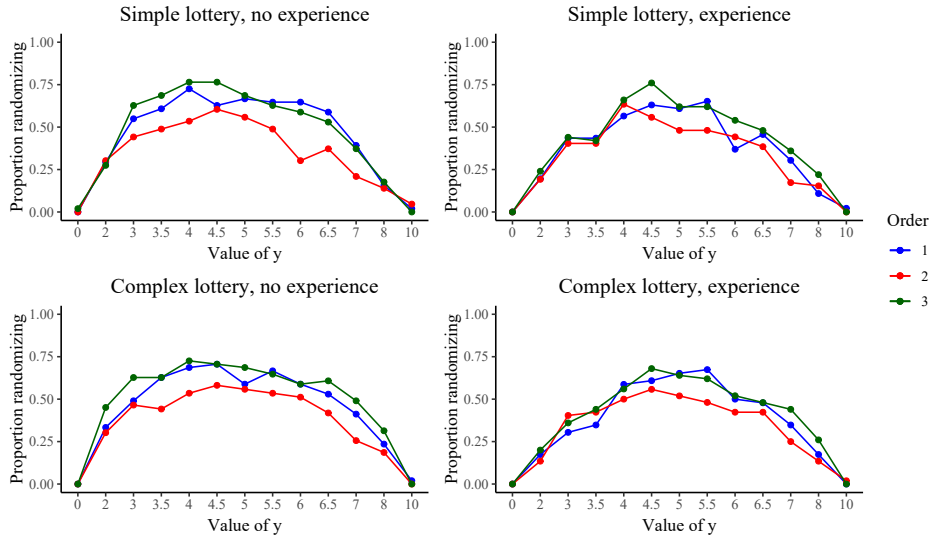


Figure A.8: The proportion of strict randomization choices ($0 < \lambda < 1$) at each value of y in the three orders.

A.6(b) shows that the above order effects were stronger in choices with high values ($y \geq 4$) than with low values ($y \leq 3.5$) of sure payments. Consistent with these findings, Figure A.8 shows that the proportion of subjects with strict randomization probabilities tends to be lower at each value of y in Order 2 (the red lines) than in the other two orders, and more so for high values of sure payments.

Apart from the lower tendency to randomize, randomization probabilities corresponded to the probabilistic confidence less well at some probabilistic confidence values (e.g., 30% x and 40% x) in Order 2 compared to the other two orders. Figure A.10 reports this result. Finally, we find that the cumulative distributions of the correlation between randomization probabilities and a self-reported decision confidence measure in Order 2 tend to be on the left of the other two orders (see Figure A.9). This implies that there were more subjects with lower correlation between randomization probabilities and self-reported decision confidence in Order 2 than in the other two orders. As a whole, the results suggest that randomized choices were affected by priming.

Despite the presence of the priming effects on randomized choices, we find support for our hypotheses when we restrict our analyses to subjects in Order 2. First, we find high correlations between randomization probabilities and self-reported confidence in Order 2, consistent with H1: the median correlation between randomization probabilities and confidence statement as well as the median correlation between randomization probabilities and probabilistic confidence in Order 2 ranges from 0.87 to 0.89 across treatments respectively. Second, subjects in Order 2 reported low decision confidence for choices around the switching range and chose randomization probabilities close to 0.5 in these decisions, supporting

H2. Table A.7 shows that subjects' median randomization probability for all choices that falls within the switching range is between 0.48 and 0.50 across treatments, consistent with their median confidence statement of "Unsure," and their median probabilistic confidence which ranges from 40% x to 60% x across treatments.²⁵ Third, like in the full sample, we find significant treatment effects in Order 2. Table A.8 reports the range of sure payments over which subjects expressed less than full decision confidence in their decisions. The ranges based on self-reported decision measures and randomization choices suggest that subjects in Order 2 had lower decision confidence for decisions involving the complex lottery than for decisions involving the simple lottery, consistent with H3. We find similar support for H4 in Order 2 as in the full sample. Among subjects in Order 2, those in the full-experience treatment reported less than full decision confidence in smaller ranges of sure payments for decisions involving the complex lottery, but not for decisions involving the simple lottery. In terms of randomization probabilities, these ranges are 2.30 vs 2.89 in Order 2. In comparison, they are 2.81 vs 3.09 in Order 1 and 2.90 vs 3.51 in Order 3. The difference is statistically significant in aggregate (2.67 vs 3.19, $p < 0.05$), but not when considering any order separately ($p > 0.10$).

Taken together, priming effects could have strengthened some of our aggregate findings. However, the findings that subjects in Order 2 randomized in ways that were broadly consistent with our hypotheses suggest that randomization probabilities and self-reported decision confidence measures are likely to share common psychological foundations.

It is worth noting that the aforementioned order effects do not automatically imply that randomization probabilities are a poorer proxy for decision confidence than self-reported decision confidence. Confidence statements and probabilistic confidence are also noisy proxies of decision confidence, and there is no objective criterion for the "right" amount of strict randomization. For example, it is possible that subjects may have not randomized too little in Order 2, but that they have randomized excessively in the other two orders due to the priming effects. Since there is no obvious benchmark to compare decision confidence measures, and decision confidence is not directly observable, we consider the value of decision confidence on the basis of its correspondence with actual choices. When decision confidence corresponds perfectly with choices, having decision confidence of $p\%$ for x would imply that x is chosen $p\%$ of time. Figure A.11 shows, across subjects and choices, the proportion of choices in which x was chosen based on subjects' binary choices by randomization probabilities as well as the probabilistic confidence in the three orders.²⁶ On average, both measures of decision confidence closely trace the proportions of x chosen in binary choices. Importantly, randomization probabilities exhibit a closer

²⁵The median randomization probability at the upper bound ($\bar{y}_b$) of the switching range is lower in Order 2 compared to the other two orders in most treatments. This is consistent with our earlier finding of the order effects that there were fewer strict randomization choices in Order 2, and more so for choices involving larger sure payments.

²⁶We did not consider confidence statements here, because they were elicited on the same decision screen as binary choices.

Confidence statements						
No experience						
Order	Simple			Complex		
	y_b	$\bar{y}_b$	$[y_b, \bar{y}_b]$	y_b	$\bar{y}_b$	$[y_b, \bar{y}_b]$
Order 1	Probably x	Probably y	Unsure	Probably x	Probably y	Unsure
Order 2	Probably x	Probably y	Unsure	Probably x	Probably y	Unsure
Order 3	Probably x	Probably y	Unsure	Probably x	Probably y	Unsure
Experience						
Order	Simple			Complex		
	y_b	$\bar{y}_b$	$[y_b, \bar{y}_b]$	y_b	$\bar{y}_b$	$[y_b, \bar{y}_b]$
Order 1	Probably x	Probably y	Unsure	Probably x	Probably y	Unsure
Order 2	Probably x	Probably y	Unsure	Probably x	Probably y	Unsure
Order 3	Probably x	Probably y	Unsure	Probably x	Probably y	Unsure
Probabilistic confidence						
No experience						
Order	Simple			Complex		
	y_b	$\bar{y}_b$	$[y_b, \bar{y}_b]$	y_b	$\bar{y}_b$	$[y_b, \bar{y}_b]$
Order 1	60% x	40% x	40% x	60% x	40% x	60% x
Order 2	70% x	40% x	60% x	60% x	40% x	40% x
Order 3	60% x	40% x	60% x	60% x	30% x	40% x
Experience						
Order	Simple			Complex		
	y_b	$\bar{y}_b$	$[y_b, \bar{y}_b]$	y_b	$\bar{y}_b$	$[y_b, \bar{y}_b]$
Order 1	60% x	40% x	50% x	60% x	40% x	40% x
Order 2	70% x	40% x	60% x	60% x	40% x	40% x
Order 3	60% x	40% x	60% x	60% x	40% x	60% x
Randomization probabilities						
No experience						
Order	Simple			Complex		
	y_b	$\bar{y}_b$	$[y_b, \bar{y}_b]$	y_b	$\bar{y}_b$	$[y_b, \bar{y}_b]$
Order 1	0.65	0.46	0.50	0.68	0.40	0.50
Order 2	0.68	0.27	0.50	0.63	0.30	0.48
Order 3	0.60	0.46	0.50	0.58	0.21	0.40
Experience						
Order	Simple			Complex		
	y_b	$\bar{y}_b$	$[y_b, \bar{y}_b]$	y_b	$\bar{y}_b$	$[y_b, \bar{y}_b]$
Order 1	0.68	0.30	0.55	0.60	0.33	0.50
Order 2	0.63	0.23	0.50	0.66	0.25	0.49
Order 3	0.70	0.43	0.51	0.61	0.40	0.50

Table A.7: Median behavior around the switching choices (y_b and $\bar{y}_b$) and within the switching range ($[y_b, \bar{y}_b]$) aggregated across lotteries and treatments for each order.

correspondence to binary choices than probabilistic confidence in all three orders. This alignment is particularly evident in Order 2, with randomization probabilities exhibiting a significantly closer match to these proportions than probabilistic confidence at nine levels of randomization probabilities/probabilistic confidence versus five levels in Order 1 and six levels in Order 3.

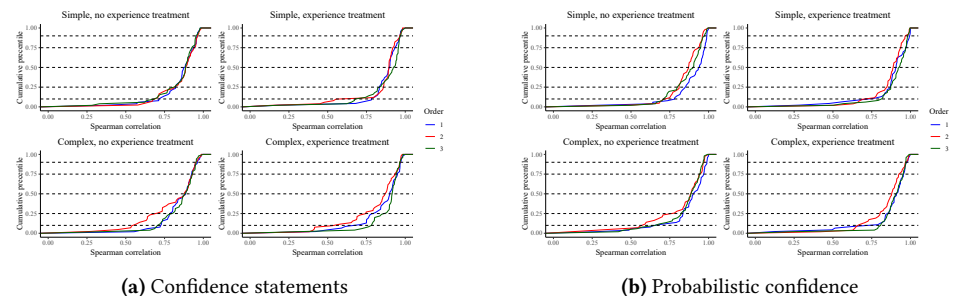


Figure A.9: ECDF for the correlation with randomization probabilities in each order across treatments.

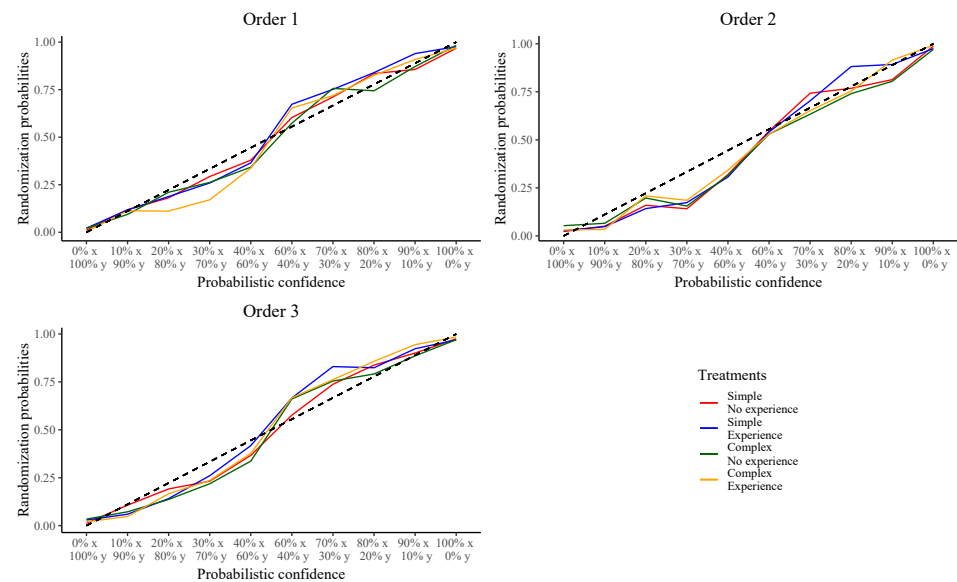


Figure A.10: The mean randomization probabilities at each self-reported probabilistic confidence level in Experiment 2 for each lottery and experience treatment in each order separately.

	Lottery	Treatment	Combined	Order 1	Order 2	Order 3
Confidence Statements	Simple	No experience	3.15	2.96	2.77	3.68
		Experience	3.04	2.84	2.92	3.35
	Complex	No experience	3.58***	3.25*	3.56***	3.93
		Experience	3.16**	3.23	3.03*	3.23*
Probabilistic Confidence	Simple	No experience	4.37	4.28	4.06	4.72
		Experience	4.38	4.24	4.40	4.50
	Complex	No experience	4.63***	4.34	4.61***	4.93*
		Experience	4.58	4.53	4.50	4.70
Randomization Probabilities	Simple	No experience	3.03	3.17	2.67	3.14
		Experience	2.74	3.00	2.34	2.93
	Complex	No experience	3.19*	3.09	2.89*	3.51**
		Experience	2.67**	2.81	2.30	2.90

Table A.8: The mean size of the range of sure payments over which subjects express that they are not fully confident about their decision based on each of the confidence measures, by the lottery and experience treatments in aggregate and in each order separately (randomization probabilities ($0.10 \leq \lambda \leq 0.90$), confidence statements (“Probably x ”, “Unsure”, “Probably y ”) and probabilistic confidence (between “90% x , 10% y ” and “10% x , 90% y)). Stars in the upper right corners of a cell denote statistical significance of Wilcoxon signed-rank tests between the simple lottery and the complex lottery, while stars in the lower right corner denote statistical significance of Wilcoxon rank-sum tests between the no experience and experience treatment: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

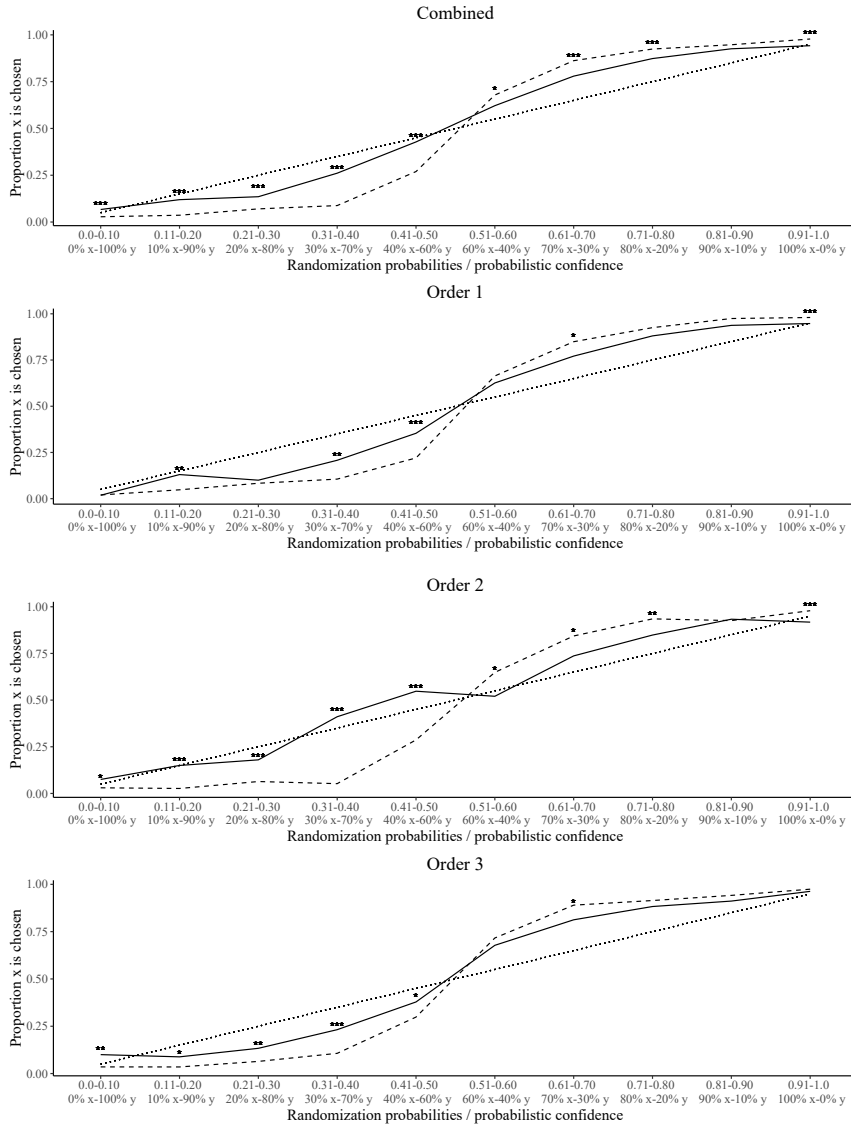


Figure A.11: Correspondence of randomization probabilities (solid line) and probabilistic confidence (dashed line) with the proportion of binary choices in which x is chosen (y -axis) across lotteries and treatments for each decision order. The dotted line represents a 45-degree line. Fisher's exact tests were performed to test the difference in choice proportions between the randomization probabilities and probabilistic confidence: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Randomization Interval	The number of subjects who chose randomization			
	0 times	1 time	2 times or more	3 times or more
Experiment 1: Simple lottery, no-experience				
$0 < \lambda < 1$	2	6	97	95
$0.10 \leq \lambda \leq 0.90$	3	6	96	93
$0.40 \leq \lambda \leq 0.60$	22	26	57	35
Experiment 1: Complex lottery, no-experience				
$0 < \lambda < 1$	4	1	100	98
$0.10 \leq \lambda \leq 0.90$	4	2	99	98
$0.40 \leq \lambda \leq 0.60$	21	15	69	46
Experiment 1: Simple lottery, partial-experience				
$0 < \lambda < 1$	6	1	93	89
$0.10 \leq \lambda \leq 0.90$	6	1	93	88
$0.40 \leq \lambda \leq 0.60$	13	24	63	38
Experiment 1: Complex lottery, partial-experience				
$0 < \lambda < 1$	3	5	92	89
$0.10 \leq \lambda \leq 0.90$	3	7	90	89
$0.40 \leq \lambda \leq 0.60$	14	19	67	46
Experiment 2: Simple lottery, no-experience				
$0 < \lambda < 1$	25	8	112	106
$0.10 \leq \lambda \leq 0.90$	26	7	112	105
$0.40 \leq \lambda \leq 0.60$	42	24	79	44
Experiment 2: Complex lottery, no-experience				
$0 < \lambda < 1$	26	6	113	100
$0.10 \leq \lambda \leq 0.90$	26	7	112	98
$0.40 \leq \lambda \leq 0.60$	37	38	70	42
Experiment 2: Simple lottery, full-experience				
$0 < \lambda < 1$	32	11	105	98
$0.10 \leq \lambda \leq 0.90$	34	11	103	96
$0.40 \leq \lambda \leq 0.60$	55	36	57	32
Experiment 2: Complex lottery, full-experience				
$0 < \lambda < 1$	35	11	102	91
$0.10 \leq \lambda \leq 0.90$	35	12	101	90
$0.40 \leq \lambda \leq 0.60$	56	25	67	38

Table A.9: The distribution of subjects who chose $0 < \lambda < 1$, $0.10 \leq \lambda \leq 0.90$, and $0.40 \leq \lambda < 0.60$ zero times, one time, two times or more, and three times or more across treatments in the two experiments.

A.2.2 Alternative interpretations of randomization probabilities

In the main text, we have shown that indifference and random errors cannot be the driving force behind subjects' randomization behavior. We elaborate why utility difference alone cannot explain randomization here.

Butler et al. (2014) call utility difference the strength of preferences: “the relative degree of difference between the two options as perceived by the decision maker” (Butler et al., 2014, p.538). For example, we can write this explicitly as a Fechnerian utility model $p = \phi[U(L) - u(y)]$, where $\phi : R \Rightarrow [0, 1]$ is a cumulative distribution function with $\phi(0) = 0.5$

(Luce and Suppes, 1965, p.334). The lower bound and the upper bound of randomization are then $u(\underline{y}) = U(L) - \phi^{-1}(0.90)$ and $u(\bar{y}) = U(L) - \phi^{-1}(0.10)$. If randomization probabilities depend only on utility differences, when $U(L)$ increases: (1) the randomization probability should increase for each value of sure payment; and (2) the lower bound and the upper bound of randomization should increase equally. Our results clearly reject these two predictions. Subjects' randomization probabilities did not shift horizontally but were instead compressed towards 0.5 when they faced the complex lottery compared to the simple lottery, and were stretched away from 0.5 in the full-experience treatment compared to the no-experience treatment. In addition, Table A.3 shows that while $\bar{y}$ for the complex lottery was significantly higher than $\bar{y}$ for the simple lottery in both experiments (Experiment 1: 6.90 vs 6.61, Wilcoxon signed-rank test, $p < 0.01$; Experiment 2: 6.38 vs 6.18, Wilcoxon signed-rank test $p < 0.01$), $\underline{y}$ for the complex lottery was significantly lower than $\underline{y}$ for the simple lottery in Experiment 1 at 10% significance level (2.84 vs 2.99, Wilcoxon signed-rank test $p < 0.10$), but was not significantly different in Experiment 2 (3.19 vs 3.16, Wilcoxon signed-rank test $p = 0.646$).

Importantly, although the difference in the mean valuation of the lottery with or without experience was similar to that of the complex lottery versus the simple lottery, Table A.5 shows that $\underline{y}$ were significantly lower for subjects in the full-experience treatment than in the no experience treatment (Simple lottery: 3.16 vs 3.44, Wilcoxon rank-sum test $p < 0.05$; Complex lottery: 3.19 vs 3.60, Wilcoxon rank-sum test $p < 0.01$) but not $\bar{y}$ (Simple lottery: 6.18 vs 6.18, Wilcoxon rank-sum test $p = 0.789$; Complex lottery: 6.27 vs 6.38, Wilcoxon rank-sum test $p = 0.369$). Further, randomization probabilities were larger at low sure payments but smaller at high sure payments when we compare the full-experience treatment with the no-experience treatment. These results highlight the central role of preference uncertainty beyond utility difference in affecting decision confidence.

One may perceive that self-reported decision confidence measures are easier to interpret than randomization probabilities because they ask about decision confidence explicitly. We show that self-reported decision confidence measures can be just as difficult to interpret by analyzing how subjects associate the two self-reported decision measures in the post-experiment questionnaire of Experiment 2. We asked subjects which confidence statement best described their probabilistic confidence $p\%$ in choosing x and $100-p\%$ in choosing y for values $p = 60, 70, 80, 90, 100$. In a separate session, we asked subjects to state the minimum level of probabilistic confidence for "Surely", and the minimum and maximum levels of probabilistic confidence for "Probably" and "Unsure" on a scale from 0% to 100%.

Table A.10 summarizes the subjects' responses to the two questions. The top panel shows the range of probabilistic confidence levels associated with each confidence statement. Although the first column shows that the median probabilistic confidence thresholds are well-ordered (the median maximum probabilistic confidence of a lower ordered statement was always smaller than the median minimum probabilistic confidence of a higher ordered statement), the standard deviations reported in the last column as well as minimum and maximum probabilistic confidence assigned to each confidence statement at different per-

Probabilistic confidence associated with each confidence statement						
Levels	Median	10th	30th	70th	90th	SD
Surely (min)	85%	70%	80%	90%	100%	16.31%
Probably (max)	80%	70%	80%	90%	99%	11.10%
Probably (min)	55%	25%	50%	60%	65%	16.59%
Unsure (max)	54%	25%	50%	60%	64%	17.58%
Unsure (min)	35%	0%	0%	40%	50%	21.12%

Confidence statements associated with each probabilistic confidence level						
Levels	Median	10th	30th	70th	90th	SD
100%	Surely <i>x</i>	Surely <i>x</i>	Surely <i>x</i>	Surely <i>x</i>	Surely <i>x</i>	0.17
90%	Surely <i>x</i>	Probably <i>x</i>	Surely <i>x</i>	Surely <i>x</i>	Surely <i>x</i>	0.54
80%	Probably <i>x</i>	Probably <i>x</i>	Probably <i>x</i>	Surely <i>x</i>	Surely <i>x</i>	0.54
70%	Probably <i>x</i>	Probably <i>x</i>	Probably <i>x</i>	Probably <i>x</i>	Probably <i>x</i>	0.34
60%	Unsure	Unsure	Unsure	Surely <i>x</i>	Surely <i>x</i>	0.54

Table A.10: The median, 10th, 30th, 70th, 90th percentile, and standard deviation of probabilistic confidence associated with each confidence statement and the median, 10th, 30th, 70th, 90th percentile, and standard deviation of confidence statements associated with each probabilistic confidence level. Consistent with Result 1, we code confidence statements of surely *x*, probably *x*, unsure, probably *y*, and surely *y* as 5, 4, 3, 2, and 1 respectively. Standard deviations are calculated accordingly.

centile levels show the presence of substantial heterogeneity in the probabilistic confidence associated with each confidence statement.

Further, we find two different interpretations of the confidence statement “Unsure”. A large group of subjects ($n=172$) reported probabilistic confidence higher than 50% as the maximum of “Unsure” and higher than 0% as the minimum of “Unsure,” while another group of subjects ($n=84$) reported a probabilistic confidence level lower than 50% as the maximum of “Unsure” and close 0% as the minimum of “Unsure.” Table A.11 shows how different these two groups were in their associations of probabilistic confidence and other confidence statements. For example, the maximum level of probabilistic confidence for the statement “Probably” ranges from 75% to 99% among subjects who reported a probabilistic confidence level higher than 50% as the maximum of “Unsure,” and it ranges from 60% to 95% among subjects who reported a probabilistic confidence level lower than 50% as the maximum of “Unsure.”

Probabilistic confidence associated with each confidence statement						
For subjects: Unsure (max) $\geq 50\%$ and Unsure(min) $> 0\%$						
Levels	Median	10th	30th	70th	90th	S.D
Surely (min)	85%	75%	80%	90%	100%	14.80%
Probably (max)	85%	75%	80%	90%	99%	9.13%
Probably (min)	60%	41%	55%	60%	70%	13.26%
Unsure (max)	60%	50%	55%	60%	65%	9.62%
Unsure (min)	40%	30%	40%	45%	50%	10.46%
For subjects: Unsure (max) $< 50\%$						
Levels	Median	10th	30th	70th	90th	S.D
Surely (min)	80%	50.3%	75%	85.5%	99%	20.22%
Probably (max)	80%	60%	75%	84%	95%	14.01%
Probably (min)	40%	20%	30%	50%	60%	17.57%
Unsure (max)	30%	10%	20%	35.4%	40%	12.94%
Unsure (min)	0%	0%	0%	0%	10%	8.46%

Table A.11: The median, 10th, 30th, 70th, 90th percentile, and standard deviation of probabilistic confidence associated with each confidence statement for subjects who fit the criteria specified in the table.

A.2.3 Reasons to randomize

At the end of the session on randomized choices in Experiment 2, we asked the subjects who had chosen to randomize at least once in the post-experiment questionnaire, what their reasons for randomizing were. Of the 120 subjects who provided an answer to this question, 22% stated that they randomized because they were unsure about their choice or found it difficult to compare the two options. Here are a few examples:

- “Because I was not completely sure whether I wanted to choose A or B.”
- “I was not sure exactly what the consequences of my decision was going to be and I was not 100% confident in choosing either A or B.”
- “Its difficult to make a decision for sure, so a combination feels more safe.”

Another group of subjects (22.5%) randomized for reasons related to hedging. Here are a few examples:

- “Even though the certain option was less valued, certainty is nice and preferred over risky options. Therefore, I chose to combine them some of the time.”
- “To hedge my bets when the expected gains of A and B were similar, gaining a small chance for big gains or loses in option A, adding some suspense.”
- “For example when I preferred A but B felt a little safer so I thought it wouldn’t hurt adding a bit more security since a B amount for sure isn’t bad.”

Around 18% stated that they chose to randomize when the sure payment amount was close to the expected value of the lottery but did not explain why randomizing is better. In contrast, most of the subjects who did not to randomize at all stated that they did not randomize because they did not want to pay the cost of 0.10 euro for randomizing and/or that they made their choices solely based on the computation of the expected value of the lottery.

A.2.4 Results for the loss lottery and the mixed lottery in Experiment 1

Treatment	Lottery	Surely x	Probably x	Unsure	Probably y	Surely y		
No-experience	Loss	Mean	0.94 (0.008)	0.73 (0.015)	0.50 (0.021)	0.29 (0.019)	0.07 (0.011)	
		Min	0.84 (0.022)	0.61 (0.023)	0.40 (0.025)	0.19 (0.019)	0 (0.002)	
		Max	1 (0.003)	0.85 (0.016)	0.60 (0.025)	0.41 (0.026)	0.15 (0.024)	
	Mixed	Mean	0.90 (0.015)	0.71 (0.015)	0.53 (0.025)	0.32 (0.020)	0.10 (0.016)	
		Min	0.75 (0.035)	0.56 (0.025)	0.44 (0.029)	0.17 (0.018)	0.02 (0.011)	
		Max	0.99 (0.003)	0.85 (0.016)	0.61 (0.030)	0.46 (0.031)	0.23 (0.031)	
	Partial-experience	Loss	Mean	0.91 (0.014)	0.70 (0.018)	0.55 (0.019)	0.34 (0.019)	0.07 (0.011)
			Min	0.81 (0.026)	0.55 (0.026)	0.42 (0.027)	0.23 (0.021)	0.01 (0.006)
			Max	0.98 (0.012)	0.82 (0.016)	0.66 (0.022)	0.46 (0.025)	0.14 (0.022)
Mixed		Mean	0.91 (0.014)	0.71 (0.019)	0.57 (0.025)	0.33 (0.023)	0.11 (0.014)	
		Min	0.78 (0.031)	0.57 (0.027)	0.47 (0.032)	0.23 (0.026)	0.01 (0.006)	
		Max	0.99 (0.006)	0.83 (0.018)	0.67 (0.026)	0.46 (0.030)	0.28 (0.035)	

Table A.12: The mean, minimum, and maximum randomization probabilities that correspond to each confidence statement for the loss and mixed lottery in both treatments in Experiment 1. The values in parentheses are the standard errors of the mean.

Treatment		Correlation between randomization probabilities and confidence statements	
		The loss lottery	The mixed lottery
No experience	10th percentile	0.67	0.35
	25th percentile	0.81	0.64
	median	0.90	0.85
	75th percentile	0.95	0.91
	90th percentile	0.97	0.96
Experience	10th percentile	0.65	0.47
	25th percentile	0.80	0.66
	median	0.87	0.81
	75th percentile	0.94	0.92
	90th percentile	0.97	0.95

Table A.13: Nonparametric Spearman correlation at the 10th percentile, 25th percentile, median, 75th percentile, and 90th percentile for the loss lottery and mixed lottery in both treatments in Experiment 1.

A.3 Experimental materials

Experiment 1

Welcome

You are invited to participate in an experiment in which we examine how individuals make decisions. Your decisions in the experiment are about choices between different options. There are no right or wrong answers. The whole experiment will take approximately 20 minutes.

You will receive a participation fee of €1 for completing the survey. In addition you will receive monetary compensation up to €10 based on the decisions you make in the experiment. Specifically, one of the questions will be randomly selected. The decision you made in this question will determine your additional compensation.

You will receive the payment (the participation fee of €1 and the additional compensation) via bank transfer. For this we will ask your IBAN number. This information will only be used for payment, and will be permanently deleted afterwards.

Thank you for your participation!

Sincerely yours,

Associate Professor Dr. Jianying Qiu and PhD student Sara Arts
The Institute of Management Research
Radboud University Nijmegen.

(a)

Voluntary participation

Your participation in this research is voluntary. This means that you can withdraw your participation and consent at any time during the survey, without giving a reason. All data we have collected from you will be deleted permanently. If you desire to withdraw, please simply close your internet browser. After completion of the survey it will not be possible to withdraw your data from the research.

What will happen to the data?

The research data we collect during this study will be used by scientists as part of data sets, articles and presentations. The anonymized research data is accessible to other scientists for a period of at least 10 years. When we share data with other researchers, these data cannot be traced back to you.

More information?

Should you want more information on this research study, please contact Sara Arts (email: s.arts@fm.ru.nl)

CONSENT:

Please select your choice below.

Checking "Agree" below indicates that:
you voluntarily agree to participate.

☒ I agree with the provided information, and I would like to proceed to the survey

☐ I do not agree with the above.

(b)

Figure A.12: Welcome screen (a) and informed consent (b) of the experiment.



Figure A.13: The introduction of the binary choices and confidence statements for the complex lottery in the no-experience treatment (a) and the simple lottery in the partial-experience treatment (b).



Figure A.14: Examples of the decision screens for the binary choices and confidence statements for the complex lottery in the no-experience treatment (a) and the simple lottery in the partial-experience treatment (b).

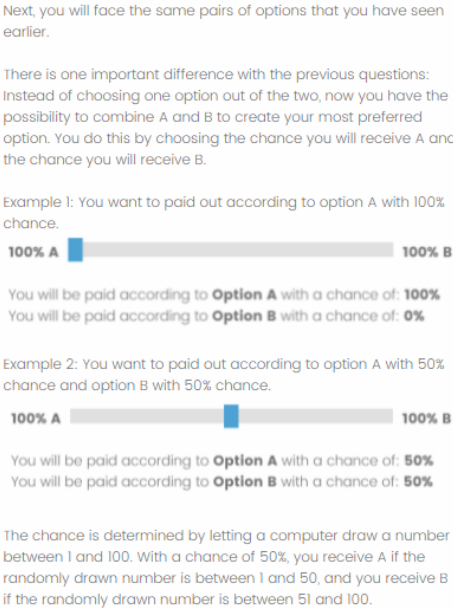


Figure A.15: Explanation of the randomized choices.



Figure A.16: Examples of the decision screens for the randomized choices for the simple lottery in the no-experience treatment (a) and the complex lottery in the partial-experience treatment (b).

The experiment is almost finished. We would like to ask you some final, general questions.

What is your gender?

☐ Male

☐ Female

☐ Other

☐ Prefer not to say

What is your current age?

What is your current field of study? (Select the category that fits best)

☐ Natural sciences

☐ Social sciences

☐ Management sciences

☐ Humanities

☐ Other

Do you have any comments?

Figure A.17: Demographic questions asked at the end of the experiment.

Experiment 2

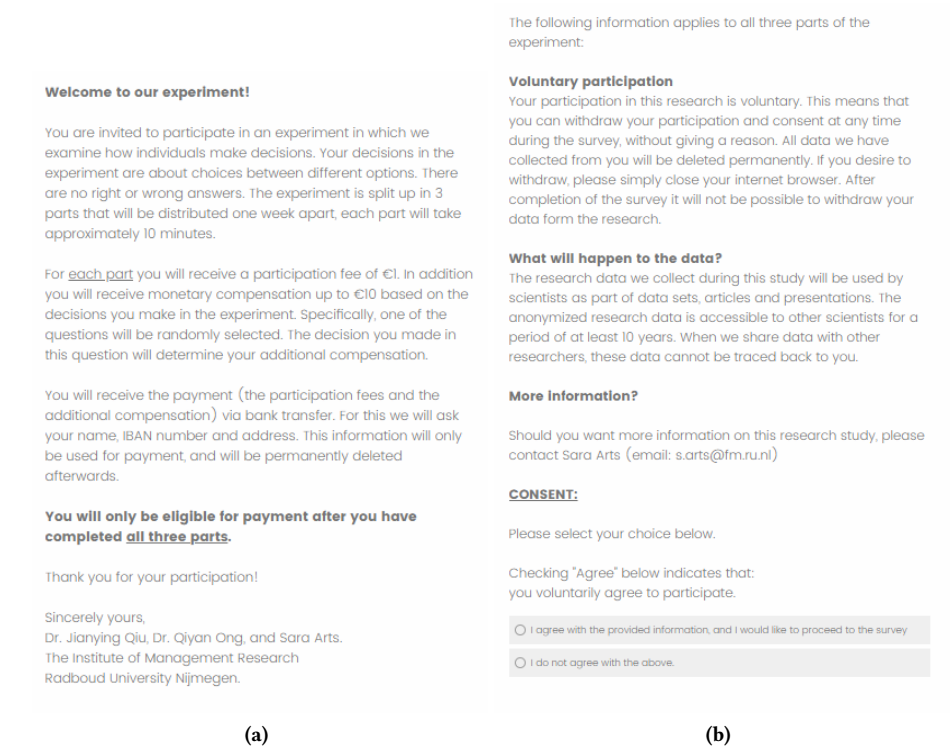


Figure A.18: Welcome screen (a) and informed consent (b) of the experiment.

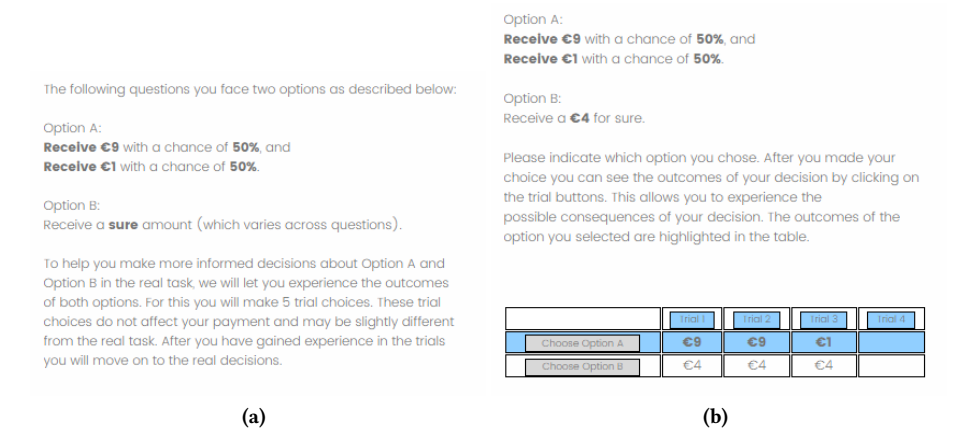


Figure A.19: The introduction (a) and an example of the hypothetical decision screens (b) of the full-experience treatment.

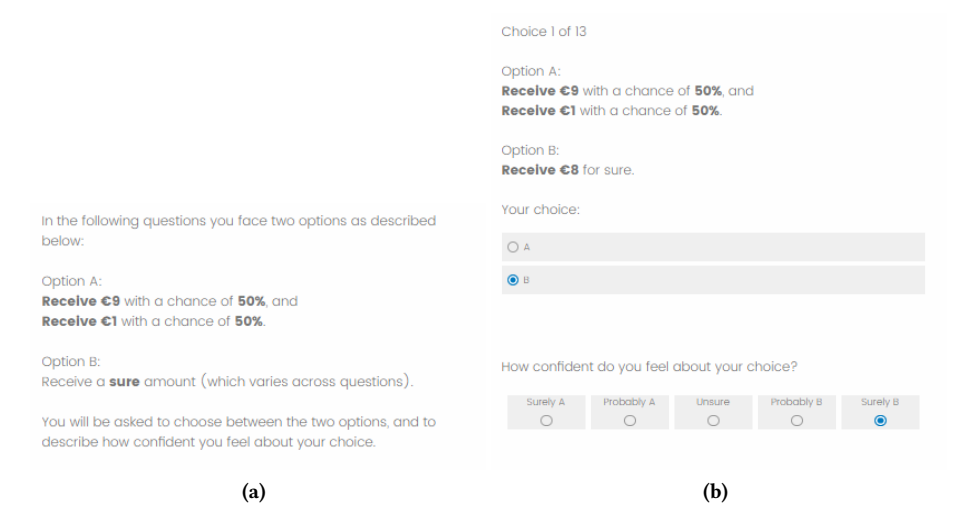


Figure A.20: The introduction (a) and an example of the decision screens (b) of binary choices and confidence statements for the simple lottery.

Please tell us how you understand the confidence statements used in this part of the experiment.

If you would have to assign a minimum confidence level to the statement '**Surely A**' or '**Surely B**', what would it be? (From 0% to 100% confident).

If you would have to assign a range of confidence levels to the statement '**Probably A**' or '**Probably B**', what would it be?

Minimum confidence level (From 0% to 100% confident):

Maximum confidence level (From 0% to 100% confident):

If you would have to assign a range of confidence levels to the statement '**Unsure**', what would it be?

Minimum confidence level (From 0% to 100% confident):

Maximum confidence level (From 0% to 100% confident):

Figure A.21: post-experiment questionnaire after the binary choices and confidence statements.

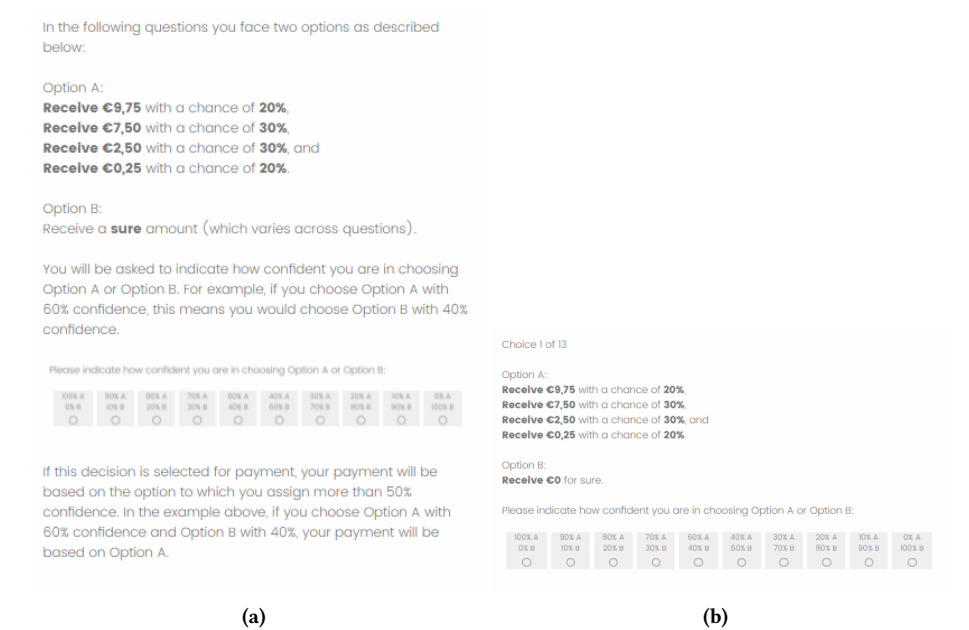


Figure A.22: The introduction (a) and an example of the decision screens (b) of probabilistic confidence choices for the complex lottery.

Which statement best describes your association with 70% confidence in choosing A and 30% confidence in choosing B:

- ☐ Surely A
- ☐ Probably A
- ☐ Unsure
- ☐ Probably B
- ☐ Surely B

Which statement best describes your association with 100% confidence in choosing A and 0% confidence in choosing B:

- ☐ Surely A
- ☐ Probably A
- ☐ Unsure
- ☐ Probably B
- ☐ Surely B

Which statement best describes your association with 90% confidence in choosing A and 10% confidence in choosing B:

- ☐ Surely A
- ☐ Probably A
- ☐ Unsure
- ☐ Probably B
- ☐ Surely B

Which statement best describes your association with 80% confidence in choosing A and 20% confidence in choosing B:

- ☐ Surely A
- ☐ Probably A
- ☐ Unsure
- ☐ Probably B
- ☐ Surely B

Which statement best describes your association with 60% confidence in choosing A and 40% confidence in choosing B:

- ☐ Surely A
- ☐ Probably A
- ☐ Unsure
- ☐ Probably B
- ☐ Surely B

Figure A.23: post-experiment questionnaire after the probabilistic confidence choices.

In the following questions you face two options as described below:

Option A:
Receive €9,75 with a chance of **20%**,
Receive €7,50 with a chance of **30%**,
Receive €2,50 with a chance of **30%**, and
Receive €0,25 with a chance of **20%**.

Option B:
Receive a **sure** amount (which varies across questions).

You can choose Option A (100% A), Option B (100% B), or **pay €0,10 and combine A and B** to create your most preferred option. You do this by choosing the chance you will receive A and the chance you will receive B.

Example 1: You want to paid out according to option A with 100% chance.



You will be paid according to **Option A** with a chance of: **100%**
You will be paid according to **Option B** with a chance of: **0%**

Example 2: You want to receive Option A with 50% chance and Option B with 50% chance.



You will be paid according to **Option A** with a chance of: **50%**
You will be paid according to **Option B** with a chance of: **50%**

The chance is determined by letting a computer draw a number between 1 and 100. In example 2, you will be paid according to Option A if the randomly drawn number is between 1 and 50, and you will be paid according to Option B if the randomly drawn number is between 51 and 100.

(a)

Choice 1 of 13

Option A:
Receive €9,75 with a chance of **20%**,
Receive €7,50 with a chance of **30%**,
Receive €2,50 with a chance of **30%**, and
Receive €0,25 with a chance of **20%**.

Option B:
Receive €4 for sure.

You can choose Option A (100% A), Option B (100% B), or **pay €0,10 and combine Option A and B** to create your most preferred option.

To make your choice, please click on the bar below and move the slider to determine the chance according to which you want to receive option A and option B.



You will be paid according to **Option A** with a chance of: **75%**
You will be paid according to **Option B** with a chance of: **25%**

Note: You can combine Option A and Option B to create your most preferred option. You do this by choosing the chance you will receive A and the chance you will receive B. For example, if you choose 100% A, your payment depends only on Option A. If you choose 50% A and 50% B, your payment depends on A with 50% chance, and your payment depends on B with 50% chance. The chance is determined by letting a computer draw a number between 1 and 100. With a chance of 50%, you will be paid according to Option A if the randomly drawn number is between 1 and 50, and you will be paid according to Option B if the randomly drawn number is between 51 and 100.

(b)

Figure A.24: The introduction (a) and an example of the decision screens (b) of randomized choices for the complex lottery.

You chose to combine Option A and Option B in one or more of the previous questions. Can you briefly tell us why?

You did not choose to combine Option A and Option B in any of the previous questions. Can you briefly tell us why?

Figure A.25: post-experiment questionnaire after the randomized choices. The first question was asked if a subject chose randomization probabilities other than 0 or 1 in at least 1 choice. The second question was asked if a subject only chose randomization probabilities of 0 or 1.

The experiment is almost finished. We would like to ask you some final, general questions.

What is your gender?

☐ Male

☐ Female

☐ Other

☐ Prefer not to say

What is your current age?

What is your current field of study? (Select the category that fits best)

☐ Natural sciences

☐ Social sciences

☐ Management sciences

☐ Humanities

☐ Other

Do you have any comments?

Figure A.26: Demographic questions asked at the end of the experiment.

Appendices to chapter 3

B.1 Additional tables

	WTA-WTP sample	Neutral frame sample
Sample size	1236	620
Female	47.17	48.06
Age		
16-34	10.76	11.13
35-49	19.74	15.97
50-64	33.25	24.52
65 and above	36.25	48.39
Education level		
Low	24.19	24.35
Medium	33.82	29.35
High	41.75	46.13
Household income		
0-2499	22.25	21.29
2500-3999	24.68	26.45
4000-5999	22.57	24.68
6000 and above	25.16	24.52
Occupation		
Paid work	50.97	42.26
House work	7.36	7.58
Retired	30.26	40.97
Others	11.41	9.19
Household composition		
Partner	66.91	63.06
# of children	0.49	0.37

Table B.1: Sample composition of the LISS panel members who participated in the experiment. The numbers represent the percentage of subjects in each demographic group, except for the number of children, which is the average. Column 1 shows subjects from the within-subject design in which they reported their WTA, WTP, and R-range. Column 2 shows the subject who completed the neutrally framed questions.

	L1	L2	Average	degenerate lottery
WTA	527.54 (243.92)	519.40 (247.26)	523.47 (226.40)	505.08 (20.48)
WTP	415.32 (252.29)	407.70 (253.78)	411.51 (234.84)	496.63 (22.66)
The WTA-WTP gap	112.22*** (275.34)	111.70*** (275.10)	111.96*** (242.99)	8.45*** (27.07)
CE neutral	476.55 (215.96)	465.12 (214.47)	470.84 (199.35)	

Table B.2: Means of WTA, WTP, WTA-WTP gaps, and CE in the neutral frame for L1, L2 and the average across the two lotteries, as well as the means of WTA, WTP, and WTA-WTP gap of the degenerate lottery. Wilcoxon signed-rank tests were performed to test the significance of the WTA-WTP gap. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Dependent variable	WTA _{sure} = 500	WTP _{sure} = 500	WTA-WTP gap _{sure} = 0
Independent variables	Log odds		
Matrix reasoning score	0.26*** (0.04)	0.26*** (0.04)	0.10*** (0.03)
Education	0.24*** (0.04)	0.14*** (0.05)	0.03 (0.04)
Self-reported understanding	0.20*** (0.05)	0.21*** (0.05)	0.16*** (0.04)
Response time	0.04** (0.02)	0.04** (0.02)	-0.02* (0.01)
Intercept	-2.97*** (0.26)	-2.92*** (0.27)	-1.01*** (0.23)
Efron's R^2	0.12	0.10	0.02

Table B.3: Logistic regressions with the optimal response as the dummy dependent variable (1 indicates the optimal response: WTA_{sure} = 500, WTP_{sure} = 500, and WTA-WTP gap_{sure} = 0). The independent variables include subjects' IQ (their score on the matrix reasoning questions), education level (categorical, from 1 to 6), self-reported understanding of the experimental questions (likert scale from 1 to 5), and time spent on the corresponding task (time on the WTA questions, time on the WTP questions, and total time on both the WTA and WTP questions for the degenerate lottery). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Lottery	Upper bound	Lower bound	R-range	Lottery	Randomizing frequency:		
					0	1	2 or more
L1	551.62 (233.61)	449.14 (239.10)	102.48 (172.42)	L1	121.05 (295.71)	67.00 (237.75)	127.19 (265.57)
L2	538.55 (229.99)	433.10 (237.23)	105.45 (174.91)	L2	118.13 (280.48)	83.65 (247.70)	121.76 (284.37)
Average	545.09 (209.22)	441.12 (213.65)	103.96 (149.59)	Randomizing frequency:			
					0	1 or 2	3 or more
				Total in	122.04	79.61	122.07
				L1 and L2	(255.09)	(226.17)	(241.32)

Table B.4: The left panel reports the mean upper bounds, lower bound, and the R-range for L1, L2, and the averages across lotteries. The right panel reports the mean WTA-WTP gap for subjects who chose the randomization option 0 times, 1 time, and 2 times or more for each of the corresponding lottery (and 0, 1 or 2, and 3 times or more, respectively, for the total times randomizing in L1 and L2). Standard deviations are presented in parentheses.

	Lottery	Before correction	Correcting		
			theoretically	for deviations	for deviations different slopes
WTA	L1	527.54 (243.92)	504.70*** (257.55)	500.09*** (217.37)	488.82*** (217.02)
	L2	519.40 (247.26)	491.02*** (274.85)	495.82 (228.26)	490.50*** (228.18)
	Average	523.47 (226.40)	497.86*** (240.35)	497.95*** (201.69)	489.66*** (201.49)
WTP	L1	415.32 (252.29)	425.50 (259.21)	431.97** (226.07)	438.02*** (226.00)
	L2	407.70 (253.78)	417.90 (264.37)	424.30** (227.86)	418.60** (227.80)
	Average	411.51 (234.84)	421.70 (238.00)	428.13** (206.49)	428.31*** (206.49)
WTA-WTP gap	L1	112.22 (275.34)	79.20*** (312.22)	68.12*** (264.27)	50.81*** (264.18)
	L2	111.70 (275.10)	73.12*** (324.00)	71.52*** (264.77)	71.90*** (265.03)
	Average	111.96 (242.99)	76.16*** (279.24)	69.82*** (230.79)	61.35*** (231.06)

Table B.5: WTA, WTP, and WTA-WTP gap before and after each of the three correction methods. Wilcoxon signed-rank tests were performed to test the significance of the difference between the WTA, WTP, and WTA-WTP gap before and after the correction, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	Lottery	Before controlling	After controlling for the	
			R-range	corrected R-range
WTA	L1	527.54	519.84***	523.27***
		(243.92)	(243.58)	(243.63)
	L2	519.40	514.30***	517.64***
		(247.26)	(247.11)	(247.21)
	Average	523.47	514.92***	519.13***
		(226.40)	(226.07)	(226.20)
WTP	L1	415.32	421.55***	421.84***
		(252.29)	(252.07)	(251.64)
	L2	407.70	413.37***	413.54***
		(253.78)	(253.60)	(253.33)
	Average	411.51	419.62***	420.44***
		(234.84)	(234.55)	(233.99)
The WTA-WTP gap	L1	112.22	98.29***	101.43***
		(275.34)	(274.34)	(273.71)
	L2	111.70	100.93***	104.10***
		(275.10)	(274.52)	(274.40)
	Average	111.96	95.30***	98.69***
		(242.99)	(241.81)	(241.17)

Table B.6: WTA, WTP, and WTA-WTP gap before and after controlling for the (mechanism complexity corrected) R-range. Wilcoxon signed-rank tests were performed to test the significance of the difference between the WTA, WTP, and WTA-WTP gap before and after controlling for the (corrected) R-range, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	Lottery	Before controlling	After controlling for both mechanism complexity and the R-range			
			Model 1	Model 2	Model 3	Model 4
WTA	L1	527.54	496.95***	486.45***	497.57***	486.62***
		(243.92)	(217.30)	(216.97)	(217.25)	(216.92)
	L2	519.40	491.81***	486.69***	493.75***	488.30***
		(247.26)	(228.16)	(228.09)	(228.19)	(228.12)
	Average	523.47	493.06***	485.40***	494.58***	486.37***
		(226.40)	(201.56)	(201.38)	(201.55)	(201.35)
WTP	L1	415.32	437.00***	442.88***	436.89***	443.26***
		(252.29)	(225.91)	(225.84)	(225.64)	(225.57)
	L2	407.70	425.90***	420.21***	426.52***	420.73***
		(253.78)	(227.85)	(227.78)	(227.79)	(227.72)
	Average	411.51	432.68***	432.60***	433.37***	433.52***
		(234.84)	(206.39)	(206.39)	(206.14)	(206.14)
WTA-WTP gap	L1	112.22	59.95***	43.57***	60.68***	43.36***
		(275.34)	(263.83)	(263.79)	(263.33)	(263.24)
	L2	111.70	62.90***	66.48***	67.23***	67.57***
		(275.10)	(264.57)	(264.83)	(264.49)	(264.74)
	Average	111.96	60.38***	52.80***	61.22***	52.84***
		(242.99)	(230.29)	(230.59)	(229.84)	(230.11)

Table B.7: WTA, WTP, and WTA-WTP gap before and after controlling for both mechanism complexity and the R-range. In Model 1 and 3 mechanism complexity is deviations from the optimal response, while in Model 2 and 4 mechanism complexity is deviations from the optimal response while allowing for different slopes for positive and negative deviations. In Model 1 and 2 the R-range is uncorrected, and in Model 3 and 4 the R-range is mechanism complexity corrected. Wilcoxon signed-rank tests were performed to test the significance of the difference between WTA, WTP, and WTA-WTP gap before the correction and after the correction, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	Mechanism complexity			No mechanism complexity		
	Valuation complexity			Valuation complexity		
	High	Low	No	High	Low	No
Proportions	23.8%	29.5%	22.5%	7.7 %	8.1%	8.4%
WTA-WTP gap	151.62 (285.83)	103.54 (234.38)	95.78 (267.05)	76.22 (170.66)	0.58 (128.25)	122.78 (195.53)
Education	3.84 (1.40)	3.58 (1.45)	3.74 (1.36)	4.27 (1.25)	4.16 (1.51)	4.71 (1.16)
MR performance	3.78 (1.49)	3.22 (1.73)	3.23 (1.78)	4.39 (1.39)	4.33 (1.64)	4.76 (1.26)
Correlation between λ and the WTA-WTP gap	0.155	0.083	-0.044	-0.068	-0.146	0.507***

Table B.8: The distribution of the subjects with/out each form of complexity, their associated average WTA-WTP gap, education, performance in the matrix reasoning questions, and loss aversion (λ), and the correlation between the WTA-WTP gap and λ . The high and low valuation complexity groups are defined by a median split of the R-range (100 points for subjects with mechanism complexity, and 131.25 points for those without). Education level is measured on a scale from 1 to 6. Standard deviations are in parentheses.

	Lottery	Before correction	Controlling λ	Combined (Model 3)	Combined (Model 3)+ λ	Combined (Model 4)	Combined (Model 4)+ λ
WTA	L1	516.82 (237.38)	517.39*** (237.38)	487.47 (211.21)	484.46*** (211.20)	479.94 (211.03)	477.18*** (211.02)
	L2	514.35 (242.82)	512.21*** (242.82)	491.52 (224.41)	486.59*** (224.38)	484.33 (224.27)	479.62*** (224.24)
	Average	515.58 (224.39)	514.80*** (224.39)	485.04 (200.34)	481.02*** (200.32)	477.98 (200.18)	474.18*** (200.16)
WTP	L1	416.02 (243.59)	444.56*** (243.20)	430.52 (216.88)	439.15*** (216.84)	435.62 (216.82)	445.16*** (216.78)
	L2	406.21 (248.39)	434.57*** (248.02)	425.84 (223.66)	433.87*** (223.63)	421.47 (223.62)	429.21*** (223.59)
	Average	411.12 (230.83)	439.56*** (230.42)	428.52 (203.32)	437.04*** (203.28)	428.82 (203.32)	437.66*** (203.28)
WTA -WTP gap	L1	100.80 (269.14)	72.83*** (268.89)	56.95 (256.87)	45.31*** (256.80)	44.32 (257.09)	32.01*** (257.00)
	L2	108.14 (274.26)	77.65*** (273.70)	65.67 (263.20)	52.72*** (263.05)	62.86 (263.52)	50.40*** (263.36)
	Average	104.47 (244.07)	75.24*** (243.62)	56.52 (230.68)	43.97*** (230.56)	49.16 (231.03)	36.52*** (230.89)

Table B.9: WTA, WTP, and WTA-WTP gap before and after controlling for both mechanism complexity and the R-range as well as the loss aversion coefficient λ . In Controlling for λ we consider only the loss aversion λ in the regression. Combined models (Model 2 and Model 4) are those defined in Table B.7. In Combined (Model 2 or 4) + λ , we include additionally λ . Wilcoxon signed-rank tests were performed to test the significance of the difference between the WTA, WTP, and WTA-WTP gap before and after the inclusion of λ , * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B.2 Mechanism complexity and valuation complexity correction with dummy variables for each deviation

In subsection 3.3.3 we describe three mechanism complexity correction methods. Here, we discuss one additional method based on the same statistical correction model 1. For this correction, we use dummy variables for each deviation of the degenerate lottery's WTA or WTP from 500 to correct for the WTA, WTP, and WTA-WTP gap of the lotteries. Specifically, in statistical model 1, $\mathbf{MC}_a = [537.5, 513, 487, 462.5]^T$ denotes a vector of dummy variables for each possible value of the WTA_{sure} other than the correct response of 500 points, and $\mathbf{MC}_p = [537.5, 513, 487, 462.5]^T$ denotes a vector of dummy variables for each possible value of the WTP_{sure} other than the correct response of 500 points. This correction relaxes the assumption that a higher deviation from 500 for the degenerate lottery corresponds to a higher mechanism complexity. In Table B.10 we present the results of the correction for mechanism complexity only and the joint correction of mechanism complexity and valuation complexity. When controlling for valuation complexity we look at both the uncorrected R-range and the R-range corrected for mechanism complexity. In the latter case, we correct the R-range using dummy variables for each deviation as described above.

Consistent with the corrections reported in the main paper, we observe a significantly smaller WTA-WTP gap after correcting for mechanism complexity with dummy variables for each deviation (64.72 vs 111.96, $p < 0.01$). This is a reduction of about 40%, comparable to the two statistical correction methods we reported in the main text. Also consistent with these earlier corrections, we see that WTA is significantly lower after correction (475.98 vs. 523.47, $p < 0.01$). WTP gives a less straightforward result, as we observe contrasting results for the two lotteries: for L1 we find a higher but insignificant WTP after correction (425.76 vs 415.32, $p > 0.10$), and for L2 we find a significantly lower WTP after correction (396.77 vs 407.70, $p < 0.01$). On average this results in a statistically significant, but only very slightly lower WTP after correction for mechanism complexity with dummy variables (411.27 vs 411.51, $p < 0.01$).

Additionally, including the R-range gives similar results as for the earlier analyses. We find a significant reduction when considering mechanism complexity together with both the uncorrected R-range (56.07 vs 111.96, $p < 0.01$) and the mechanism complexity corrected R-range (55.01 vs 111.96, $p < 0.01$), which corresponds to a reduction of about 50%.

	Lottery	Before correction	Correcting for each deviation		
			mechanism complexity only	with uncorrected R-range	with corrected R-range
WTA	L1	527.54	478.27***	474.56***	474.45***
		(243.92)	(214.06)	(213.95)	(213.88)
	L2	519.40	473.70***	469.42***	470.10***
		(247.26)	(224.31)	(224.19)	(224.19)
	Average	523.47	475.98***	470.46***	470.43***
		(226.40)	(197.74)	(197.57)	(197.49)
WTP	L1	415.32	425.76	429.99**	431.35***
		(252.29)	(222.41)	(222.28)	(222.11)
	L2	407.70	396.77***	397.04	396.79
		(253.78)	(220.69)	(220.69)	(220.69)
	Average	411.51	411.27***	414.39***	415.42***
		(234.84)	(200.78)	(200.72)	(200.65)
WTA-WTP gap	L1	112.22	52.51***	44.56***	43.10***
		(275.34)	(264.23)	(263.73)	(263.29)
	L2	111.70	76.92***	72.38***	73.30***
		(275.10)	(264.12)	(263.99)	(263.02)
	Average	111.96	64.72***	56.07***	55.01***
		(242.99)	(230.66)	(230.18)	(229.90)

Table B.10: WTA, WTP, and WTA-WTP gap before and after correcting for mechanism complexity by including each deviation as a dummy variable. The first model includes mechanism complexity only, the second and third model additionally control for valuation complexity by including the uncorrected R-range and the corrected R-range, respectively. The corrected R-range controls for mechanism complexity with dummy variables for each deviation. Wilcoxon signed-rank tests were performed to test the significance of the difference between the WTA, WTP, and WTA-WTP gap before and after the correction, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B.3 Correcting the R-range for mechanism complexity

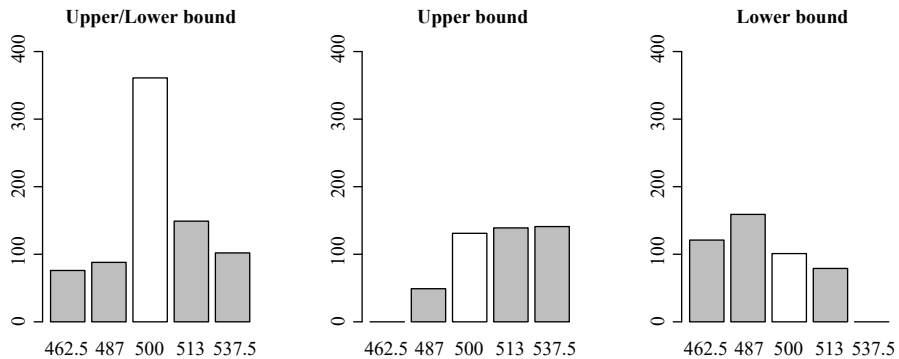


Figure B.1: Distribution of the upper bound and lower bound of the degenerate lottery. The left panel shows the distribution of the bounds for subjects who do not randomize for the degenerate lottery. The right two panels show the distribution of the bounds for subjects who randomize at least once for the degenerate lottery. The white bars show the optimal response of 500 points for the upper and lower bound.

To correct the R-range for mechanism complexity, we use subjects' two bounds of the R-range for the degenerate lottery. The optimal upper bound and lower bound for the degenerate lottery are both 500, and subjects should never choose the randomization option. Deviating from these optimal responses implies mistakes or biases. When subjects never chose the randomization option but had non-optimal bounds for the degenerate lottery, it is unclear how these non-optimal responses affect the R-range systematically. The left histogram in Figure B.1 suggests that these non-optimal responses have no systematic pattern. We, therefore, make no correction of the R-range for these deviations from the optimal response. On the other hand, when subjects chose the randomization option at least once and thus have an R-range for the degenerate lottery, this could represent a systematic bias and may increase the R-range for the lotteries if subjects would exhibit the same tendency in the R-range task for the lotteries. Indeed, the right two histograms in Figure B.1 suggest that these non-optimal responses are systematically biased: The upper bound is systematically biased upward and the lower bound is systematically biased downward.

We correct the R-range for mechanism complexity by using deviations from the optimal responses. In these statistical correction regressions, we exclude subjects who never chose the randomization option but reported non-optimal bounds for the degenerate lottery. As discussed above, their non-optimal responses do not systematically bias the R-range, and it is unclear how to correct these R-ranges. We use their uncorrected R-range in further analysis.

	Lottery	Before correction	Correcting for deviations
Upper Bound	L1	551.62 (233.61)	527.69** (226.85)
	L2	538.55 (229.99)	513.96** (223.89)
	Average	545.09 (209.22)	519.83** (202.08)
Lower Bound	L1	449.14 (239.10)	467.84*** (233.02)
	L2	433.10 (237.23)	451.94*** (230.98)
	Average	441.12 (213.65)	459.89*** (206.76)
R-range	L1	102.48 (172.42)	57.85*** (160.64)
	L2	105.45 (174.91)	62.02*** (160.66)
	Average	103.96 (149.59)	59.94*** (134.24)

Table B.11: The upper bound, the lower bound, and the R-range before and after correction for mechanism complexity. We only correct the upper bound and lower bound for subjects who chose the randomization option at least once for the degenerate lottery. For subjects who did not choose the randomization option the uncorrected upper and lower bound is included in the computation of the mean and standard deviation after correction. Wilcoxon signed-rank tests were performed to test the significance of the difference between the upper bound, the lower bound, and the R-range before and after the correction for mechanism complexity, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.11 reports the corrected bounds and R-range of the two lotteries separately and the averages. We see that mechanism complexity correction significantly lowers the upper bound (average of the two lotteries: 519.83 vs 545.09, $p < 0.05$) and increases the lower bound (average of the two lotteries: 459.89 vs 441.12, $p < 0.01$). This leads to significantly smaller R-ranges (average of the two lotteries: 59.94 vs 103.96, $p < 0.01$).

B.4 Subsample analysis

We performed further analyses by taking a median split of our sample based on the performance on the matrix reasoning questions (median = 3.5) and performed the analysis separately for those subjects who performed above median ($N = 618$) and those who performed below median ($N = 618$).

Above median in matrix reasoning performance				
	L1	L2	Average	Sure
WTA	524.19 (210.59)	514.95 (214.93)	519.57 (195.33)	504.30 (18.37)
WTP	401.27 (217.04)	393.22 (219.37)	397.25 (201.21)	496.05 (20.85)
The WTA-WTP gap	122.92*** (252.77)	121.72*** (261.23)	122.32*** (230.37)	8.25*** (24.83)
Below median in matrix reasoning performance				
	L1	L2	Average	Sure
WTA	530.89 (273.35)	523.85 (275.93)	527.37 (253.80)	505.86 (22.38)
WTP	429.37 (282.67)	422.17 (283.49)	425.77 (263.63)	497.21 (24.33)
The WTA-WTP gap	101.52*** (296.02)	101.68*** (288.17)	101.60*** (254.76)	8.65*** (29.15)

Table B.12: Means of the WTA, the WTP, and the WTA-WTP gap for L1, L2, the average across the two lotteries, and the degenerate lottery. Wilcoxon signed-rank tests were performed to test the significance of the difference between WTA and WTP. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Above median in matrix reasoning performance					
Lottery	Before correction	Correcting mechanism complexity			
		theoretically	for deviations	for deviations different slopes	
WTA	L1	524.19	502.68***	502.95***	495.83
		(210.59)	(223.98)	(190.02)	(189.77)
	L2	514.95	489.42***	496.48*	492.10
		(214.93)	(239.42)	(199.92)	(199.83)
	Average	519.57	496.05***	499.72*	493.96
		(195.33)	(209.61)	(175.94)	(175.77)
WTP	L1	401.27	417.13***	418.42	416.10
		(217.04)	(230.91)	(197.24)	(197.22)
	L2	393.22	409.87**	409.59	399.58***
		(219.37)	(236.48)	(201.61)	(201.28)
	Average	397.25	413.50***	414.01	407.84
		(201.21)	(209.43)	(180.70)	(180.56)
WTA-WTP gap	L1	122.92	85.55***	84.53***	79.73***
		(252.77)	(289.90)	(242.00)	(242.25)
	L2	121.72	79.55***	86.89***	92.51***
		(261.23)	(310.20)	(251.21)	(251.53)
	Average	122.37	82.55***	85.71***	86.12***
		(230.37)	(267.50)	(218.72)	(219.00)
Below median in matrix reasoning performance					
Lottery	Before correction	Correcting mechanism complexity			
		theoretically	for deviations	for deviations different slopes	
WTA	L1	530.89	506.72***	497.37***	478.93***
		(273.35)	(287.40)	(241.55)	(241.00)
	L2	523.85	492.63***	495.29***	488.43***
		(275.93)	(306.40)	(253.48)	(253.40)
	Average	527.37	499.67***	496.33	483.68***
		(253.80)	(267.75)	(224.45)	(224.16)
WTP	L1	429.37	433.87	444.31*	469.96***
		(282.67)	(284.66)	(250.77)	(249.96)
	L2	422.17	425.94*	437.45***	446.16***
		(283.49)	(289.56)	(250.12)	(250.03)
	Average	425.77	429.90	440.88***	458.06***
		(263.63)	(263.41)	(228.23)	(227.83)
WTA-WTP gap	L1	101.52	72.85	53.06***	8.97***
		(296.02)	(333.17)	(284.87)	(283.46)
	L2	101.68	66.69*	57.83***	42.27
		(288.17)	(337.36)	(278.28)	(278.78)
	Average	101.60	69.77*	55.45***	25.62***
		(254.76)	(290.58)	(242.59)	(242.58)

Table B.13: WTA, WTP, and WTA-WTP gap before and after each of the three correction methods. Wilcoxon signed-rank tests were performed to test the significance of the difference between the WTA, WTP, and WTA-WTP gap before the correction and after the correction, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Above median in matrix reasoning performance				
	Lottery	Before controlling for R-range	After controlling for	
			R-range	mechanism complexity corrected R-range
WTA	L1	524.19 (210.59)	515.01*** (210.08)	519.59*** (210.33)
	L2	514.95 (214.93)	516.00*** (214.92)	518.03*** (214.80)
	Average	519.57 (195.33)	512.68*** (195.10)	517.36*** (195.28)
WTP	L1	401.27 (217.04)	401.85*** (217.04)	405.62*** (216.81)
	L2	393.22 (219.37)	403.55*** (218.74)	404.60*** (217.66)
	Average	397.25 (201.21)	404.74*** (200.95)	408.44*** (199.98)
WTA-WTP gap	L1	122.92 (252.77)	113.16*** (252.28)	113.97*** (251.94)
	L2	121.72 (261.23)	112.45*** (260.80)	113.42*** (260.46)
	Average	122.37 (230.37)	107.94*** (229.50)	108.91*** (228.82)
Below median in matrix reasoning performance				
	Lottery	Before controlling for R-range	After controlling for	
			R-range	mechanism complexity corrected R-range
WTA	L1	530.89 (273.35)	524.43*** (273.11)	526.39 (272.88)
	L2	523.85 (275.93)	513.12*** (275.31)	517.01 (275.26)
	Average	527.37 (253.80)	517.22*** (253.35)	520.20*** (253.13)
WTP	L1	429.37 (282.67)	440.16*** (282.02)	435.71*** (281.78)
	L2	422.17 (283.49)	423.44*** (283.48)	422.48*** (283.49)
	Average	425.77 (263.63)	434.06*** (263.33)	431.19*** (263.26)
WTA-WTP gap	L1	101.52 (296.02)	84.27*** (294.43)	90.68 (293.52)
	L2	101.68 (288.17)	89.68*** (287.44)	94.53*** (287.48)
	Average	101.60 (254.76)	83.17*** (253.27)	89.01*** (252.69)

Table B.14: The WTA, the WTP, and the WTA-WTP gap before and after controlling for the (corrected) R-range. Wilcoxon signed-rank tests were performed to test the significance of the difference between the WTA, WTP, and WTA-WTP gap before and after controlling for R-range, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Above median in matrix reasoning performance						
	Lottery	Before controlling	After controlling for both forms of complexity			
			Model 1	Model 2	Model 3	Model 4
WTA	L1	524.19	496.83***	491.39***	498.26***	492.04***
		(210.59)	(189.76)	(189.58)	(189.72)	(189.51)
	L2	514.95	499.66*	495.26	499.15	494.81
		(214.93)	(199.85)	(199.75)	(199.82)	(199.73)
	Average	519.57	496.28***	491.46***	497.10***	491.60***
		(195.33)	(175.88)	(175.73)	(175.87)	(175.70)
WTP	L1	401.27	417.97	415.62	420.72***	418.63***
		(217.04)	(197.24)	(197.22)	(197.17)	(197.15)
	L2	393.22	418.4***	408.23***	418.52***	408.78***
		(219.37)	(201.10)	(200.74)	(200.40)	(200.09)
	Average	397.25	419.83***	413.63***	422.07***	416.45***
		(201.21)	(180.51)	(180.37)	(179.95)	(179.84)
WTA-WTP gap	L1	122.92	78.87***	75.76***	77.54***	73.41***
		(252.77)	(241.75)	(242.01)	(241.46)	(241.72)
	L2	121.72	81.21***	87.03***	80.62***	86.03***
		(261.23)	(251.07)	(251.41)	(250.68)	(250.98)
	Average	122.37	76.45***	77.83***	75.03***	75.15***
		(230.37)	(218.28)	(218.63)	(217.63)	(217.92)
Below median in matrix reasoning performance						
	Lottery	Before controlling	After controlling for both forms of complexity			
			Model 1	Model 2	Model 3	Model 4
WTA	L1	530.89	496.92***	478.58***	496.19***	477.88***
		(273.35)	(241.55)	(240.99)	(241.51)	(240.96)
	L2	523.85	484.48***	477.48***	488.39***	481.46***
		(275.93)	(252.80)	(252.72)	(252.74)	(252.67)
	Average	527.37	490.04***	477.44***	491.75***	479.22***
		(253.80)	(224.24)	(223.96)	(224.11)	(223.84)
WTP	L1	429.37	453.88***	478.10***	449.86***	474.21***
		(282.67)	(250.19)	(249.44)	(249.98)	(249.24)
	L2	422.17	431.90*	441.19***	433.18	442.82***
		(283.49)	(249.94)	(249.83)	(249.82)	(249.71)
	Average	425.77	443.72***	460.04***	442.58***	459.10***
		(263.63)	(228.19)	(227.81)	(228.18)	(227.80)
WTA-WTP gap	L1	101.52	43.05***	0.49***	46.33***	3.67***
		(296.02)	(284.27)	(282.95)	(283.72)	(282.43)
	L2	101.68	52.58***	36.29***	55.21***	38.64***
		(288.17)	(278.28)	(278.83)	(278.39)	(278.94)
	Average	101.60	46.32***	17.40***	49.17***	20.12***
		(254.76)	(242.09)	(242.13)	(241.88)	(241.93)

Table B.15: The WTA, WTP, and WTA-WTP gap before and after controlling for both forms of complexity in two groups based on their performance in matrix reasoning questions. In Model 1 and 3 mechanism complexity is deviations from the optimal response, while in Model 2 and 4 mechanism complexity is deviations from the optimal response while allowing for different slopes for positive and negative deviations. In Model 1 and 2 the R-range is uncorrected, and in Model 3 and 4 the R-range is mechanism complexity corrected. Wilcoxon signed-rank tests were performed to test the significance of the difference between the WTA, WTP, and WTA-WTP gap before the correction and after the correction, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Above median in matrix reasoning performance (N = 306)						
Lottery	Before controlling	Controlling λ	Combined (Model 3)	Combined (Model 3)+ λ	Combined (Model 4)	Combined (Model 4)+ λ
WTA	L1	518.91 (213.48)	525.39*** (213.43)	494.15 (189.36)	495.34*** (189.35)	490.16 (189.28)
	L2	516.18 (216.70)	504.59*** (216.55)	507.93 (200.84)	491.44*** (200.52)	501.92 (200.71)
	Average	517.54 (198.57)	514.99*** (198.56)	498.42 (177.70)	491.59*** (177.64)	492.59 (177.54)
						485.66*** (177.48)
WTP	L1	402.04 (209.94)	462.96*** (208.13)	413.57 (185.89)	450.85*** (185.12)	416.48 (185.86)
	L2	395.83 (222.43)	455.95*** (220.76)	423.42 (197.70)	452.51*** (197.25)	410.61 (197.22)
	Average	398.94 (200.63)	459.46*** (198.75)	421.26 (174.95)	456.02*** (174.23)	416.60 (174.88)
						439.05*** (174.17)
WTA -WTP gap	L1	116.87 (247.44)	62.43*** (247.14)	80.59 (233.24)	44.48*** (233.06)	73.67 (233.28)
	L2	120.34 (257.05)	48.64*** (254.78)	84.51 (250.08)	38.93*** (248.72)	91.31 (250.75)
	Average	118.61 (224.42)	55.53*** (223.01)	77.16 (212.53)	35.57*** (211.66)	76.00 (212.96)
						34.06*** (212.01)
Below median in matrix reasoning performance (N = 227)						
Lottery	Before controlling	Controlling λ	Combined (Model 3)	Combined (Model 3)+ λ	Combined (Model 4)	Combined (Model 4)+ λ
WTA	L1	513.99 (266.69)	507.81*** (266.65)	487.38 (238.10)	479.48*** (238.03)	467.87 (237.42)
	L2	511.89 (274.61)	520.30*** (274.54)	478.13 (250.40)	483.01*** (250.38)	468.66 (250.25)
	Average	512.94 (255.54)	514.06*** (255.54)	482.17 (227.88)	481.00*** (227.88)	468.88 (227.55)
						468.42*** (227.55)
WTP	L1	434.86 (282.08)	425.88*** (282.05)	447.53 (251.76)	424.66*** (251.48)	461.46 (251.49)
	L2	420.21 (279.50)	412.41*** (279.47)	425.16 (251.86)	401.06*** (251.56)	446.46 (251.25)
	Average	427.53 (265.76)	419.15*** (265.73)	435.32 (235.35)	411.37*** (235.03)	454.57 (234.79)
						431.56*** (234.52)
WTA -WTP gap	L1	79.13 (295.06)	81.94*** (295.16)	39.85 (285.48)	54.82*** (285.67)	6.41 (285.64)
	L2	91.69 (295.68)	107.88*** (295.57)	52.98 (279.10)	81.94*** (279.03)	22.20 (279.79)
	Average	85.41 (267.62)	94.91*** (267.67)	46.85 (253.13)	69.63*** (253.24)	14.32 (253.52)
						36.86*** (253.59)

Table B.16: WTA, WTP, and WTA-WTP gap before and after controlling for mechanism complexity and the (mechanism complexity corrected) R-range as well as the loss aversion coefficient λ . In Controlling for λ we consider only the loss aversion λ in the regression. Combined models (Model 3 and Model 4) are those defined in Table D.4, which control for both mechanism complexity and the (corrected) R-range. In Combined (Model 3 or 4) + λ , we include additionally λ . The sample used for this part consist of 306 and 227 subjects, above and below median performance respectively, who overlap with the study from which we obtained the loss aversion measure. Wilcoxon signed-rank tests were performed to test the significance of the difference between the WTA, WTP, and WTA-WTP gap before and after the inclusion of λ , * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B.5 Theoretical analysis of mechanism complexity and the R-range

As explained in the experimental design, we elicited a measure of mechanism complexity and valuation complexity for each subject. Below we provide an illustration of how these two factors relate to the WTA-WTP gap and how to control the WTA-WTP gap for these two measures.

B.5.1 Correcting theoretically the WTA-WTP gap for mechanism complexity

For analytical convenience, we assume a linear utility function. The decision maker perceives it as a loss when giving up the lottery and as a gain when receiving the lottery. It can then be shown that $WTP = EV_X/\lambda$ and $WTA = \lambda EV_X$, where EV_X denote the expected value of the lottery and λ the loss aversion coefficient. In line with Cason and Plott (2014), we assume the decision maker is subject to mechanism complexity of the first price auction when reporting WTA and WTP. Let $0 \leq \alpha_i \leq 1$ denote the degree of mechanism complexity, with a small α indicating a weaker mechanism complexity, and $i=a$ for WTA and p for WTP. Let p denote a price, $\underline{p}$ and $\bar{p}$ the minimum and the maximum prices in the price list. The expected payoff (π) of stating WTP or WTA in the price list is, respectively:

$$\begin{aligned} E(\pi_{WTP}) &= \frac{1}{\bar{p} - \underline{p}} \int_{\underline{p}}^{WTP} [EV_X - \lambda \alpha_p WTP - (1 - \alpha_p) \lambda p] dp \\ &= \frac{1}{\bar{p} - \underline{p}} \left\{ (WTP - \underline{p}) [EV_X - \lambda \alpha_p WTP] \right. \\ &\quad \left. - (1 - \alpha_p) \lambda (WTP^2/2 - \underline{p}^2/2) \right\}, \\ E(\pi_{WTA}) &= \frac{1}{\bar{p} - \underline{p}} \int_{WTA}^{\bar{p}} [\alpha_a WTA + (1 - \alpha_a) p - \lambda EV_X] dp \\ &= \frac{1}{\bar{p} - \underline{p}} \left\{ (\bar{p} - WTA) [\alpha_a WTA - \lambda EV_X] \right. \\ &\quad \left. - (1 - \alpha_a) (\bar{p}^2/2 - WTA^2/2) \right\}. \end{aligned}$$

The optimal WTA and WTP that the decision maker reports are then:

$$\begin{aligned} WTP &= \frac{1}{(1 + \alpha_p)} EV_X/\lambda + \frac{\alpha_p}{(1 + \alpha_p)} \underline{p} \\ WTA &= \frac{1}{1 + \alpha_a} \lambda EV_X + \frac{\alpha_a}{1 + \alpha_a} \bar{p} \end{aligned}$$

When X is a sure payment, the decision maker exchanges between sure payments and there should be no loss aversion. Thus, if the decision maker correctly understands the incentives in the price list, she should report $WTA=WTP=500$ for the degenerate lottery. Deviating from 500 implies mechanism complexity. We can infer the decision maker's

degree of mechanism complexity about the price list (α_i) from the reported WTA_{sure} and WTP_{sure} for the degenerate lottery as:

$$\begin{aligned}\alpha_p &= \frac{500 - WTP_{sure}}{WTP_{sure} - 400} \\ \alpha_a &= \frac{WTA_{sure} - 500}{600 - WTA_{sure}},\end{aligned}$$

where 400 is $\underline{p}$ and 600 $\bar{p}$ of the degenerate lottery. Correcting for mechanism complexity, the preference-based WTP^* and WTA^* for the lottery are respectively:

$$\begin{aligned}WTP^* &= EV_X / \lambda = WTP + \alpha_p(WTP - \underline{p}) \\ WTA^* &= \lambda EV_X = WTA - \alpha_a(\bar{p} - WTA),\end{aligned}$$

where $\underline{p} = 0$ and $\bar{p} = 1000$ for the two lotteries in our experiment. Inserting these values in the above equations, the underlying WTP^* and WTA^* corrected for mechanism complexity are:

$$\begin{aligned}WTP^* &= \frac{100}{WTP_{sure} - 400} WTP \\ WTA^* &= \frac{100}{600 - WTA_{sure}} WTA - \frac{WTA_{sure} - 500}{600 - WTA_{sure}} 1000.\end{aligned}$$

Note that $WTP^* = WTP$ and $WTA^* = WTA$ when $WTP_{sure} = 500$ and $WTA_{sure} = 500$.

B.5.2 What the R-range reveals

There are a number of ways to relate the R-range to valuation complexity. For example, the literature of incomplete preferences assumes that decision makers may have incomplete preferences over some choices, and they choose to randomize between the options when indecisive (Bewley, 2002; Eliaz and Ok, 2006; Cettolin and Riedl, 2019; Halevy et al., 2023). Below we discuss how the R-range captures valuation complexity in the framework of Fudenberg et al. (2015). We then motivate our statistical analysis about how to control the WTA-WTP gap for valuation complexity.

When facing a lottery X and a price y , a decision maker behaving according to Fudenberg et al.'s (2015) functional form chooses the optimal randomization probability as

$$p^* = \operatorname{argmax}_p \quad pU(X) + (1 - p)U(y) - c(p) - c(1 - p),$$

where $U(\cdot)$ is the Von Neumann–Morgenstern expected utility, $u(\cdot)$ is the Bernoulli utility, and $c(p)$ is a convex and continuously differentiable cost function. Fudenberg et al. (2015) show that their representation corresponds to a form of an uncertainty averse decision maker who is unsure about her true utility and uses randomization to hedge her preference uncertainty. When the decision maker faces the lottery X , a price y , and a randomization option where the decision maker receives either option with probability 0.5, she may strictly

prefer the randomization option over X or y . To see this, notice that the first order condition of the above equation is simply $c'(p) - c'(1-p) = U(X) - u(y)$, which predicts that the optimal p is close to 0.5 when $U(X) - U(y)$ is small.

We can infer the level of valuation complexity that the decision maker perceives over the choice problem between X and y from the range of prices y that the decision maker chooses the randomization option. Define the smallest ($\underline{y}$) and largest ($\bar{y}$) prices of y that such that the decision maker prefers the randomization option as

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

The term $2[c(1.0) + c(0) - 2c(0.5)] = \frac{c(1.0)-c(0.5)}{1-0.5} - \frac{[c(0.5)-c(0)]}{0.5-0}$ measures the convexity of the cost function, which, according to Fudenberg et al. (2015), is positively related to valuation complexity. This is most clear when the cost function takes the form of $c(p) = p \log(p)/\eta$. Simple calculation shows that the optimal randomization probability is the familiar logit/logistic choice rule: $p^* = \frac{e^{\eta U(X)}}{e^{\eta U(X)} + e^{\eta u(y)}}$. As shown by Holman and Marley, the parameter $1/\eta$ can be linked to the variance of the i.i.d. Gumbel preference shocks in a random utility representation (Luce and Suppes, 1965, p.338), with a larger η corresponding to a smaller valuation complexity. In light of Fudenberg et al. (2015), the cost function may also relate to a measure of control desirability because, as Fudenberg et al. (2015) point out, a preference for randomization arises because the decision maker trades off the probability of errors against the cost of making the desired choice. A decision maker who prefers to maintain controls may perceive the implementation cost to be small and is less willing to randomize.

Let $\Delta = 2[c(1.0) + c(0) - 2c(0.5)]$. It follows that we can estimate a measure of valuation complexity Δ from the R-range as $\Delta = \frac{u(\bar{y})-u(\underline{y})}{2}$. Note that the “true” valuation of the lottery is $U(X) = [u(\bar{y}) + u(\underline{y})]/2$. When the decision maker perceives valuation complexity about the lottery and is asked for WTP and WTA, she behaves cautiously and understates her WTP and overstates her WTA, respectively (Cerreia-Vioglio et al., 2024), resulting a positive WTA-WTP gap. This could arise even when the decision maker is not loss averse. One way to capture the above mechanism is:

$$\begin{aligned} u(WTP) &= U(X) - \delta_p \Delta \\ u(WTA) &= U(X) + \delta_a \Delta, \end{aligned}$$

where $\delta_i, i = a, b$ denotes the decision maker's cautious attitude toward valuation complexity when stating WTP and WTA. This motivates the following analysis in the main text:

$$WTA_X = I + \beta_1 R\text{-range}_X + \epsilon$$

$$WTP_X = I + \beta_2 R\text{-range}_X + \epsilon,$$

where I is the intercept, X is L1, L2, or the average of the two lotteries.

B.6 Regressions to control WTA and WTP for mechanism complexity or/and valuation complexity

Explanatory variables	WTA		
	L1	L2	Average
$(WTA_{sure} - 500)$	5.40*** (0.30)	4.64*** (0.32)	5.02*** (0.28)
Intercept	500.09*** (6.37)	495.82*** (6.69)	497.95*** (5.91)
Adjusted R^2	0.21	0.15	0.21
Explanatory variables	WTP		
	L1	L2	Average
$(500 - WTP_{sure})$	-4.94*** (0.28)	-4.93*** (0.29)	-4.94*** (0.26)
Intercept	431.97*** (6.50)	424.30*** (6.56)	428.13*** (5.94)
Adjusted R^2	0.20	0.19	0.23

Table B.17: Regression results for WTA and WTP correction for mechanism complexity using the deviation from the optimal responses for the WTA and WTP of the degenerate lottery as a continuous variable. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Explanatory variables	WTA		
	L1	L2	Average
$(WTA_{sure} - 500)$	6.05*** (0.44)	4.94*** (0.46)	5.50*** (0.41)
$D_{neg}(WTA_{sure} - 500)$	-1.70** (0.85)	-0.80 (0.90)	-1.25 (0.79)
Intercept	488.82*** (8.51)	490.50*** (8.95)	489.66*** (7.90)
Adjusted R^2	0.21	0.15	0.21
Explanatory variables	WTP		
	L1	L2	Average
$(500 - WTP_{sure})$	-5.27*** (0.46)	-4.63*** (0.47)	-4.95*** (0.50)
$D_{neg}(500 - WTP_{sure})$	0.74 (0.84)	-0.70 (0.84)	0.02 (0.76)
Intercept	438.02*** (9.41)	418.60*** (9.49)	428.31*** (8.60)
Adjusted R^2	0.20	0.19	0.23

Table B.18: Regression results for WTA and WTP correction for mechanism complexity using the deviation from the optimal responses for the WTA and WTP of the degenerate lottery as a continuous variable while allowing for different slopes for positive and negative deviations. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Explanatory variables	WTA		
	L1	L2	Average
R-range	0.08* (0.04)	0.05 (0.04)	0.08* (0.04)
Intercept	519.84*** (8.06)	514.30*** (8.21)	514.92*** (7.84)
Adjusted R^2	0.002	0.000	0.002
Explanatory variables	WTP		
	L1	L2	Average
R-range	-0.06 (0.04)	-0.05 (0.04)	-0.08* (0.04)
Intercept	421.55*** (8.35)	413.37*** (8.43)	419.62*** (8.13)
Adjusted R^2	0.001	0.001	0.002

Table B.19: Regression results for controlling WTA and WTP for the R-Range. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Explanatory variables	Upper bound		
	L1	L2	Average
(Upper _{sure} - 500)	4.96*** (0.46)	4.71*** (0.45)	4.84*** (0.40)
Intercept	510.30*** (7.74)	498.61*** (7.59)	504.45*** (6.75)
Adjusted R^2	0.12	0.12	0.15
Explanatory variables	Lower bound		
	L1	L2	Average
(500-Lower _{sure})	-4.14*** (0.48)	-4.18*** (0.47)	-4.16*** (0.42)
Intercept	459.92*** (7.76)	443.78*** (7.58)	451.85*** (6.72)
Adjusted R^2	0.08	0.09	0.11

Table B.20: Regression results for R-range correction for mechanism complexity using the deviation from the optimal responses for the randomization task of the degenerate lottery as a continuous variable. For this regression subjects who did not choose the randomization option for the sure lottery and reported an upper bound and lower bound other than 501 and 499, respectively, are excluded. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Explanatory variables	WTA		
	L1	L2	Average
R-range after correction	0.07* (0.04)	0.03 (0.04)	0.07 (0.05)
Intercept	523.27*** (7.37)	517.64*** (7.54)	519.13*** (7.04)
Adjusted R^2	0.002	0.000	0.001
Explanatory variables	WTP		
	L1	L2	Average
R-range after correction	-0.11** (0.04)	-0.09** (0.04)	-0.15*** (0.05)
Intercept	421.84*** (7.61)	413.54*** (7.73)	420.44*** (7.29)
Adjusted R^2	0.004	0.003	0.006

Table B.21: Regression results for controlling WTA and WTP for the R-Range after the correction for mechanism complexity using deviations from the optimal upper and lower bound of the degenerate lottery as a continuous variable. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Explanatory variables	WTA		
	L1	L2	Average
(WTA _{sure} - 500)	5.39*** (0.30)	4.64*** (0.32)	5.00*** (0.28)
R-range (uncorrected)	0.03 (0.04)	0.04 (0.04)	0.05 (0.04)
Intercept	496.95*** (7.31)	491.81*** (7.74)	493.06*** (7.10)
Adjusted R^2	0.21	0.15	0.21
Explanatory variables	WTP		
	L1	L2	Average
(500 - WTP _{sure})	-4.94*** (0.28)	-4.92*** (0.29)	-4.92*** (0.26)
R-range (uncorrected)	-0.05 (0.04)	-0.02 (0.04)	-0.04 (0.04)
Intercept	437.00*** (7.53)	425.90*** (7.61)	432.68*** (7.19)
Adjusted R^2	0.20	0.19	0.23

Table B.22: Regression results for WTA and WTP correction for mechanism complexity (continuous deviations from the optimal WTA and WTP of the degenerate lottery) and the uncorrected R-Range combined. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Explanatory variables	WTA		
	L1	L2	Average
(WTA _{sure} – 500)	6.01*** (0.44)	4.93*** (0.46)	5.46*** (0.41)
D _{neg} (WTA _{sure} – 500)	–1.65* (0.85)	–0.78 (0.90)	–1.20 (0.79)
R-range (uncorrected)	0.03 (0.04)	0.04 (0.04)	0.04 (0.04)
Intercept	486.45*** (9.10)	486.69*** (9.71)	485.40*** (8.71)
Adjusted R ²	0.21	0.15	0.21

Explanatory variables	WTP		
	L1	L2	Average
(500–WTP _{sure})	–5.26*** (0.46)	–4.61*** (0.47)	–4.92*** (0.42)
D _{neg} (500–WTP _{sure})	0.73 (0.84)	–0.72 (0.84)	–0.01 (0.76)
R-range (uncorrected)	–0.05 (0.04)	–0.02 (0.04)	–0.04 (0.04)
Intercept	442.88*** (10.12)	420.21*** (10.13)	432.60*** (9.41)
Adjusted R ²	0.20	0.19	0.23

Table B.23: Regression results for WTA and WTP correction for mechanism complexity (continuous deviations from the optimal WTA and WTP of the degenerate lottery while allowing for different slopes for positive and negative deviations) and the uncorrected R-Range combined. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Explanatory variables	WTA		
	L1	L2	Average
WTA _{sure} – 500	5.39*** (0.30)	4.64*** (0.32)	5.01*** (0.28)
R-range after mechanism complexity correction	0.04 (0.04)	0.03 (0.04)	0.06 (0.04)
Intercept	497.57*** (6.73)	493.75*** (7.15)	494.58*** (6.43)
Adjusted R ²	0.21	0.15	0.21

Explanatory variables	WTP		
	L1	L2	Average
(500–WTP _{sure})	–4.92*** (0.28)	–4.91*** (0.29)	–4.90*** (0.26)
R-range after mechanism complexity correction	–0.09** (0.04)	–0.04 (0.04)	–0.09** (0.04)
Intercept	436.89*** (6.88)	426.52*** (6.99)	433.77*** (6.46)
Adjusted R ²	0.20	0.19	0.23

Table B.24: Regression results for WTA and WTP correction for mechanism complexity (continuous deviations from the optimal WTA and WTP of the degenerate lottery) and the R-Range after mechanism complexity correction (continuous deviations from the optimal upper and lower bound of the degenerate lottery) combined. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Explanatory variables	WTA		
	L1	L2	Average
$(WTA_{sure} - 500)$	6.02*** (0.44)	4.95*** (0.46)	5.49*** (0.41)
$D_{neg}(WTA_{sure} - 500)$	-1.67* (0.85)	-0.82 (0.90)	-1.24 (0.79)
R-range after mechanism complexity correction	0.04 (0.04)	0.03 (0.04)	0.06 (0.04)
Intercept	486.62*** (8.74)	488.30*** (9.33)	486.37*** (8.29)
Adjusted R^2	0.21	0.15	0.21
Explanatory variables	WTP		
	L1	L2	Average
$(500 - WTP_{sure})$	-5.26*** (0.46)	-4.60*** (0.47)	-4.91*** (0.42)
$D_{neg}(500 - WTP_{sure})$	0.78 (0.83)	-0.72 (0.84)	0.02 (0.76)
R-range after mechanism complexity correction	-0.09** (0.04)	-0.04 (0.04)	-0.09** (0.04)
Intercept	443.26*** (9.70)	420.73*** (9.76)	433.52*** (8.96)
Adjusted R^2	0.20	0.19	0.23

Table B.25: Regression results for WTA and WTP correction for mechanism complexity (continuous deviations from the optimal WTA and WTP of the degenerate lottery while allowing for different slopes for positive and negative deviations), and the R-range after mechanism complexity correction (continuous deviations from the optimal upper and lower bound of the degenerate lottery) combined.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Explanatory variables	WTA		
	L1	L2	Average
λ	-0.26 (6.40)	0.95 (6.55)	0.35 (6.04)
Intercept	517.39*** (17.71)	512.21*** (18.12)	514.80*** (16.74)
Adjusted R^2	0.000	0.000	0.000
Explanatory variables	WTP		
	L1	L2	Average
$1/\lambda$	50.24 (38.70)	49.90 (39.47)	50.07 (36.67)
Intercept	444.56*** (24.38)	434.57*** (24.87)	439.56*** (23.10)
Adjusted R^2	0.001	0.001	0.002

Table B.26: Regression results for WTA and WTP correction for loss aversion. WTA is corrected for λ and WTP for $1/\lambda$. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Explanatory variables	WTA		
	L1	L2	Average
λ	1.28 (5.71)	2.20 (6.06)	1.79 (5.42)
$WTA_{sure} - 500$	5.37*** (0.46)	4.63*** (0.49)	4.99*** (0.44)
R-range after mechanism complexity correction	0.03 (0.05)	0.01 (0.06)	0.06 (0.06)
Intercept	489.03*** (16.37)	488.99*** (17.38)	486.79*** (15.69)
Adjusted R^2	0.20	0.14	0.20
Explanatory variables	WTP		
	L1	L2	Average
$1/\lambda$	15.62 (34.70)	14.60 (35.76)	15.53 (32.51)
$(500 - WTP_{sure})$	-4.96*** (0.43)	-4.76*** (0.44)	-4.86*** (0.40)
R-range after mechanism complexity correction	-0.05 (0.05)	-0.11 (0.06)	-0.09 (0.06)
Intercept	439.10*** (22.16)	432.12*** (22.73)	436.39*** (20.83)
Adjusted R^2	0.20	0.19	0.22

Table B.27: Regression results for WTA and WTP correction for loss aversion, mechanism complexity (continuous deviations from the optimal WTA and WTP of the degenerate lottery), and the R-range after mechanism complexity correction (continuous deviations from the optimal upper and lower bound of the degenerate lottery) combined. WTA is corrected for λ and WTP for $1/\lambda$. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Explanatory variables	WTA		
	L1	L2	Average
λ	1.18 (5.71)	2.12 (6.07)	1.70 (5.42)
$WTA_{sure} - 500$	5.90*** (0.69)	5.08*** (0.73)	5.47*** (0.65)
$D_{neg}(WTA_{sure} - 500)$	-1.32 (1.28)	-1.11 (1.35)	-1.22 (1.21)
R-range after mechanism complexity correction	0.03 (0.05)	0.01 (0.06)	0.06 (0.06)
Intercept	480.83*** (18.18)	481.78*** (19.46)	479.06*** (17.45)
Adjusted R^2	0.20	0.14	0.20
Explanatory variables	WTP		
	L1	L2	Average
$1/\lambda$	16.82 (34.79)	13.87 (35.89)	15.74 (32.60)
$(500 - WTP_{sure})$	-5.29*** (0.70)	-4.55*** (0.73)	-4.92*** (0.66)
$D_{neg}(500 - WTP_{sure})$	0.72 (1.22)	-0.44 (1.26)	0.13 (1.15)
R-range after mechanism complexity correction	-0.05 (0.05)	-0.11* (0.06)	-0.09 (0.06)
Intercept	445.41*** (24.65)	428.17*** (25.40)	437.50*** (23.22)
Adjusted R^2	0.20	0.19	0.22

Table B.28: Regression results for WTA and WTP correction for loss aversion, mechanism complexity (continuous deviations from the optimal WTA and WTP of the degenerate lottery while allowing for different slopes for positive and negative deviations), and the R-range after mechanism complexity correction (continuous deviations from the optimal upper and lower bound of the degenerate lottery) combined. WTA is corrected for λ and WTP for $1/\lambda$. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B.7 Experimental materials

U krijgt aan het begin van deze vraag een loterij. Deze loterij is nu in uw bezit. De loterij geeft u:

- 50% kans op 600 punten.
- en 50% kans op 100 punten.

U hebt steeds 2 mogelijkheden:

1. U houdt de loterij
2. U verkoopt de loterij voor een bepaald aantal punten.

Geef hieronder aan wat u voor ieder aantal punten kiest. Het aantal punten neemt in iedere keuze toe. De loterij blijft steeds hetzelfde. De computer vult automatisch de keuzes aan als u in een rij een keuze maakt. Controleert u daarna of u voor de overige rijen inderdaad die keuzes zou maken.

(om de loterij nogmaals te bekijken kunt u in iedere keuze met uw muis over 'loterij' bewegen).

	Optie 1	Optie 2	
	Houd de <u>loterij</u>	Verkoop de loterij voor 0 punten	
	Houd de <u>loterij</u>	Verkoop de loterij voor 100 punten	
	Houd de <u>loterij</u>	Verkoop de loterij voor 200 punten	
	Houd de <u>loterij</u>	Verkoop de loterij voor 250 punten	
	Houd de <u>loterij</u>	Verkoop de loterij voor 300 punten	
	Houd de <u>loterij</u>	Verkoop de loterij voor 325 punten	
	Houd de <u>loterij</u>	Verkoop de loterij voor 350 punten	
	Houd de <u>loterij</u>	Verkoop de loterij voor 375 punten	
	Houd de <u>loterij</u>	Verkoop de loterij voor 400 punten	
	Houd de <u>loterij</u>	Verkoop de loterij voor 425 punten	
	Houd de <u>loterij</u>	Verkoop de loterij voor 450 punten	
	Houd de <u>loterij</u>	Verkoop de loterij voor 475 punten	
	Houd de <u>loterij</u>	Verkoop de loterij voor 500 punten	
	Houd de <u>loterij</u>	Verkoop de loterij voor 525 punten	
	Houd de <u>loterij</u>	Verkoop de loterij voor 550 punten	
	Houd de <u>loterij</u>	Verkoop de loterij voor 600 punten	
	Houd de <u>loterij</u>	Verkoop de loterij voor 700 punten	
	Houd de <u>loterij</u>	Verkoop de loterij voor 800 punten	
	Houd de <u>loterij</u>	Verkoop de loterij voor 900 punten	
	Houd de <u>loterij</u>	Verkoop de loterij voor 1000 punten	

De beste manier om uw keuze te maken is om bij de eerste rij te beginnen. Zolang u de mogelijkheid links aantrekkelijker vindt dan de mogelijkheid rechts gaat u steeds een rij verder. Als u de mogelijkheid rechts aantrekkelijker vindt dan de mogelijkheid links kiest u in die rij de mogelijkheid rechts. De computer vult dan de rest van de keuzes aan. Controleert u daarna of u inderdaad die keuzes zou maken.

[Verder](#)




Figure B.2: Example of the WTA task for L1.

U heeft aan het begin van deze vraag 1000 punten gekregen. U kunt met deze punten een loterij kopen. Deze loterij geeft u:

- 50% kans op 900 punten,
- en 50% kans op 100 punten.

U hebt steeds 2 mogelijkheden:

1. U houdt de 1000 punten
2. U koopt de loterij voor een bepaald aantal punten en houdt de rest van de 1000 punten.

Bijvoorbeeld: Als u de loterij koopt voor 200 punten, houdt u 800 punten over. Daarnaast ontvangt u de winst van de loterij.

Geef hieronder aan wat u voor ieder aantal punten kiest. Het aantal punten neemt in iedere keuze toe. De loterij blijft steeds hetzelfde. De computer vult automatisch de keuzes aan als u in een rij een keuze maakt. Controleert u daarna of u voor de overige rijen inderdaad die keuzes zou maken.

(om de loterij nogmaals te bekijken kunt u in iedere keuze met uw muis over 'loterij' bewegen).

Optie 1		Optie 2	
 Koop de <u>loterij</u> voor 0 punten en houd de overgebleven 1000 punten	Houd de 1000 punten		
 Koop de <u>loterij</u> voor 100 punten en houd de overgebleven 900 punten	Houd de 1000 punten		
 Koop de <u>loterij</u> voor 200 punten en houd de overgebleven 800 punten	Houd de 1000 punten		
 Koop de <u>loterij</u> voor 250 punten en houd de overgebleven 750 punten	Houd de 1000 punten		
 Koop de <u>loterij</u> voor 300 punten en houd de overgebleven 700 punten	Houd de 1000 punten		
 Koop de <u>loterij</u> voor 325 punten en houd de overgebleven 675 punten	Houd de 1000 punten		
 Koop de <u>loterij</u> voor 350 punten en houd de overgebleven 650 punten	Houd de 1000 punten		
 Koop de <u>loterij</u> voor 375 punten en houd de overgebleven 625 punten	Houd de 1000 punten		
 Koop de <u>loterij</u> voor 400 punten en houd de overgebleven 600 punten	Houd de 1000 punten		
 Koop de <u>loterij</u> voor 425 punten en houd de overgebleven 575 punten	Houd de 1000 punten		
 Koop de <u>loterij</u> voor 450 punten en houd de overgebleven 550 punten	Houd de 1000 punten		
 Koop de <u>loterij</u> voor 475 punten en houd de overgebleven 525 punten	Houd de 1000 punten		
 Koop de <u>loterij</u> voor 500 punten en houd de overgebleven 500 punten	Houd de 1000 punten		
 Koop de <u>loterij</u> voor 525 punten en houd de overgebleven 475 punten	Houd de 1000 punten		
 Koop de <u>loterij</u> voor 550 punten en houd de overgebleven 450 punten	Houd de 1000 punten		
 Koop de <u>loterij</u> voor 600 punten en houd de overgebleven 400 punten	Houd de 1000 punten		
 Koop de <u>loterij</u> voor 700 punten en houd de overgebleven 300 punten	Houd de 1000 punten		
 Koop de <u>loterij</u> voor 800 punten en houd de overgebleven 200 punten	Houd de 1000 punten		
 Koop de <u>loterij</u> voor 900 punten en houd de overgebleven 100 punten	Houd de 1000 punten		
 Koop de <u>loterij</u> voor 1000 punten en houd de overgebleven 0 punten	Houd de 1000 punten		

De beste manier om uw keuze te maken is om bij de eerste rij te beginnen. Zolang u de mogelijkheid links aantrekkelijker vindt dan de mogelijkheid rechts gaat u steeds een rij verder. Als u de mogelijkheid rechts aantrekkelijker vindt dan de mogelijkheid links kiest u in die rij de mogelijkheid rechts. De computer vult dat de rest van de keuzes aan. Controleert u daarna of u inderdaad die keuzes zou maken.

[Verder](#)




Figure B.3: Example of the WTP task for L1.

U krijgt aan het begin van deze vraag een loterij. Deze loterij is nu in uw bezit. De loterij geeft u:

- 100% kans op 500 punten.

U hebt steeds 2 mogelijkheden:

1. U houdt de loterij
2. U verkoopt de loterij voor een bepaald aantal punten.

Geef hieronder aan wat u voor ieder aantal punten kiest. Het aantal punten neemt in iedere keuze toe. De loterij blijft steeds hetzelfde. De computer vult automatisch de keuzes aan als u in een rij een keuze maakt. Controleert u daarna of u voor de overige rijen inderdaad die keuzes zou maken.

(om de loterij nogmaals te bekijken kunt u in iedere keuze met uw muis over 'loterij' bewegen).

	Optie 1	Optie 2	
☐	Houd de loterij	Verkoop de loterij voor 450 punten	☐
☒	Houd de loterij	Verkoop de loterij voor 475 punten	☐
☒	Houd de loterij	Verkoop de loterij voor 499 punten	☐
☐	Houd de loterij	Verkoop de loterij voor 501 punten	☒
☐	Houd de loterij	Verkoop de loterij voor 525 punten	☒
☐	Houd de loterij	Verkoop de loterij voor 550 punten	☒

De beste manier om uw keuze te maken is om bij de eerste rij te beginnen. Zolang u de mogelijkheid links aantrekkelijker vindt dan de mogelijkheid rechts gaat u steeds een rij verder. Als u de mogelijkheid rechts aantrekkelijker vindt dan de mogelijkheid links kiest u in die rij de mogelijkheid rechts. De computer vult dan de rest van de keuzes aan. Controleert u daarna of u inderdaad die keuzes zou maken.

Verder

U heeft aan het begin van deze vraag 1000 punten gekregen. U kunt met deze punten een loterij kopen. Deze loterij geeft u:

- 100% kans op 500 punten.

U hebt steeds 2 mogelijkheden:

1. U houdt de 1000 punten
2. U koopt de loterij voor een bepaald aantal punten en houdt de rest van de 1000 punten.

Bijvoorbeeld: Als u de loterij koopt voor 200 punten, houdt u 800 punten over. Daarnaast ontvangt u de winst van de loterij.

Geef hieronder aan wat u voor ieder aantal punten kiest. Het aantal punten neemt in iedere keuze toe. De loterij blijft steeds hetzelfde. De computer vult automatisch de keuzes aan als u in een rij een keuze maakt. Controleert u daarna of u voor de overige rijen inderdaad die keuzes zou maken.

(om de loterij nogmaals te bekijken kunt u in iedere keuze met uw muis over 'loterij' bewegen).

	Optie 1	Optie 2	
☒	Koop de loterij voor 450 punten en houd de overgebleven 550 punten	Houd de 1000 punten	☐
☒	Koop de loterij voor 475 punten en houd de overgebleven 525 punten	Houd de 1000 punten	☐
☒	Koop de loterij voor 499 punten en houd de overgebleven 501 punten	Houd de 1000 punten	☐
☐	Koop de loterij voor 501 punten en houd de overgebleven 499 punten	Houd de 1000 punten	☒
☐	Koop de loterij voor 525 punten en houd de overgebleven 475 punten	Houd de 1000 punten	☒
☐	Koop de loterij voor 550 punten en houd de overgebleven 450 punten	Houd de 1000 punten	☒

De beste manier om uw keuze te maken is om bij de eerste rij te beginnen. Zolang u de mogelijkheid links aantrekkelijker vindt dan de mogelijkheid rechts gaat u steeds een rij verder. Als u de mogelijkheid rechts aantrekkelijker vindt dan de mogelijkheid links kiest u in die rij de mogelijkheid rechts. De computer vult dat de rest van de keuzes aan. Controleert u daarna of u inderdaad die keuzes zou maken.

Verder

Figure B.4: Example of the WTA and WTP task for the degenerate lottery.

De loterij in deze vraag geeft u:

- 50% kans op 600 punten,
- en 50% kans op 100 punten.

U hebt steeds 3 mogelijkheden:

1. U kiest de loterij.
2. U laat de computer bepalen door het opgooien van een virtuele munt: bij kop krijgt u de loterij en bij munt het aantal punten.
3. U kiest het aantal punten.

Geef hieronder aan wat u voor ieder aantal punten kiest. Het aantal punten neemt in iedere keuze toe. De loterij blijft steeds hetzelfde. De computer vult automatisch de keuzes aan als u in een rij een keuze maakt. Controleert u daarna of u voor de overige rijen inderdaad die keuzes zou maken.

(om de loterij nogmaals te bekijken kunt u in iedere keuze met uw muis over 'loterij' bewegen).

Optie 1	Optie 2	Optie 3
☐ Kies de loterij	☐ De computer bepaalt	Kies 0 punten ☐
☒ Kies de loterij	☐ De computer bepaalt	Kies 100 punten ☐
☒ Kies de loterij	☐ De computer bepaalt	Kies 200 punten ☐
☒ Kies de loterij	☐ De computer bepaalt	Kies 250 punten ☐
☒ Kies de loterij	☐ De computer bepaalt	Kies 300 punten ☐
☒ Kies de loterij	☐ De computer bepaalt	Kies 325 punten ☐
☒ Kies de loterij	☐ De computer bepaalt	Kies 350 punten ☐
☒ Kies de loterij	☐ De computer bepaalt	Kies 375 punten ☐
☐ Kies de loterij	☒ De computer bepaalt	Kies 400 punten ☐
☐ Kies de loterij	☒ De computer bepaalt	Kies 425 punten ☐
☐ Kies de loterij	☒ De computer bepaalt	Kies 450 punten ☐
☐ Kies de loterij	☐ De computer bepaalt	Kies 475 punten ☒
☐ Kies de loterij	☐ De computer bepaalt	Kies 500 punten ☒
☐ Kies de loterij	☐ De computer bepaalt	Kies 525 punten ☒
☐ Kies de loterij	☐ De computer bepaalt	Kies 550 punten ☒
☐ Kies de loterij	☐ De computer bepaalt	Kies 600 punten ☒
☐ Kies de loterij	☐ De computer bepaalt	Kies 700 punten ☒
☐ Kies de loterij	☐ De computer bepaalt	Kies 800 punten ☒
☐ Kies de loterij	☐ De computer bepaalt	Kies 900 punten ☒
☐ Kies de loterij	☐ De computer bepaalt	Kies 1000 punten ☒

De beste manier om uw keuze te maken is om bij de eerste rij te beginnen. Zolang u de mogelijkheid links aantrekkelijker vindt dan de mogelijkheden in het midden en rechts gaat u steeds een rij verder. In de laatste rij waar u de mogelijkheid links het meest aantrekkelijk vindt selecteert u die optie. De linkse kolom wordt nu aangevuld. Daarna kiest u in welke rij u de mogelijkheid rechts aantrekkelijker vindt dan de mogelijkheden in het midden en links. De computer vult dan de rest van de keuzes aan. U selecteert dus eerst een rij in de linkse kolom en daarna in de middelste of rechtse kolom. Controleert u daarna of u inderdaad die keuzes zou maken.




Figure B.5: Example of the R-range task for L1.

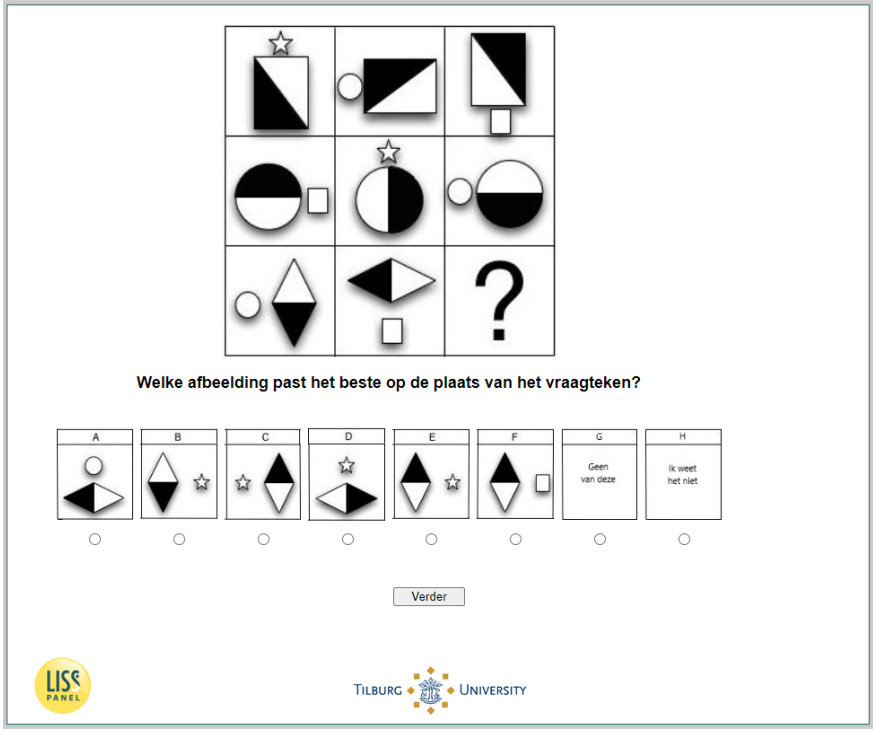


Figure B.6: Example of the matrix reasoning questions.

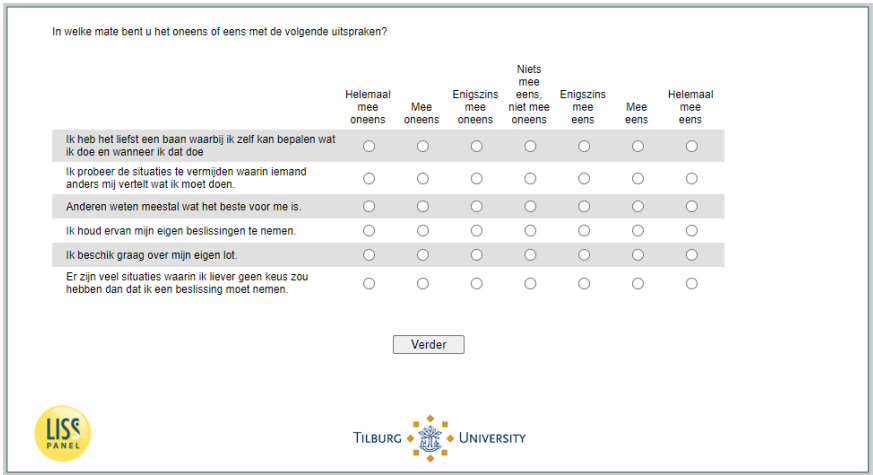
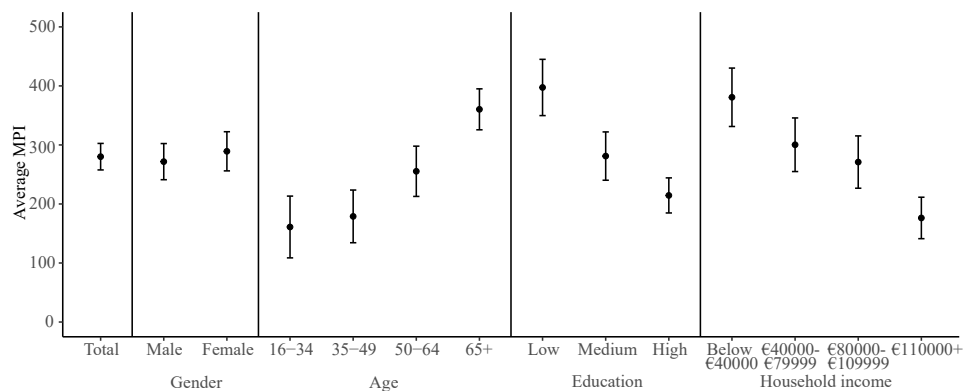


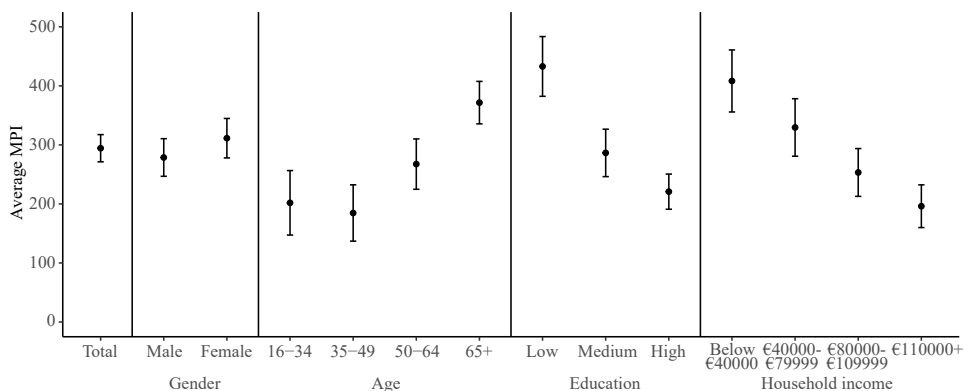
Figure B.7: Desirability of Control statements.

Appendices to chapter 4

C.1 Additional tables and figures



(a) Lottery 1



(b) Lottery 2

Figure C.1: Mean MPI for each lottery in each demographic group. Error bars present 95 percent confidence intervals. The four categories of gross yearly household income are relatively equally distributed: below 40000 euros, 40000-79999 euros, 80000-109999 euros, and 110000 and above.

		Average MPI	
Cognitive ability		−51.05***	(10.38)
Female		20.23	25.83
		(20.64)	(20.52)
Age	35 - 49	18.97	17.43
		(33.97)	(33.30)
	50 - 64	84.70**	72.14**
		(33.62)	(32.93)
	65+	164.47***	149.61***
		(41.95)	(40.37)
Education	Medium	−69.74**	−52.51*
		(30.98)	(28.13)
	High	−124.98***	−96.66***
		(30.29)	(30.36)
Gross household income	40000 - 79999	−43.55	−40.79
		(30.43)	(31.10)
	80000-109999	−39.88	−34.71
		(32.95)	(32.35)
	110000 and above	−98.22**	−87.49***
		(33.16)	(32.23)
Occupation	Housekeeping	12.60	0.26
		(45.44)	(43.88)
	Retired	−23.67	−33.59
		(39.68)	(38.65)
	Others	−1.97	−7.25
		(36.46)	(35.17)
Household composition	Partner	−19.09	−17.26
		(22.71)	(22.41)
	Children at home	7.85	4.14
		(24.76)	(24.07)
Constant		318.71***	309.76***
		(41.72)	(44.32)
Adjusted R^2		0.153	0.187
Observations		571	571

Table C.1: Results of the linear regressions with the average MPI across lotteries as the dependent variable and socioeconomic variables (column 1) and additionally cognitive ability (column 2) as explanatory variables. The variable of cognitive ability is standardized. The benchmark (omitted) categories are male, low education, below 40000, paid work, no partner, and no children at home respectively. The significance *, **, and *** denotes $p < 0.10$, $p < 0.05$, and $p < 0.01$, respectively.

	Log of household wealth					
	Total sample	Household head	Excluding:			
			Wealth outliers	Dominated choosers	Younger than 35	Retirees
MPI	−0.157** (0.073)	−0.251*** (0.083)	−0.146** (0.073)	−0.200** (0.089)	−0.160** (0.075)	−0.182* (0.104)
Female	−0.184 (0.131)	−0.296* (0.169)	−0.157 (0.130)	−0.048 (0.134)	−0.176 (0.141)	−0.050 (0.160)
Age	−0.141 (0.121)	−0.096 (0.139)	−0.123 (0.124)	−0.149 (0.124)	−0.389 (0.247)	−0.061 (0.138)
Age ²	0.004 (0.002)	0.002 (0.003)	0.003 (0.002)	0.004 (0.002)	0.007* (0.004)	0.001 (0.003)
Age ³	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000* (0.000)	0.000 (0.000)
Education						
Medium	0.523*** (0.200)	0.449* (0.241)	0.548*** (0.201)	0.372* (0.216)	0.503** (0.207)	0.773*** (0.286)
High	0.892*** (0.186)	0.933*** (0.221)	0.922*** (0.187)	0.749*** (0.197)	0.897*** (0.191)	1.135*** (0.275)
Log of household income	1.164*** (0.145)	1.100*** (0.162)	1.069*** (0.149)	1.135*** (0.144)	1.082*** (0.162)	1.240*** (0.166)
Occupation						
Housekeeping	0.587** (0.280)	1.042* (0.594)	0.544* (0.278)	0.256 (0.295)	0.486* (0.278)	0.544* (0.314)
Retired	0.249 (0.233)	0.566** (0.278)	0.182 (0.232)	0.139 (0.242)	0.117 (0.241)	
Others	−0.228 (0.308)	−0.011 (0.328)	−0.293 (0.310)	−0.082 (0.315)	−0.416 (0.322)	−0.129 (0.314)
Household composition						
Partner	0.453*** (0.162)	0.518*** (0.174)	0.406*** (0.163)	0.467*** (0.163)	0.493*** (0.172)	0.516** (0.225)
Children	0.107 (0.160)	0.079 (0.199)	0.152 (0.159)	0.106 (0.163)	0.067 (0.174)	0.157 (0.172)
Constant	−1.448 (2.239)	−1.240 (2.660)	−0.846 (2.229)	−1.085 (2.306)	4.612 (4.590)	−3.703 (2.583)
Adjusted R^2	0.385	0.399	0.380	0.374	0.344	0.456
Observations	511	374	505	443	468	305

Table C.2: Results of the OLS regression model of the log of household wealth on the average MPI across the two lotteries. MPI is standardized. Socioeconomic background variables are included as control variables. The omitted categories are male, low education, paid work, no partner, and no children respectively. Age is included as a flexible function. The significance *, **, and *** denotes $p < 0.10$, $p < 0.05$, and $p < 0.01$, respectively.

	Log of household wealth					
	Total sample	Household head	Excluding:			
			Wealth outliers	Dominated choosers	Younger than 35	Retirees
Lottery 1						
MPI	−0.143** (0.068)	−0.232*** (0.082)	−0.125* (0.072)	−0.190** (0.082)	−0.134* (0.074)	−0.146* (0.101)
Controls	Yes	Yes	Yes	Yes	Yes	
Adjusted R^2	0.384	0.396	0.379	0.377	0.342	0.454
Observations	511	374	505	466	468	305
Lottery 2						
MPI	−0.136* (0.065)	−0.251*** (0.083)	−0.146** (0.073)	−0.200** (0.089)	−0.160** (0.075)	−0.182* (0.104)
Controls	Yes	Yes	Yes	Yes	Yes	
Adjusted R^2	0.383	0.399	0.380	0.374	0.344	0.456
Observations	511	374	505	443	468	305

Table C.3: Results of the OLS regressions of the log of household wealth on MPI for each of the two lotteries. MPI is standardized. Socioeconomic background variables are included as control variables. The omitted categories are male, low education, paid work, no partner, and no children respectively. Age is included as a flexible function. The significance *, **, and *** denotes $p < 0.10$, $p < 0.05$, and $p < 0.01$, respectively.

	Log of household wealth			
	(1)	(2)	(3)	(4)
MPI	-0.157** (0.073)	-0.132* (0.074)	-0.198** (0.078)	-0.174** (0.079)
Cognitive ability		0.128** (0.063)		0.132** (0.063)
Risk attitude			0.130** (0.062)	0.134** (0.062)
Female	-0.184 (0.131)	-0.196 (0.131)	-0.176 (0.130)	-0.189 (0.130)
Age	-0.141 (0.121)	-0.147 (0.122)	-0.136 (0.122)	-0.143 (0.123)
Age ²	0.004 (0.002)	0.003 (0.002)	0.003 (0.002)	0.003 (0.002)
Age ³	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Education				
Medium	0.523*** (0.200)	0.490** (0.199)	0.499** (0.201)	0.464** (0.200)
High	0.892*** (0.186)	0.836*** (0.188)	0.897*** (0.186)	0.840*** (0.187)
Log of household income	1.164*** (0.145)	1.159*** (0.146)	1.145*** (0.146)	1.140*** (0.147)
Occupation				
Housekeeping	0.587** (0.280)	0.623** (0.279)	0.597** (0.278)	0.634** (0.276)
Retired	0.249 (0.233)	0.267 (0.231)	0.286 (0.231)	0.306 (0.230)
Others	-0.228 (0.308)	-0.210 (0.303)	-0.207 (0.307)	-0.187 (0.302)
Household composition				
Partner	0.453*** (0.162)	0.448*** (0.162)	0.492*** (0.163)	0.487*** (0.162)
Children	0.107 (0.160)	0.130 (0.159)	0.149 (0.157)	0.174 (0.156)
Constant	-1.448 (2.239)	-1.307 (2.264)	-1.337 (2.234)	-1.188 (2.260)
Adjusted R ²	0.385	0.388	0.389	0.393
Observations	511	511	511	511

Table C.4: Results of the OLS regression model of the log of household wealth on the average MPI across the two lotteries. Columns 2 and 3 include in addition to the baseline specification cognitive ability and risk attitude, respectively, and column 4 includes both. MPI, cognitive ability, and risk attitude are all standardized. Socioeconomic background variables are included as control variables. The omitted categories are male, low education, paid work, no partner, and no children respectively. Age is included as a flexible function. The significance *, **, and *** denotes $p < 0.10$, $p < 0.05$, and $p < 0.01$, respectively.

Log of household wealth				
	(1)	(2)	(3)	(4)
Lottery 1				
MPI	-0.143** (0.064)	-0.120* (0.065)	-0.180** (0.066)	-0.158** (0.072)
Cognitive ability		0.131** (0.063)		0.138** (0.062)
Risk attitude			0.122** (0.065)	0.129* (0.066)
Controls	Yes	Yes	Yes	Yes
Adjusted R^2	0.384	0.388	0.387	0.392
Observations	511	511	511	511
Lottery 2				
MPI	-0.136* (0.072)	-0.114 (0.073)	-0.158** (0.075)	-0.135* (0.076)
Cognitive ability		0.134** (0.063)		0.136** (0.063)
Risk attitude			0.102 (0.067)	0.104 (0.067)
Controls	Yes	Yes	Yes	Yes
Adjusted R^2	0.383	0.387	0.385	0.390
Observations	511	511	511	511

Table C.5: Results of the OLS regression model of the log of household wealth on MPI for each of the two lotteries. Columns 2 and 3 include in addition to the baseline specification cognitive ability and risk attitude, respectively, and column 4 includes both. MPI, cognitive ability, and risk attitude are all standardized. Socioeconomic background variables are included as control variables. The omitted categories are male, low education, paid work, no partner, and no children respectively. Age is included as a flexible function. The significance *, **, and *** denotes $p < 0.10$, $p < 0.05$, and $p < 0.01$, respectively.

C.2 Types of choice inconsistency

Table C.6 summarises the proportions of different choice patterns in the experiment. Figure C.2 shows the mean MPI for each category of subjects based on the type of inconsistency. From these means we can see that types of inconsistency do not rank well the economic cost of choice inconsistency. There exists substantial heterogeneity in the economic cost of the type of inconsistency, as evidenced by the standard deviations of MPI within each category. This illustrates the challenge of categorizing subjects' based on the type of choice inconsistency.

L1					
	No MSB and no MR	Only MSB	Only MR	Mix of MSB and MR	Total
No dominated	18.88%	12.06%	11.71%	48.78%	91.14%
	0	160.33	417.40	353.91	263.42
	(0)	(151.68)	(314.85)	(213.93)	(251.17)
Dominated	3.32%	1.75%	0%	3.50%	8.57%
	0	585.00	-	830.84	458.51
	(0)	(249.50)	-	(173.33)	(409.73)
L2					
	No MSB and no MR	Only MSB	Only MR	Mix of MSB and MR	Total
No dominated	19.06%	11.19%	12.41%	48.25%	90.91%
	0	119.73	401.72	380.15	271.36
	(0)	(106.85)	(312.94)	(214.53)	(256.84)
Dominated	2.80% ⁹	1.92%	0%	4.37%	9.09%
	0	563.64	-	843.68	524.84
	(0)	(214.58)	-	(96.45)	(387.28)

Table C.6: Proportions (of the 572 subjects), mean MPI, and standard deviation of MPI of each choice pattern for the respective lottery. MSB stands for multiple switching behavior between the lottery and the sure payment. MR stands for choosing the randomization option more than once. The numbers with % are proportions, and the numbers in parentheses are standard deviations of MPI. The cell "Dominated, No MSB and No MR" includes subjects who selected only the lottery or only the sure payment for all decisions, which results in choosing one of the dominated options.

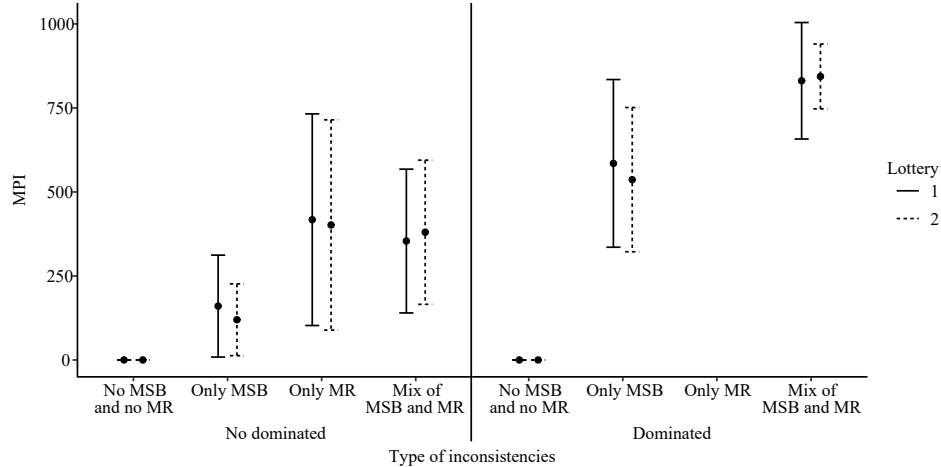


Figure C.2: Mean MPI and corresponding standard deviations for each type of choice inconsistency. The error bars represent and standard deviation above and below the mean.

C.3 Using self-reported household wealth

We obtain the measure of household wealth based on data about housing, financial, and other assets. For this we obtain responses to two longitudinal studies from the LISS panel. The first asks about the housing assets and is fielded yearly. In this questionnaire, subjects are asked to answer questions about the house they are living in: If they own it or rent it, what the current price would be if they were to sell, what the remaining debt of the mortgage(s) is, etc. The second asks about all other assets and is fielded every two years. In this questionnaire, subjects are first asked to indicate which type of asset they possess (e.g. savings, financial investments, real-estate) and subsequently how much each indicated asset is worth.²⁷

As the household data contains a lot of missing data we construct three definitions of household data to trade off between the quality of the data and the number of observations.

1. The first definition of household wealth excludes the top and bottom 1% of subjects based on household wealth.
2. The second definition, additionally, excludes households in which a member did not participate in the asset questionnaire.
3. The final definition includes only households that do not have any missing wealth data.

Table C.8 reports the results of the regressions with household wealth as the dependent variable and the average MPI across lotteries as the main explanatory variable. We control for socioeconomic characteristics. The results show that, for each definition of household wealth, a higher money pumping potential is related to lower household wealth. Importantly, as remarked in the main text, the coefficient of MPI estimated using self-reported household wealth is about twice the size of the coefficient estimated using administrative records.

Since the use of self-reported data is still prevalent, below we provide some insights into how self-reported data may lead to different results from administrative records. One evidence has been shown in Figure 21. To more formally test the possible interaction between reporting bias and MPI, we created a binary variable indicating if subjects' self-reported wealth falls within the range of 10% above or 10% below the administrative record of household wealth. The results in Table C.9 show that subjects with a higher MPI are less likely to accurately report their household wealth. Specifically, one standard deviation increase in MPI decreases the odds of reporting one's household wealth within the accurate bands of 10% by 37.4%. This suggests that the biased estimation of the association between MPI and self-reported household wealth could partially be driven by the interaction between MPI and the biased reporting of household wealth. One intuitive explanation is that subjects

²⁷For each study, we use the questionnaire fielded closest to the moment of our data collection, which provides us with data on housing for 2020 and data on assets at the end of 2019. Our experimental data was collected in September 2021.

	Complete sample	Household wealth		
		Definition 1	Definition 2	Definition 3
Sample size	620	609	416	286
Female	48.06	48.77	48.80	50.35
Age				
16-34	12.42	12.48	14.90	14.33
35-49	15.48	15.60	14.42	8.04
50-64	25.32	25.45	22.12	19.93
65 and above	46.77	46.47	48.58	57.69
Education level				
Low	24.52	24.79	25.24	26.92
Medium	30.16	30.38	30.05	29.02
High	45.16	44.66	44.47	44.06
Household income				
0-2499	21.77	21.84	27.40	33.22
2500-3999	26.61	27.09	27.88	30.07
4000-5999	23.06	23.15	20.67	19.23
6000 and above	24.84	24.14	20.67	15.38
Occupation				
Paid work	42.90	42.53	41.83	32.87
Housekeeping	7.26	7.39	6.73	8.04
Retired	38.87	38.92	40.38	48.25
Others	10.97	11.17	11.06	10.84
Household composition				
Partner	64.19	63.88	56.49	47.20
# of children	0.41	0.39	0.26	0.03

Table C.7: Sample distribution

with a higher MPI may have a worse overview of their financial situation and are therefore more likely to make errors in their self-reports.

	Household wealth		
	Definition 1	Definition 2	Definition 3
Exploitation potential	−0.315** (0.127)	−0.520*** (0.164)	−0.627*** (0.173)
Female	−0.154 (0.246)	−0.180 (0.315)	−0.101 (0.338)
Log age	2.770*** (0.527)	3.510*** (0.675)	3.793*** (0.736)
Education level			
Medium	0.266 (0.335)	0.297 (0.432)	0.193 (0.445)
High	1.355*** (0.331)	1.545*** (0.430)	1.425*** (0.449)
Log household income	0.600*** (0.117)	0.476*** (0.132)	0.157 (0.147)
Occupation			
Housekeeping	−1.001** (0.507)	−0.676 (0.688)	−0.913 (0.714)
Retired	−0.415 (0.363)	−0.772 (0.472)	−0.567 (0.513)
Others	−0.723 (0.456)	−798 (0.594)	−1.208* (0.665)
Household composition			
Partner	1.017*** (0.262)	0.865*** (0.332)	1.205*** (0.354)
# of children	−0.003 (0.162)	0.261 (0.254)	−0.631 (0.782)
Constant	−6.208*** (2.124)	−8.180*** (2.636)	−6.717** (2.814)
Adjusted R^2	0.237	0.242	0.269
Observations	505	339	248

Table C.8: Results of the regression models where the log of household wealth is the dependent variable and the average money pumping potential across the two lotteries is the independent variable (the results for the money pumping range and for the two measures for each lottery separately can be found in Appendix C.1). Gender, log age, log household income. The significance *, **, and *** denotes $p < 0.10$, $p < 0.05$, and $p < 0.01$, respectively. We run the regression for each of the three definitions of household wealth. The omitted categories are no dominated choice in L1, no dominated choice in L2, male, and low education respectively.

	Accurate self-reported household wealth (0.90 - 1.10)
MPI	-0.468** (0.227)
Controls	Yes
Nagelkerke R^2	0.226
Observations	265

Table C.9: Results of the logistic regressions of the binary variable indicating accurate self-reported household wealth on the average MPI across lotteries. MPI is standardized. Controls are the same as in Table 8. We restrict the sample to those who meet the criteria of the most restrictive definition of self-reported household wealth. The significance *, **, and *** denotes $p < 0.10$, $p < 0.05$, and $p < 0.01$, respectively.

C.4 Exploring the channels underlying the association between MPI and household wealth

Below, we shed some light on the underlying mechanisms that contribute to the observed association between overall household wealth and MPI. First, we look at the different components of household wealth and their association with MPI. We categorize household wealth into three categories: housing assets, investment assets, and bank assets. Housing assets include only the house that the household is living in, whereas other real-estate holdings are included in the category of investments assets together with financial investments, business assets, and other investments. Finally, bank assets are limited to the household's bank savings.

For the top panel of Table C.10, we created for each category of wealth a binary outcome variable. For housing and investment assets, this binary variable indicates whether the household has any assets that fall into the respective category. For bank assets, the binary variable indicates whether all the assets of the household are limited to this category (since 99.7% of households have bank assets and a binary variable similar to the other two categories would be meaningless).

We ran a logistic regression of the respective binary variable for each category on MPI while controlling for socioeconomic variables. The results show that one standard deviation increase in MPI increases the odds of only holding bank assets by 29.5%, whereas it decreases the odds of having housing assets by 21.3%. For investment assets, the coefficient is also negative but statistically insignificant. In the bottom panel of Table C.10, we look at the same three sources of household wealth. Instead of focusing on whether households have certain assets, we look at the fraction of each category of wealth within the total wealth of the household. We find a similar trend, one standard deviation increase in MPI increases the fraction of bank assets by 0.045 and decreases the fraction of housing assets by 0.031.

Next, we look at household income, which is a different economic outcome, but closely related to household wealth. In section 4.3.1 we already established that MPI is negatively correlated to household income. In the top panel of Table C.11 we elaborate on this by regressing the log of gross household income on MPI while controlling for other socioeconomic controls. We find a significant negative association between MPI and household income, although the estimate is smaller than what we find for the association with household wealth, with one standard deviation increase in MPI corresponding to a decrease in gross household income of 3.5%.²⁸

Since the gross household income measure includes all types of income, such as income from labor, capital, and other sources (e.g., social security subsidies, other forms of transfers), this may create noise in our investigation into the correspondence between income and decision-making quality. To overcome this potential source of noise and to investigate

²⁸We perform the regressions for gross household income in 2020 and for the average across 2016 to 2020, since income can be volatile across years. We try to control for changes in household composition by restricting the sample to those subjects who have the same household identifier across the years as we did for Figure 20.

	Housing assets (binary)	Investment assets (binary)	Only bank assets (binary)
MPI	-0.240** (0.121)	-0.135 (0.112)	0.258** (0.130)
Controls	Yes	Yes	Yes
Nagelkerke R^2	0.406	0.209	0.399
Observations	561	561	561
	Fraction housing assets (1 – 99%)	Fraction investment assets (1 – 99%)	Fraction bank assets (1 – 99%)
MPI	-0.031* (0.017)	-0.007 (0.008)	0.045** (0.019)
Controls	Yes	Yes	Yes
Adjusted R^2	0.165	0.031	0.094
Observations	548	550	553

Table C.10: Top panel: Results of the logistic regressions of the binary indicator of having each of the subcategories of household wealth on the average MPI across the two lotteries. The binary indicator for bank assets indicates whether this is the only type of assets the household holds. Bottom panel: Results of the OLS regressions of the ratio of the subcategories of household wealth to total household wealth on the average MPI across the two lotteries. The socioeconomic background variables are the same as those in Table 8. MPI, cognitive ability, and risk attitude are all standardized. The significance *, **, and *** denotes $p < 0.10$, $p < 0.05$, and $p < 0.01$, respectively.

in more detail the mechanism that underlies the association between MPI and household income, we look at two sources of income that are of particular interest when it comes to their link to decision-making quality. Specifically, we look at income from labor and income from capital. Bastani et al. (2023) show, using military enlistment data and administrative records on income, that cognitive ability is better at predicting capital income compared to labor income later in life. Following this distinction between types of income, we show in the bottom panel of C.11 similar results for the association between the two types of income and MPI. For income from labor, we find no significant association between income and MPI. In contrast, for income from capital we find a significant and large association with MPI, with one standard deviation increase in MPI corresponding to a decrease in household income from capital of 26.5%. This points to the idea that more consistency is associated with more income from capital, which naturally relates to the overall level of wealth, both in terms of higher wealth accumulation and as a source of income.

	Gross household income		Household income from labor		Household income from capital	
	2020	Average 2016-2020	2020	Average 2016-2020	2020	Average 2016-2020
MPI	-0.036** (0.018)	-0.035* (0.019)	-0.174 (0.159)	-0.156 (0.182)	-0.345** (0.163)	-0.308** (0.138)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.552	0.574	0.679	0.583	0.233	0.260
Observations	499	467	499	467	498	467

Table C.11: Results of the OLS regressions of different household income measures on the average MPI across the two lotteries. The socioeconomic background variables are the same as those in Table 8. MPI, cognitive ability, and risk attitude are all standardized. The significance *, **, and *** denotes $p < 0.10$, $p < 0.05$, and $p < 0.01$, respectively.

C.5 Experimental materials

In deze vragenlijst zijn we geïnteresseerd in uw voorkeuren, dit betekent dat er geen goede of foute keuzes zijn. Naast de vergoeding die u ontvangt voor het invullen van de vragenlijst kunt u een extra vergoeding verdienen. Deze extra vergoeding is afhankelijk van de keuzes die u maakt. **Kies daarom altijd de optie die werkelijk uw voorkeur heeft!**

Nadat alle ingevulde vragenlijsten zijn verwerkt, kiest de computer willekeurig een aantal panelleden uit die een extra vergoeding ontvangen op basis van uw keuzes. De kans dat u wordt gekozen voor deze extra vergoeding is 1 op 10. Voor de extra vergoeding wordt willekeurig 1 van de keuzes die u heeft gemaakt geselecteerd. Omdat achteraf pas duidelijk is of u naast de normale vergoeding ook een extra vergoeding krijgt, wordt de vergoeding achteraf bepaald.

In de vragenlijst gebruiken we punten in plaats van euro's: **100 punten = €0.50.**

Let op: U kunt in deze vragenlijst niet terug. Telkens wanneer u antwoord hebt gegeven en u op 'Verder' klikt, is uw antwoord dus definitief. Klikk u pas op 'Verder' wanneer u zeker bent van uw antwoord.

In de meeste vragen in deze vragenlijst moet u een keuze maken tussen een loterij waarin u punten kunt winnen of een aantal punten dat u zeker ontvangt.

De loterijen worden beschreven als de kans dat u een bepaald aantal punten ontvangt.

Bijvoorbeeld: een loterij waarin u **50% kans** heeft dat u **400 punten** ontvangt en **50% kans** dat u **0 punten** ontvangt.

U kunt zich deze loterij voorstellen als een computer die willekeurig een nummer van 1 tot 100 kiest. U ontvangt 0 punten als een nummer van 1 tot 50 wordt gekozen en u ontvangt 400 punten als een nummer van 51 tot 100 wordt gekozen.

Figure C.3: Introduction to the experiment.

U begint nu aan deel 1.

U maakt deze keuze steeds voor 20 verschillende aantallen punten. De loterij blijft steeds hetzelfde.

In dit deel maakt u steeds keuzes tussen een loterij en een aantal punten. Bij elke keuze hebt u drie mogelijkheden:

1. U kiest de loterij.
2. U laat het de computer bepalen door het opgooien van een virtuele munt: bij kop krijgt u de loterij en bij munt het aantal punten.
3. U kiest het aantal punten.

Het is belangrijk dat u iedere keuze als een opzichzelfstaande beslissing ziet.

Extra vergoeding
Als dit deel wordt gekozen voor de extra vergoeding, zal 1 van uw keuzes willekeurig worden gekozen voor de extra vergoeding.

Verder



De komende vragen gaan over de volgende loterij:

- 50% kans op 900 punten
- 50% kans op 100 punten

Verder



De loterij in deze vraag geeft u:

- 50% kans op 900 punten
- en 50% kans op 100 punten.

U hebt 3 mogelijkheden:

Kies de loterij De computer bepaalt Kies de 100 punten

☒ ☐ ☐

Verder



Figure C.4: Example of a question from the choice inconsistency task for L1.

U begint nu aan deel 1.

U maakt deze keuze steeds voor 20 verschillende aantallen punten. De loterij blijft steeds hetzelfde.

In dit deel maakt u steeds keuzes tussen een loterij en een aantal punten. Bij elke keuze hebt u drie mogelijkheden:

1. U kiest de loterij.
2. U laat het de computer bepalen door het opgooien van een virtuele munt: bij kop krijgt u de loterij en bij munt het aantal punten.
3. U kiest het aantal punten.

Het is belangrijk dat u iedere keuze als een opzichzelfstaande beslissing ziet.

Extra vergoeding
Als dit deel wordt gekozen voor de extra vergoeding, zal 1 van uw keuzes willekeurig worden gekozen voor de extra vergoeding.

Verder

De komende vragen gaan over de volgende loterij:

- 20% kans op 975 punten,
- 30% kans op 850 punten,
- 30% kans op 150 punten,
- en 20% kans op 25 punten.

Verder

De loterij in deze vraag geeft u:

- 20% kans op 975 punten,
- 30% kans op 850 punten,
- 30% kans op 150 punten,
- en 20% kans op 25 punten.

U hebt 3 mogelijkheden:

Kies de loterij De computer bepaalt Kies de 425 punten

☐ ☒ ☐

Verder

Figure C.5: Example of a question from the choice inconsistency task for L2.

U begint nu aan deel 2.

In dit deel moet u een aantal puzzels oplossen. Voor ieder goed antwoord ontvangt u 100 punten.

Extra vergoeding
Als dit deel wordt gekozen voor de extra vergoeding, ontvangt u 100 punten maal het aantal goede antwoorden.

Verder



Welke afbeelding past het beste op de plaats van het vraagteken?

A	B	C	D	E	F	G	H
							
☐	☐	☐	☐	☐	☐	☐	☐

Verder



Figure C.6: Example of a matrix reasoning question.

U begint nu aan deel 3.

In dit deel maakt u in iedere vraag een keuze tussen een loterij of een bepaald aantal punten.

U maakt deze keuze steeds voor 20 verschillende aantallen punten. De loterij blijft steeds hetzelfde.

Om u te helpen de vragen efficiënt te beantwoorden vult de computer uw keuzes aan als u in één van de rijen de loterij of het aantal punten kiest. De beste manier om uw keuze te maken is om bij de eerste rij te beginnen. Zolang u de linkerkant aantrekkelijker vindt dan de rechterkant gaat u steeds een rij verder. Als u de rechterkant aantrekkelijker vindt dan de linkerkant kiest u in die rij de rechterkant. De computer vult dat de rest van de keuzes aan. Controleert u daarna of u inderdaad die keuzes zou maken.

Het is belangrijk dat u iedere keuze als een opzichzelfstaande beslissing ziet.

Extra vergoeding
Als dit deel wordt gekozen voor de extra vergoeding, zal 1 van uw keuzes willekeurig worden gekozen voor de extra vergoeding.

Verder



De keuze(s) in deze vraag geeft u:

- 50% kans op 500 punten
- en 50% kans op 100 punten

U heeft steeds 2 mogelijkheden:

1. U kiest de loterij

2. U kiest een bepaald aantal punten

Deel 3: Hierbij gaat het om een voor ieder aantal punten kiest. Het aantal punten neemt in iedere keuze toe. De loterij blijft steeds hetzelfde. De computer vult automatisch de keuzes aan als u in een rij een keuze maakt. Controleert u daarna of u voor de overige rijen inderdaad die keuzes zou maken.

(om de loterij ingevuld te krijgen kunt u in iedere keuze met een muis over 'loterij' bewegen)

	Optie 1	Optie 2
☐	Kies de loterij	Kies 0 punten
☒	Kies de loterij	Kies 100 punten
☒	Kies de loterij	Kies 200 punten
☒	Kies de loterij	Kies 300 punten
☒	Kies de loterij	Kies 400 punten
☒	Kies de loterij	Kies 500 punten
☒	Kies de loterij	Kies 600 punten
☒	Kies de loterij	Kies 700 punten
☒	Kies de loterij	Kies 800 punten
☒	Kies de loterij	Kies 900 punten
☒	Kies de loterij	Kies 1000 punten
☐	Kies de loterij	Kies 1100 punten
☐	Kies de loterij	Kies 1200 punten
☐	Kies de loterij	Kies 1300 punten
☐	Kies de loterij	Kies 1400 punten
☐	Kies de loterij	Kies 1500 punten
☐	Kies de loterij	Kies 1600 punten
☐	Kies de loterij	Kies 1700 punten
☐	Kies de loterij	Kies 1800 punten
☐	Kies de loterij	Kies 1900 punten
☐	Kies de loterij	Kies 2000 punten

De beste manier om uw keuze te maken is om bij de eerste rij te beginnen. Zolang u de linkerkant aantrekkelijker vindt dan de rechterkant gaat u steeds een rij verder. Als u de rechterkant aantrekkelijker vindt dan de linkerkant kiest u in die rij de rechterkant. De computer vult dan de rest van de keuzes aan. Controleert u daarna of u inderdaad die keuzes zou maken.

Verder



Figure C.7: Risk attitude task for L1.

Acknowledgements

Completing this dissertation has been a long and rewarding journey, and it would not have been possible without the support, guidance, and encouragement of many people.

First and foremost, I am especially grateful to my co-promotor and daily supervisor, **Jianying Qiu**. Thank you for your continuous support, thoughtful feedback, and close involvement in my research. I greatly appreciate the many discussions we had, your patience in reading and commenting on numerous versions of my work, and your guidance during both the challenging and rewarding stages of the PhD.

Next, I would like to express my sincere gratitude to my promotor, **Stefan Zeisberger**. While your guidance was from a bit more distance academically, I knew I could always approach you if I were in need for advice or support. Furthermore, the feedback and encouragement I received were of great value.

I would also like to thank the members of my dissertation committee, **Etiënne Rouwette, Nicolas Jacquemet, and Zhihua Li**, for taking the time to read and evaluate my dissertation. My gratitude also extends to the members of the Examination Board, **Kim Fairley, and Binglin Gong**, for their time, expertise, and contribution to the final stage of this doctoral journey.

I am grateful to all my colleagues at the **Department of Economics and Business Economics**, and particularly to my colleagues within the **Chair of Finance**. Thank you for creating a stimulating and supportive working environment. The conversations during IdeaLabs, seminars, lunches, coffee breaks, board game nights, and drinks at the Yard contributed not only to my research but also to making daily life at the department enjoyable. I have greatly appreciated your willingness to exchange ideas, offer advice, and share both academic and non-academic moments.

A special word of thanks goes to my co-authors, **Qiyang Ong, Jana Vyrastekova, and Songfa Zhong**. Working with you has been an invaluable learning experience. Thank you for sharing your knowledge, challenging my ideas, and contributing your time and expertise to our joint projects. I have learned a great deal from our collaboration and from the many discussions that shaped our work.

I would also like to thank my office mate, **Ludger Buijs**, for the companionship, conversations, and welcome distractions during the working day. Being able to work in silence but also have the necessary coffee or tea breaks made for a very pleasant work environment.

My thanks also go to the **other PhD students** at the Institute for Management Research. It was reassuring and motivating to share this experience with people who understood the particular joys and frustrations of working towards a PhD. Thank you for the discussions, advice, coffee breaks, social activities, and support. Both the organized meetings, seminars, and writing retreats as well as the impromptu dinners or drinks provided a valuable sense of community throughout the PhD.

Beyond the university, I am deeply grateful to my friends. Thank you for your encouragement, understanding, and interest in my work, but also for reminding me that there is a world outside the dissertation. Your friendship and ability to provide distraction when needed have meant a great deal to me.

Finally, I would like to thank my family. **Pap en mam**, bedankt voor jullie hulp, interesse en trots, zowel tijdens mijn PhD en daarvoor. Zonder jullie had ik het zeker niet zover geschopt! **Jochem**, lieve schat, ook jij enorm bedankt voor je steun tijdens de afgelopen jaren en alle leuke, mooie momenten daarbuiten die zo nu en dan voor een welkome afleiding zorgden. En als laatste lieve **Isa en Jake**, jullie zijn voor mij de grootste inspiratie (ondanks de gebroken nachten en het slaapttekort).

And to anyone who I might have forgotten and supported me along the way: **Thank you!**

Nederlandse Samenvatting

Mensen nemen elke dag talloze beslissingen, van het kiezen van boodschappen in de supermarkt tot het afsluiten van een hypotheek of pensioenplan. Zulke keuzes vragen vaak om het beoordelen van risico's, het vergelijken van opties, het maken van lastige afwegingen en het omgaan met onzekerheid. Economen proberen al lang te begrijpen hoe mensen dergelijke beslissingen nemen en gaan er doorgaans van uit dat mensen, gegeven bepaalde beperkingen, opties kunnen vergelijken en de beste optie kunnen kiezen. In de praktijk geven mensen echter vaak aan weinig vertrouwen te hebben in hun keuzes, blijven ze uit voorzichtigheid bij standaardopties, veranderen ze van mening of stellen ze beslissingen uit. Dit proefschrift bevat drie hoofdstukken over één belangrijke factor die het beslisgedrag beïnvloedt: beslissingscomplexiteit.

Het centrale idee is dat veel beslissingen moeilijk zijn. Mensen begrijpen de keuzeomgeving soms niet volledig (bijv. wat precies de implicaties van een belastinghervorming zijn). Of ze voelen zich onzeker over de waarde die ze aan opties toekennen (bijv. bij complexe financiële producten). Dergelijke vormen van complexiteit kunnen ertoe leiden dat mensen heuristieken (beslisregels) gebruiken, die tot systematische biases (vertekeningen) en beslisfouten kunnen leiden. Dit proefschrift onderzoekt de rol van complexiteit in drie stappen, elk in een afzonderlijk hoofdstuk.

Hoofdstuk twee stelt een simpele vraag: hoe krijgen we informatie over het vertrouwen dat mensen hebben in de keuzes die zij maken? Beslissingsvertrouwen heeft de potentie om uiteenlopende systematische biases te verklaren, maar is subjectief en niet direct te observeren. Traditioneel wordt vertrouwen gemeten met niet-geïncentiveerde zelfrapporten. Dit hoofdstuk introduceert een nieuwe, geïncentiveerde methode: deelnemers mogen “randomiseren” tussen opties. Concreet: wanneer ze kiezen tussen twee opties, wijzen deelnemers een randomisatiekans toe aan elke optie. Die randomisatiekansen bleken sterk hun zelfgerapporteerde beslissingsvertrouwen te weerspiegelen. Wanneer mensen onzeker waren, randomiseerden ze meer; wanneer ze zeker waren, kozen ze vaker één optie. Deze methode biedt een robuuste, geïncentiveerde manier om beslissingsvertrouwen te meten.

Hoofdstuk drie draagt bij aan een klassieke puzzel in de economie: de kloof tussen willingness-to-accept (WTA) en willingness-to-pay (WTP). Mensen vragen vaak meer geld om iets af te staan dat ze bezitten (WTA) dan ze bereid zijn te betalen om het te verkrijgen (WTP). Deze kloof is lang gezien als bewijs voor verliesaversie. Recente literatuur laat echter zien dat de WTA-WTP-kloof complexer is dan gedacht en niet door verliesaversie alleen kan worden verklaard. Dit hoofdstuk laat zien dat complexiteit een cruciale rol speelt. Twee vormen van complexiteit worden onderzocht: mechanism complexity (hoe moeilijk het is de keuzeomgeving te begrijpen) en valuation complexity (hoe onzeker mensen zijn over de waardering van opties). Op basis van een representatieve Nederlandse steekproef

blijkt dat deze twee vormen van complexiteit samen ongeveer 50% van de geobserveerde WTA–WTP-kloof verklaren.

Het laatste hoofdstuk onderzoekt de gevolgen van beslissingscomplexiteit in de praktijk. In experimenten maken veel mensen inconsistente keuzes wanneer zij complexiteit ervaren, maar het is onduidelijk in hoeverre dit doorwerkt in echte uitkomsten. In dit hoofdstuk gebruiken we een methode die het potentiële financiële verlies door inconsistente keuzes kwantificeert. Door deze experimentele methode te koppelen aan gedetailleerde administratieve gegevens van het Centraal Bureau voor de Statistiek (CBS), vinden we een duidelijk patroon: mensen met hogere inconsistentiescores hebben gemiddeld een lager huishoudvermogen. Een toename van één standaarddeviatie in inconsistentie hing samen met bijna 15% minder vermogen. Dit verband blijft bestaan na correctie voor opleiding, inkomen, risicohouding en cognitieve vermogens, wat erop wijst dat hoe mensen in experimenten met complexiteit omgaan, weerspiegeld wordt in hun financiële uitkomsten in het dagelijks leven.

Gezamenlijk laten deze hoofdstukken zien dat complexiteit een belangrijke rol speelt in besluitvorming. Complexiteit beïnvloedt hoe zeker mensen zich voelen over hun keuzes, helpt lang bestaande puzzels in economisch gedrag te verklaren en heeft tastbare effecten op de financiën van huishoudens. Het erkennen van de rol van complexiteit kan onderzoekers helpen betere experimenten te ontwerpen, beleidsmakers helpen effectievere interventies te ontwikkelen en individuen helpen hun eigen beslisproces beter te begrijpen.

English Summary

People need to make many decisions every day, from choosing groceries in the supermarket to deciding on a mortgage or pension plan. These decisions often involve weighing risks, comparing options, trading off conflicting outcomes, and dealing with uncertainty. Economists have long tried to understand how people make such decisions, typically assuming that people, given certain constraints, know how to compare options and choose the best one. Yet in practice, people often express a lack of confidence in their decisions, behave overly cautiously by sticking to the default choices, change their minds about their decisions, or delay the decisions. This dissertation contains three chapters that focus on one important factor that affects decision-making: Decision complexity.

The central idea is that people may find many decisions difficult to make. People may not fully understand the choice environment (e.g. what exactly are the implications of a tax reform). Or they may feel unsure about the value they attach to choice options (e.g. complicated financial products). These forms of complexity can cause people to use decision heuristics, leading to systematic biases and decision errors. This dissertation investigates the role of complexity in decision-making in three steps, each covered in a separate chapter.

Chapter two asks a simple question: how can we obtain information about the confidence that people have in the choices they make? Decision confidence has the potential to explain a range of systematic biases, but it is subjective and cannot be observed directly. Traditionally, confidence is measured by non-incentivized self-reports. This chapter introduces a new behavioral measure that is incentivized: allowing participants to “randomize” between options. Specifically, when choosing between two options, participants assign a randomization probability to each option. The randomization probabilities turned out to strongly reflect their self-reported confidence levels. When people felt uncertain, they randomized more and when they felt sure, they stuck to one option. This new method provides a robust and incentive-compatible way of measuring decision confidence.

Chapter three contributes to a classic puzzle in economics: the willingness-to-accept and willingness-to-pay gap. People often demand more money to give up something they own (WTA) than they are willing to pay to acquire it (WTP). This gap has been seen as evidence for loss aversion. However, a lot of recent research has shown that the WTA-WTP gap is more complex than originally thought and that loss aversion alone cannot explain it. This chapter shows that complexity plays a crucial role. Two forms of complexity were studied: mechanism complexity (how difficult it is to understand the decision environment) and valuation complexity (how uncertain people are about the valuation of the choice options). Using a representative Dutch sample, we find that together these forms of complexity explain about 50% of the observed WTA–WTP gap.

The final chapter examines the real-world consequences of decision complexity. In experimental settings, many individuals make inconsistent choices when perceiving complexity.

However, it is unclear to what extent this translates to real-life outcomes. In this chapter we use a measure that quantifies the potential financial loss from inconsistent choices. Linking this experimental measure to detailed administrative data from Statistics Netherlands (CBS) we find a clear pattern: individuals with higher inconsistency scores tend to have lower household wealth. A one standard deviation increase in inconsistency corresponded to nearly 15% less wealth. This link remained even after accounting for education, income, risk attitudes, and cognitive ability, showing that the way people deal with complexity in experiments reflects real-life financial outcomes.

Together, these chapters highlight that complexity is an important factor in decision-making. It shapes how confident people feel about their decisions, explains long-standing puzzles in economic behavior, and has tangible effects on household finances. Recognizing the role of complexity can help researchers design better experiments, policymakers craft more effective interventions, and individuals better understand their own decision-making.

Research Data Management

The data management in this dissertation complies with the regulations and guidelines described by Radboud University, in accordance with the General Data Protection Regulations (GDPR).

Data were collected from several sources: (1) through experiments involving students from the Individual Decision-Making Lab at Radboud University, (2) through experiments conducted at the Longitudinal Internet studies for the Social Sciences (LISS) panel, (3) existing (background) data from the LISS data archive, and (4) administrative data obtained from Statistics Netherlands (CBS). Participants of our experiments were informed of the data privacy guidelines via consent forms provided at the beginning of each experiment. Existing data from the LISS panel was used in compliance with the LISS data user statement. Similarly, access to the administrative data from Statistics Netherlands was obtained in compliance with guidelines and conditions for the use of the CBS microdata.

Privacy-sensitive data collected during experiments for payment purposes were deleted once administrative tasks were completed. Anonymous experimental research data were used by scientists as part of data sets, articles and presentations and were stored securely on Radboud University's servers, in compliance with university protocol. The anonymized research data will be accessible to other scientists for a period of a minimum of 10 years.

Curriculum Vitae

Sara Arts was born on 11 February 1995 in Boxmeer, the Netherlands. In 2016, she completed a bachelor's degree in Liberal Arts and Sciences at University College Utrecht (Utrecht University), majoring in Economics and Psychology and minoring in Statistics. In 2019, she completed the Research Master's program in Social and Behavioral Sciences at Tilburg University.

Following her Master's, she began her PhD project in the Department of Economics and Business Economics at Radboud University in Nijmegen, under the supervision of Prof. Dr. Stefan Zeisberger and Dr. Jianying Qiu. She currently works as a postdoctoral researcher at the Erasmus School of Health Policy and Management at Erasmus University Rotterdam.