Ranges of Randomization*

Marina Agranov† Pietro Ortoleva‡

Abstract

A growing literature has shown that people sometimes prefer to randomize between two options. We investigate how prevalent this behavior is in an experiment using a novel and simple method. Subjects face a list of questions in which one of the alternatives is fixed and the other varies, like a Multiple Price List, but in each row they can randomize between the options. We find that the majority of subjects chose to randomize in the majority of questions, and notably, they did so for ranges of values that were "very large."

Key words: Preference for Randomization, Incomplete Preferences, Non-Expected Utility.

JEL: C91, D81, D90

*We thank Francesco Bogliacino, Mark Dean, David Dillenberger, Efe Ok, and Leeat Yariv, as well as the editor and four anonymous referees for very useful comments and suggestions. Ortoleva gratefully acknowledges the financial support of NSF Grant SES-1763326.

†Division of the Humanities and Social Science, California Institute of Technology. Email: magranov@hss.caltech.edu.

‡Department of Economics, Princeton University and SPIA. Email: pietro.ortoleva@princeton.edu.

1 Introduction

A growing body of research has demonstrated that individuals often choose to *randomize* between two options. For example, when asked to choose between a sure amount \$$x$ and a lottery that pays \$20 or \$0 with equal chances, for some values of x the individual may prefer a randomization between the two alternatives over either of the two. The choice to randomize has been documented in many environments and connected theoretically to violations of Expected Utility, convexity, and incomplete preferences. (Both the empirical and the theoretical literature are discussed below.)

To our knowledge, however, most studies focused on demonstrating the *existence* of choices in which subjects want to randomize. This paper studies their *prevalence*. For example, in the choice above between \$$x$ and a lottery that pays \$20 or \$0, we ask: How big is the *range* of values of x for which people want to randomize? Is it large or is it small? We propose a novel, simple method to capture this range—a modified Multiple-Price-List—and we apply it to standard economic questions like certainty equivalents and lottery equivalents. Then, we relate our results to choices in standard tasks, Certainty Bias, non-monotonicities, and other individual characteristics.

Understanding the prevalence of preferences for randomization is crucial to assess their significance. If such preferences exist solely on rare, smartly-constructed, knife-edge cases, they may be of theoretical interest but may not have a significant impact on many economic issues. On the other hand, if preferences for randomization are more widespread, affecting a wide range of questions and values, then ignoring them can be problematic: if individuals do prefer to randomize, it should be taken into account in modeling and in interpreting data.

Eliciting Ranges. To illustrate our approach, recall a standard technique to elicit the certainty equivalent of a lottery, the Multiple Price List (MPL): a list of questions with the lottery fixed on the left, while on the right is an amount of dollars that increases as we proceed down the rows. In each line, the agent has to choose either the left (lottery) or the right (money) option; the highest value against which the lottery is chosen indicates the certainty equivalent. We make a simple twist: in our experiment, instead of left or right, in each line subjects need to indicate a number between 0 and 10. If they choose 0, they get the left option; if 10, the right one; but 5 gives them each option with probability 50%; 3 means getting the left option with probability 30%, *etc*. That is, in each line subjects can choose left, right, or one of many lotteries between them.

Individuals who follow Expected Utility and reduce compound lotteries should not report numbers other than 0 or 10, except for one line if they are indifferent: because Expected Utility is linear in probabilities, there is no benefit in choosing probabilities other than 0 or 1. However, once we depart from Expected Utility, we can have randomization in more than one row. In Section 2 we discuss how it may arise from complete preferences that have regions of strict convexity or from incomplete (or imprecise) preferences.

Our method allows us to elicit ranges of values for which subjects choose to randomize. We use it to capture ranges of *i*) certainty equivalents of a lottery, *ii*) lottery equivalents of a sure amount, and *iii*) lottery equivalents of another lottery. We also ask the same questions using standard MPLs, allowing a direct comparison; and we measure individual characteristics, including risk attitudes, Certainty Bias, measures of IQ, and overconfidence.

Results. *Most subjects report ranges.* About three-quarters of subjects report in at least two out of three questions ranges of values in which they want to randomize; only 21% never do. This is confirmed by other measures of randomization.

Second, *ranges are 'very large'*. For example, when asked to choose between $x for sure and $20 or $0 with equal chances, on average subjects want to randomize for all *x*s between $5.3 to $11.9. That is, there is a large breadth of values, covering many possible risk attitudes, where subjects choose to randomize. Similar results hold for other questions. The same is true if we look at the number of rows where subjects choose to randomize: when subjects randomize, on average they do so for about *half* of the rows in each question. Ranges remain widespread and sizable also if we take more restrictive definitions—for example, if we define a range as giving at least 40% of chances to both options.

Third, *ranges tend to involve values in the risk-seeking domain*: for example, in the question above the top of the range exceeds the risk-neutral value for 73% of subjects. We discuss how this is incompatible with risk aversion and signals more complex risk attitudes. In general, ranges encompass both risk-averse and risk-seeking values, although they are asymmetric and extend much further into risk-averse areas.

Fourth, *choices in standard MPLs on average fall around the middle of ranges*. When taking more restrictive definitions of ranges (at least 40% to both options), answers to standard MPLs *fall instead towards more risk-averse values*.

Fifth, *virtually all subjects who exhibit non-monotone choices in standard MPLs also have* *ranges*, and non-monotonicities and ranges are related in multiple ways. Instead of mistakes, non-monotonicities could be signs of more complex preferences.

Overall, our key finding is that the majority of our subjects exhibit preferences for randomization that span wide ranges encompassing values typically used in economic problems. Far from being restricted to knife-edge cases, preferences for randomization are prevalent and widespread. Ignoring them in modeling or in interpreting data may substantially misrepresent choices and preferences.

Related Literature. A large empirical literature documents preferences for randomization: see Agranov and Ortoleva (2022) for a recent review. (We discuss the theoretical literature in the next Section.) Several papers find strict convexity of preferences under risk (Becker et al., 1964; Sopher and Narramore, 2000; Feldman and Rehbeck, 2022), preference for randomization with objective lotteries (Agranov and Ortoleva, 2017; Dwenger et al., 2018), ambiguity (Cettolin and Riedl, 2019), time preferences (Agranov and Ortoleva, 2017), social preferences (Kircher et al., 2013; Agranov and Ortoleva, 2017; Miao and Zhong, 2018; Chew et al., 2022), and even with dominated options (Rubinstein, 2002); others document this behavior in the field for university choices (Dwenger et al., 2018) or donation (Zhang and Zhong, 2020). Agranov et al. (2022) finds high rates of preferences for randomization in different domains and shows that most persist after explicit training, substantiating the belief that they represent a fundamental trait of preferences and not just an error. Permana (2020) and one of the treatments in Cettolin and Riedl (2019) use modified MPLs where subjects are also allowed to express "I am not sure what to choose" or "I am indifferent," respectively, which gives an even draw, but find little preference for randomization.[1] Feldman and Rehbeck (2022) use the convex budgets method, allowing for more nuanced randomization, and document that many subjects choose non-degenerate mixtures. Arts et al. (2021) present several choices between fixed lotteries and sure amounts in random order and allow subjects to specify randomization probabilities; again, they document substantial and frequent choices of randomization. Finally, the marketing literature offers two methods to elicit the willingness to pay for a product that includes some aspects of randomization, called ICERANGE and BDM-range. While very different from ours, significant ranges of values for which participants choose to randomize are also found using these methods; see Dost and Wilken (2012) and references therein.[2]

This paper is also related to the growing literature that aims to capture preference incompleteness, preference imprecision, or cognitive uncertainty using variants of MPLs. The techniques used are reminiscent of ours but do not involve randomization and are not incentivized. Cohen et al. (1985, 1987) use MPLs in which, in any row, subjects are allowed to answer that they "do not know." They found that this option is often used; however, this is not incentivized because, when this option is used, payments are computed taking as a switching point the middle of the 'do not know' interval. Other papers measure preference imprecision—a notion distinct from, but related to, incomplete pref-

[1] This may be due to the non-neutral wording, or to the fact that only the 50/50 mixture was offered.

[2] In BDM-range, participants state two prices for an object: a floor price and a ceiling price. Like in BDM, a price is randomly drawn. If the drawn price is above the ceiling, the product is not purchased. If it is below the floor, the product is purchased, and the participant pays the price. If the price is in-between, the participant is asked to choose whether to buy or not: if they choose not to, no purchase occurs; if they choose to, only with 50% probability do they receive the product and pay the price. ICERANGE adds a third price and a more complex procedure (Wang et al., 2007). Both methods differ from ours in several aspects and may relate to a desire to delay decisions; in our case, participants are not asked again and directly choose the probabilities of randomization.

erences: Dubourg et al. (1994, 1997); Butler and Loomes (2007, 2011); Cubitt et al. (2015). In these papers, subjects are either asked to choose between options but also to report the strength of their preferences (which is inconsequential); or they face an MPL in which they can choose not to make a choice in some rows, but also have to separately report their switching point (only the latter matters for payment). These papers document sizable preference imprecision and link it to features like violations of Independence and preference reversal (Butler and Loomes, 2007, 2011). More recently, Enke and Graeber (2023, 2022) measure "cognitive uncertainty:" in a series of questions including risk and ambiguity preferences, subjects first choose from an MPL; then, in a second screen, they indicate two bounds they are "certain" about; these are inconsequential for payment; they are also unrelated to randomization. They find sizable ranges and relate them to behavior in different areas, including risk and ambiguity. Like these papers, we also aim to measure ranges; but we do so using the desire to randomize (and thus have an incentivized measure).

Finally, we connect randomization to violations of monotonicity in MPLs. We are not the first to suggest deliberate randomization as a possible explanation or document such a link. In two large-scale experiments, Chew et al. (2022) study the determinants of non-monotone behavior in MPLs. They distinguish such violations into 'irregular' and 'regular.' They find that irregular violations correlate with violations of first-order stochastic dominance, pointing toward noise and mistakes. Instead, regular violations relate to non-expected utility behavior, correct reduction of compound lottery, and, importantly, deliberate randomization in repeated choice. Our analysis on this point was inspired by their work; our results confirm and elaborate on these findings.

2 Theoretical Background

In our experiment, subjects can report numbers between 0 and 10 to express their desire to randomize between options. What do theories predict? We begin with Expected Utility and how these choices relate to risk attitudes. While standard, both observations are important in our analysis below. We then briefly discuss several approaches that generalize Expected utility and allow for preferences for randomization. Readers not interested in this discussion may skip this section and proceed to the next.

No Range with Expected Utility. Agents who reduce compound lotteries and whose choices maximize complete, monotone preferences that follow Expected Utility will not choose numbers other than 0 or 10 in more than one row. Under Expected Utility, if p is strictly preferred to q, it is also strictly preferred to any mixture between them, a direct implication of the Independence axiom. Only if p and q are indifferent is the individual also indifferent between any randomization. Since in MPLs one of the options is fixed and the other becomes strictly better, we can have indifference *in at most one row*. Thus, if we observe numbers other than 0 or 10 in more than one row, we are documenting a violation of this model. (In fact, it is a violation of a property weaker than Independence, Betweenness, Dekel 1986.)

Randomization and Risk Attitudes. For any monetary lottery p, denote by $E[p]$ its expected value; denote by δ_y the degenerate lottery that pays y. Recall that an agent is *risk-averse* if the expected value of a lottery is preferred to the lottery, i.e., $\delta_{E[p]} \succeq p$ for all p. In general, risk attitudes and attitudes towards randomization are unrelated: the former is linked to the shape of the utility function over money (the Bernoulli index in Expected Utility), while the latter is related to the shape in probabilities. Once we depart from Expected Utility, it is easy to construct examples where the agent is risk-averse or risk-seeking yet weakly, and sometimes strictly, prefers to randomize, or vice-versa.[3]

However, preferences for randomization can be informative about risk attitudes even outside of the Expected Utility framework: if $x > E[p]$, randomizing between p and δ_x is a violation of risk aversion. This is because choosing to randomize means $q := \alpha p + (1-\alpha)\delta_x \succeq \delta_x, p$ for some $\alpha \in (0,1)$. If $x > E[p]$, then $x > E[q]$, thus $\delta_x \succ \delta_{E[q]}$. It follows that $q \succ \delta_{E[q]}$, violating risk aversion. This will be of particular relevance below.

We now discuss several approaches that allow for preferences for randomization.

Complete Preferences exhibiting Convexity. Consider individuals with complete preferences over lotteries, who reduce the compound lottery created by randomization but who, contrary to Expected Utility, have regions of strict convexity in probability: for some $\alpha \in (0,1)$ and some p and q, $\alpha p + (1-\alpha)q \succ p, q$. In these regions, these individuals will exhibit preferences for randomization. Several models allow for regions of strict convexity. In Rank Dependent Expected Utility, or Cumulative Prospect Theory (Quiggin, 1982; Kahneman et al., 1991), preferences can have areas of strict convexity if probability weighting is not always pessimistic.[4] Under Cautious Expected Utility, preferences are

[3]For example, under Cautious Expected Utility, if all utilities in the representation are concave (resp. convex), then preferences are risk-averse (seeking). Yet, these preferences are convex, possibly strictly so.

[4]There are no preferences for randomization when probability weighting is pessimistic (e.g., see Section 2.3 in Dillenberger and Raymond 2019); in fact, preferences may exhibit the opposite tendency (the mixture between two indifferent lotteries is strictly worse). A preference for randomization may arise if there are areas of optimistic probability weighting. For example, suppose small probabilities are overweighted and large ones underweighted, with probabilities of $\frac{1}{4}$, $\frac{1}{2}$, and $\frac{3}{4}$, weighted to $\frac{3}{10}$, $\frac{1}{3}$, and $\frac{1}{2}$, respectively. Suppose that the utility index is the identity function. Then $\frac{1}{2}\delta_9 + \frac{1}{2}\delta_0 \sim \delta_3$ and $\frac{1}{2}\left(\frac{1}{2}\delta_9 + \frac{1}{2}\delta_0\right) + \frac{1}{2}\delta_3 \succ \delta_3$.

in general weakly convex (Cerreia-Vioglio et al., 2015), and typically have areas of strict convexity (e.g., when the set of utilities is finite; Cerreia-Vioglio et al. 2020). However, this model does not allow for randomization with degenerate lotteries.[5]

This view of randomization is related to the several papers that derive stochastic choice (the observation that the same subject may give different answers to the same question) as the outcome of deliberate randomization: see Machina (1985); Agranov and Ortoleva (2017); Cerreia-Vioglio et al. (2019). In these models, stochastic behavior is due to non-Expected Utility and strict convexity. In this light, our question format provides an external randomization device that allows individuals to generate the desired stochasticity.

Finally, note that any randomization creates a compound lottery. Robust evidence shows that individuals typically fail to reduce compound lotteries and are averse to them, pushing against randomization. This implies that we need the forces pushing in its favor, like convexity, to be sufficiently strong to compensate for the aversion to the induced compound lotteries.

Incomplete Preferences. Alternatively, evidence of randomization can be considered evidence of preference *incompleteness*. Intuitively, agents may want to randomize when they are unsure of what to choose; they use randomizations as ways to resolve the uncertainty. Therefore, the choice of randomization indicates that they do not know how to rank the two options.[6] Indeed, a large and growing literature studied incomplete pref-

[5] That is, we cannot have p, δ_x, α such that $\alpha\delta_x + (1-\alpha)p \succ p, \delta_x$. This is directly implied by the Negative Certainty Independence axiom. See Cerreia-Vioglio et al. (2015, p. 697).

[6] This could be formalized, relating randomization and incompleteness following Ghirardato et al. (2004), Cerreia-Vioglio (2010), Cerreia-Vioglio et al. (2015). For any preference relation $\succeq$ over lotteries

erences and their possible completions: see, among many, Bewley (1986); Dubra et al. (2004); Ghirardato et al. (2004); Gilboa et al. (2010); Ok et al. (2012); Cerreia-Vioglio et al. (2015), and, more recently, Ok and Nishimura (2019) and references therein.

It is worth noting how, according to this interpretation, subjects randomize to resolve a difficulty in choosing and may not perceive this action as creating a complex compound lottery. As such, preferences for randomization may not stand in contrast with the standard aversion to compound lotteries.

If we understand ranges of randomization as indicating the region of incompleteness, we can look at where forced choices in regular MPLs stand with respect to this range. Several models propose rules on how incomplete preferences should be completed when subjects are forced to make a choice: in the context of risk, see Cerreia-Vioglio (2010); Cerreia-Vioglio et al. (2015); Ok and Nishimura (2019). Cerreia-Vioglio et al. (2015) propose cautious completions: if preferences are incomplete, when forced to choose individuals should select a sure amount if available ('when in doubt go with certainty'). They show (Th. 5) that such a rule implies that there is a unique (well-behaved) completion and that is such that it can be represented by Cautious Expected Utility. As we will discuss in Section 5, this completion rule will be useful in the interpretation of our results.

define its linear core $\succeq'$ as its largest subrelation that satisfies independence, i.e., $p \succeq' q$ iff $\lambda p + (1-\lambda)r \succeq \lambda q + (1-\lambda)r$ for all $\lambda \in (0,1)$ and lottery r. The relation $\succeq'$ is naturally incomplete if $\succeq$ violates Independence; the rankings it includes may be understood as those that the agent is sure about, as they are preserved when both options are mixed with others. If an individual prefers the mixture between p and q, then p and q must be *incomparable* according to the linear core $\succeq'$. Thus, evidence of randomization can be interpreted as evidence of incompleteness of preferences. It is also worth noting that, in our context, it is not immediate how to formally set apart incompleteness with specific completions and non-Expected Utility. For example, Cerreia-Vioglio et al. (2015) derived Cautious Expected Utility both starting from complete, non-EU preferences; and as a completion of incomplete preferences.

Preference Imprecision and Cognitive Uncertainty. As mentioned above, the ranges elicited in our tasks can also be related to preference imprecision (see Butler and Loomes, 2011 for a formal treatment) or to the notion of cognitive uncertainty introduced by Enke and Graeber (2023). In turn, this is related to theoretical models of cognitive imprecision: see Khaw et al. (2021, 2022) and many references therein. While all of these approaches are related to incompleteness, in their standard formulation, none of these theories gives preference for randomization, as they typically assume Expected Utility conditional on the signals received (which gives no incentive to randomize). Modifying these theories to include such preference would obtain models related to Cautious Expected Utility.

Regret. Finally, preference for randomization may come from fear of regret or decision avoidance: delegating their decision to chance, individuals may feel less responsible if they obtain a lower-paying outcome. It is worth noting that Dwenger et al. (2018) investigated this possibility in their final questionnaire but found it explains the behavior of a small group of subjects; the majority indicated answers pointing to incompleteness.

3 Design

The experiment included 10 main questions and several control tasks; complete instructions and the screenshots are presented in Appendix C. The main questions involved choices between monetary lotteries with objective probabilities. These questions were of two types: standard multiple price lists (MPL, hereafter) and *range* multiple price-list (r-MPL, hereafter).

A standard MPL consists of several rows, each including two options, Left and Right. The Left option is the same in all rows, while the Right one changes, becoming more attractive as we go down the rows. Subjects are required to select one option in each row.

Range MPL, or r-MPL, are almost identical, but in each row subjects are asked, instead of Left or Right, to indicate an integer number from 0 to 10. This number corresponds to the probability of receiving the Left option: specifying an integer i means getting the Left option with probability $10i\%$ and the Right option with the remaining probability of $(100-10i)\%$. Thus, indicating 3 means getting the Left option with probability 30%. This enriches the standard MPL technique by allowing subjects to choose any combination of two options while maintaining the ability to choose either option for sure.

Table 1 lists the 10 main questions, where the last column includes the range of values of the Right option (the exact steps are listed in Appendix A). Note that Q1r-2r-3r correspond to Q1-2-3, except that they use r-MPLs instead of MPLs. This allows us to compare behavior across types of questions. To investigate if responses are sensitive to specific kinds of questions, we purposely vary Left-Right combinations: fixed lottery vs. sure amount, fixed sure amount vs. lottery, and lottery vs. lottery. Questions Q4-Q7 measure risk attitudes and Certainty Bias—note that Q4 and Q5, and Q6 and Q7, have common-ratio-type variations.

[[TABLE 1 HERE]]

In addition, subjects completed two investment tasks (Gneezy and Potters, 1997), measures of IQ (ICAR, Condon and Revelle, 2014) and overconfidence, and a non-incentivized questionnaire.[7] Appendix A.8 discusses the answer to this questionnaire, finding them broadly consistent with choices.

Subjects' payment consisted of two parts in addition to the participation fee ($7). First, one of the 10 main questions or two investment tasks was selected at random; if the chosen question was a MPL or r-MPL, one of its rows was selected and the choice implemented; for r-MPLs, if subjects chose a non-degenerate lottery, each option was given with the specified probability. Second, one control question (IQ, overconfidence) was randomly selected for payment.

4 Results

Preliminaries. A total of 165 subjects participated in an experiment run at the University of California, Irvine; all subjects were undergraduate students at that institution. We focus on the 148 subjects who made non-dominated choices in all ten questions.[8]

In discussing behavior in standard MPLs, we distinguish between subjects with *monotone* and *non-monotone* choices. The former switch from the Left to the Right option at most once; the latter switch multiple times. With monotone choices, the key measure is the dollar amount linked to the switching point: following standard practice, we use the average dollar amount between the last Left and the first Right choice.

To analyze choices in r-MPLs, we use two types of measures. First of all, we have mea-

[7] To test for order effects, subjects were randomly assigned to one of the two possible orders of questions. In Appendix B we present the two orders and show that our message remains unchanged in either order, with the majority of subjects exhibiting our behaviors of interests in both, even though the order had some effects.

[8] The remaining 17 subjects are either not paying attention or have non-monotone preferences on money, making the analysis difficult. Including them does not change significantly any result; we omit it for brevity.

sures of ranges. Specifically, we denote by *range91* the range of dollar amounts in which a subject chooses both options with positive probability: the range from the smallest to the largest value for which they indicate numbers between 9 and 1.[9] We say that a subject *exhibits a range91* if numbers other than 0 or 10 appear in more than one row (as choosing it in one row alone could be due to indifference.) We construct *range64* in a similar way, except that we focus on the range in which both options are chosen with probability at least 40%—indicating 4, 5, or 6. While range91 captures all values for which the subject prefers to mix, range64, obviously a subset, indicates values for which the subject assigns high weight to both options. In the main body, we restrict our attention to these two ranges. Appendix A.2 replicates our analysis for ranges 8-2, 7-3, and 5-5, defined similarly, finding coherent results.

While range91 and range64 identify where subjects want to randomize, they do not take into account the chosen probabilities. Our final measure, which we call the q-measure, does so.[10] Let us illustrate it using question Q1r. Denote by $p(x)$ probability specified by a participant in a particular step of the r-MPL and by $q(x) = min\{p(x), 1-p(x)\}$; for each x, $q(x)$ captures how even is the randomization chosen by the agent. Then, we can define

$$q := \frac{2}{\text{Nb of non-FOSD steps in the Q1r}} \sum_{x>0}^{x<20} q(x).$$

In words, our q-measure is the sum of all $q(x)$ normalized by the number of rows in which no first-order stochastic dominance relation exists between the left and the right options

[9] To be consistent with how we code behavior in standard MPLs, one extreme of range91 is the average dollar amount between the last row in which the subject chose 10 and the first with values in [1, 9]; similarly, the other extreme is the average dollar amount between the last row with values in [1, 9] and the next.

[10] We thank one anonymous referee for suggesting this measure.

("Nb of non-FOSD steps in the Q1r" in the formula above). Thus, it takes a value of 0 if a participant never wants to randomize and keeps a low value if the chosen randomizations tend to be skewed, with values like .1 or .9. On the other hand, the q-measure takes a value of 1 if participants chose a 50/50 randomization in every step (excluding steps with first-order stochastic dominance), and, in general, is higher the more frequent the randomization and the closer it is to 50/50. As such, the q-measure takes into account the exact probabilities specified by a participant, speaking to the aggregate magnitude of randomization.

4.1 Measures of Randomization: Frequency and Size

People randomize *a lot*, report ranges *very often*, and these ranges are *big*. Table 2 shows the fraction of subjects who exhibit either kind of range and have positive q-measure in each question. This table also reports basic summary statistics of all three measures of randomization.

[[TABLE 2 HERE]]

Ranges are Frequent. In all questions, between two-thirds and three-quarters of subjects exhibit range91; more than 50% do so for range64. Furthermore, the vast majority exhibit ranges in at least one question: only 21% of subjects never exhibit range91 in any of the three questions, and only 28% never exhibit range64. Instead, most exhibit ranges91 in multiple occasions: 55% in all three questions, 19% in two, and 5% in one. Similarly, 30% exhibit ranges64 in all three questions, 28% in two, and 14% in one. That is: most subjects exhibit ranges in most questions; this is true even if we take restrictive definitions of ranges, like range64.

Ranges are Big. Conditional of exhibiting a range, ranges are '*very large*:' if we look at Q1r, subjects who exhibit range91 (72%) have an average range of size \$6.6, from \$5.3 to \$11.9. The top of the range is more than twice the bottom. This is a remarkable span, especially because this question measures the certainty equivalent of a lottery that pays \$20 or \$0 with equal chances. Ranges remain 'very large' also for Q2r and Q3r (the former is larger, also reflecting the different scale of payoffs).

Ranges are big also if we look at the number of rows with randomization instead of the span in dollars. On average, subjects want to randomize in 9.3 rows in Q1r, 12.7 in Q2r, and 8.6 in Q3r (medians are 10, 14, and 9).[11] These are about *half or more* of the rows in each question (there are 19, 23, and 17 rows, respectively). It shows that the large span of range91 is not an artifact of large gaps; rather, subjects consistently choose to randomize in many rows. This should alleviate concerns that ranges are chosen by mistake or due to the demand effect; we elaborate on these points in Section 5.

Range64 is also consistently large, with a mean of \$2.7 for Q1r. The top of the range is more than 30% above the bottom. We have large ranges also for the other questions. These ranges are also big in terms of the number of rows: on average, 4.4, 7.6, and 4.1 rows in Q1r, Q2r, and Q3r, respectively (medians are 4, 6, and 3). Therefore, there is a sizable span of dollar values and a sizable number of rows for which subjects not only want

[11] Moreover, the large majority of these answers are monotone, in the sense that subjects report weakly decreasing numbers within the range as we proceed down the rows. We have monotone answers for 85% of subjects with ranges in Q1r (n = 106), 74% in Q2r (n = 109), and 85% in Q3r (n = 94). Violations of monotonicity are highly related to violations in standard MPLs.

to randomize but do so giving at least a 40% chance to both options. Overall, subjects are creating very substantial randomizations, with high probabilities for many options.

q-measures are Large. In all questions, between two-thirds and three-quarters of subjects have strictly positive q-measure. The average and the median values of q-measure for subjects with positive q's are about one-third. (Figure A.1 in Online Appendix A.3 shows the histograms.) This value is large: it corresponds to a situation in which the average participant randomizes 50/50 between the left and the right options in roughly a third of all rows presented on the screen or randomizes with more skewed probabilities towards one option in more than a third of all rows.

Another way to visualize the extent of the preference for randomization is to look at the values taken by $q(x)$ for each x; recall that, if $p(x)$ is the probability in the chosen randomization, then $q(x)$ is equal to $\min\{p(x), 1-p(x)\}$, and is therefore between 0 (no randomization) and 0.5 (even randomization). Figure 1 displays a violin plot of these values for question Q1r, focusing on subjects who exhibit a range in this question, including the median (hollow dots) and mean (filled dots) values. (Figure A.2 in Appendix A.3 do the same for Q2r and Q3r.) This graph shows that many subjects pick randomization probabilities close to 50/50 for several values. For example, consider $x = 10$, the risk-neutral value. The median $q(10)$ is .3, which means that 50% of subjects who randomize in this question do so with a probability of at least 30%. Overall, this graph shows, once again, that subjects who randomize do so by picking probabilities that give non-trivial randomization and doing so for many values of x.

[[FIGURE 1 HERE]]

Risk Attitude. We showed that subjects exhibit ranges and that these ranges are big. Where are they located? From Table 2 it is clear that they span substantially *above and below the risk-neutral value*. Consider again Q1: the expected value of the lottery is \$10; the average range goes from \$5.3 to \$11.9. The range includes the risk-neutral value in its interior and extends in both directions. This is not symmetric: ranges extend much deeper into the risk-averse direction (lower numbers). But they do span into the risk-seeking area: not only the average top of the range is \$11.9, but of the subjects that exhibit range91 for Q1, 73% have the top part of the range strictly above \$10. Similar results hold for Q2 and Q3, where ratios are 91% and 77%, respectively. (For Q2r, this means values below \$36; for Q3r, values above \$18.)

Another way to see this is by looking again at Figure 1. Substantial randomization is also present for xs above the risk-neutral value of 10: the average $q(x)$ is above 0.2 well into the risk-seeking area. Distributions are, however, not symmetric around the risk-neutral values, and we have more randomization in the risk-averse area. (The same holds for Q2r and Q3r, as shown in Figure A.2 in Appendix A.3.)

Taken together, these observations paint a clear picture: conditional on having a range, the majority of our subjects prefer to randomize between a lottery and sure amounts also *above* the expected value of the lottery. We have seen (Section 2) that this is incompatible with risk aversion, even though choices in regular MPLs are typically risk-averse. At the same time, ranges extend deeper in the risk-averse area, where subjects randomize more.

One interpretation is that, while choices in the standard MPLs tend to be risk-averse, r-MPLs allow us to acquire a more nuanced view. If, in line with some of the models discussed above, we interpret ranges as boundaries of the values the agent is sure about—or of the utility functions that are being considered—our results suggest that agents consider values that fall both in the risk-averse *and* in the risk-seeking domain. It is as if they were not fully sure they should be risk-averse. When asked to make a precise choice in standard MPLs, however, they tend to fall in the risk-averse area. We discuss these choices next.

4.2 Ranges and standard MPLs

We use both standard MPLs and r-MPLs for the same three questions. This allows us to ask: Where does the choice in the standard MPL fall with respect to the range expressed in the corresponding r-MPL? In the middle or biased toward one end?

For subjects without ranges. We begin with a sanity check. For each question, consider subjects with no range in the r-MPL and with monotonic answers in the MPL.[12] Do answers coincide? This is informative of whether r-MPLs bias responses in a particular way.

We find that answers are highly related and show no particular bias. In all questions, for subjects without a range there is a very high correlation between switching points in MPLs and r-MPLs: correlations are 0.84, 0.91, and 0.72 for Q1, Q2, and Q3 and are significant at 1% level.[13] Moreover, the differences in switching points are distributed around zero, suggesting there is no bias. In other words, subjects who don't express ranges make consistent choices in both formats.[14]

[12]We have 36 participants with monotonic choices in the Q1 question and no ranges in the corresponding Q1r question, 28 such participants in Q2-Q2r, and 50 in Q3-Q3r.

[13]As a benchmark, it may be worth recalling that even identical standard MPLs have similar or often lower correlations when repeated in different parts of the same experiment. For example, in Chapman et al. (2022), the correlation between two conceptually identical MPLs (riskAversionGain1 and riskAversionUrn1) appearing in different randomized parts of the experiment is 0.45.

[14]Moreover, their choices are also not concentrated around salient values. Appendix A.4 presents an

Defining a measure. We now turn to subjects with monotone choices in MPLs but also a range in r-MPLs. (We discuss subjects with non-monotone choices below.) To shed light on their behavior, for each subject in this group and relevant question Q_i, define $\lambda_{91}^{Q_i} \in \mathbb{R}$ by

$$\text{Standard MPL}^{Q_i} = \lambda_{91}^{Q_i} \cdot \text{Top range91}^{Q_i} + (1 - \lambda_{91}^{Q_i}) \cdot \text{Bottom range91}^{Q_i},$$

where, as names suggest, 'Standard MPL' is the switch point in standard MPLs; 'Top range91' is the boundary of range91 in the direction of less risk aversion; 'Bottom range91' is the other.[15] In words, λ_{91} represent where the choice in a standard MPL fall w.r.t. the range expressed in the corresponding r-MPL. Values are in $[0,1]$ if and only if the former falls inside the range, and lower (higher) values indicate choices towards more (less) risk-averse options. Thus, $\lambda_{91} = 0$ means that the choice in the MPL is exactly at the most risk-averse point of range91. We define λ_{64} analogously.

Distribution of λs. Figure 2 plots the distributions of average λ_{91} and λ_{64} across subjects, where we take the average across questions (eliminating 2 outliers for clarity; Figure A.5 in Appendix plots them by question). The picture is clear.

First of all, λ_{91} tends to be symmetric around 0.5 and concentrated in $[0,1]$. Indeed, it is within these bounds for 82% of subjects in Q1, 75% in Q2, and 67% in Q3. This means most subjects' choices in standard questions fall within the range91.

By contrast, λ_{64} is far from symmetric and tends to revolve around zero, with more than half the subjects (53%) below 0.25. Moreover, choices in standard MPLs tend to fall

in-depth analysis.

[15] For Q1r and Q3r 'Top range91' and 'Bottom range91' are the highest and lowest numbers in the ranges, respectively. Instead, for Q2r, 'Bottom range91' is the highest number while the 'Top range91' is the lowest.

outside this more restrictive range: the fraction of subjects with λ_{64} in $[0,1]$ are 47%, 48%, and 32% in Q1, Q2, and Q3, respectively.

Overall, this means that: i) choices in standard MPLs are in the middle of range91; but they are also ii) closer to the *bottom* of range64—the more risk-averse choice.

[[FIGURE 2 HERE]]

We can connect these results with our analysis of risk attitudes and ranges. We have seen that ranges typically extend to the risk-seeking domain, but they also tend to be asymmetric, extending further into risk-averse areas. If choices in standard MPLs tend to fall in the middle of range91, or at the bottom (more risk-averse) of range64, then they will be risk-averse. As we suggested above, one possible interpretation is that subjects are unsure of how to evaluate lotteries and contemplate both risk-averse and risk-seeking options; but when it comes to selecting one option in standard MPLs, they pick a risk-averse one, either being cautious or because the range they consider is asymmetric.

4.3 Ranges and Individual Characteristics

How do ranges relate to individual characteristics such as risk attitudes, Certainty Bias, IQ, and overconfidence?

Measures. The answers to standard MPLs in Q1-2-3 give us three continuous measures of risk aversion: the certainty equivalent of a lottery (Q1), the lottery equivalent of a sure amount (Q2), and the lottery equivalent of a lottery (Q3). Following Chapman et al. (2022), we treat these measures separately. In addition, we have the two investment tasks a' la Gneezy and Potters (1997), denoted Inv. Task 1 and 2. For ease of comparison, values are normalized such that higher numbers mean more risk aversion.

The difference between the answers to Q4 and Q5, and to Q6 and Q7, give us two measures of the common ratio effect, or Certainty Bias (Allais, 1953). We take the weighted average (dividing by the fixed amount on the MPLs) and denote it *C-bias*. Higher numbers denote higher certainty bias.[16]

[[TABLE 3]]

Results. Table 3 shows the correlation between these measures with: an indicator on whether the subject exhibits the range at all (Ind. Range); the number of questions with ranges (Range Freq.); the average dollar size of ranges[17] (Range Size); and the q-measure averaged across three r-MPL questions. The p-values are adjusted for multiple hypotheses using the method of Benjamini and Hochberg (1995) (see Anderson 2008), which naturally reduces the significance; table A.2 in the Appendix presents the uncorrected version. The results are clear.

First, there are some but limited relations between risk attitudes and both ranges and the *q*-measure. Risk Q2 is weakly related to range91 and the q-measure, with more risk-averse subjects reporting ranges in more questions, reporting larger ranges, and featuring larger q-measures. At the same time, this does not extend to other measures of risk and many measures of range64.[18]

[16] The variable C-Bias is constructed by $\frac{1}{16}(Q4-Q5)+\frac{1}{14}(Q6-Q7)$, where $Q4$ is value in Q4, *etc*. For subjects for whom we have this measure (which requires monotone answers), 59%, 6%, and 35% have positive, zero, and negative values.

[17] This is the weighted average of range sizes, normalized by the expected value of the Left Option. In Appendix A.6, we show that the results remain the same if we normalize by the maximum possible size of the range in each question (last column of Table 1).

[18] The fact that risk measures from Q1 and Q2 relate differently is not surprising in light of Chapman

Second, Certainty Bias relates negatively to the frequency and size of ranges and the q-measure. Subjects with more such bias have less frequent and smaller ranges and lower q-measures. There is, therefore, weak evidence of a connection between these two forms of violations of Expected Utility, but in the direction opposite of what some may expect—with ranges and randomization related to a smaller certainty bias.

Third, we can look at the relationship between Certainty Bias and our measures of λ (not included in Table 3). We find some evidence of a negative relation between λ_{64} and Certainty Bias ($corr = -0.22$ with $p = 0.06$, uncorrected), but not using λ_{91} ($corr = -0.02$ with $p = 0.85$, uncorrected). This means that subjects who exhibit a Certainty Bias tend to make choices that fall in the *bottom* (more risk averse) part of the range of higher randomization, range64.

IQ and Overconfidence. We also analyze the relationship between ranges and IQ and measures of overconfidence (overestimation, overplacement, and overprecision). We find no robust evidence of any relation; details are discussed in Appendix A.7.

4.4 Ranges and Non-Monotone Choices

We conclude with a connection with non-monotone choices. Experiments that use standard MPLs usually find a sizable fraction of non-monotone answers, i.e., multiple switches between the left and right options. This is true in our sample as well: 28% of subjects

et al. (2022, 2023), where only the latter is related to several demographics and other characteristics. It is worth noting also that measures of risk preferences from investment tasks are weakly related to the others (again, not surprising in light of the literature and the fact that these measures bunch risk-seeking behavior): while Inv. Task 1 and 2 have a correlation of 0.80 with each other, this decreases to 0.25, 0.18, and 0.13 for the correlation of Inv. Task 1 with Risk Q1, Q2, and Q3, respectively; and to 0.17, 0.12, and 0.11 for the correlation of Inv. Task 2 with Risk Q1, Q2, and Q3, respectively.

display non-monotonic choices at least once in Q1, Q2, or Q3. The typical approach is to disregard these choices, treating them as solely noise. But what if they are instead informative? For example, subjects may be randomizing in each line and thus exhibit non-monotone behavior—as already suggested by Chew et al. (2022).

Our first observation is that non-monotone behavior is very strongly related to ranges. This appears in the last row of Table 3, showing that violations of monotonicity very strongly correlate with *i*) exhibiting ranges, *ii*) doing so more often, *iii*) having larger ranges, and iv) randomizing more intensely as captured by the q-measure.

A second approach is to classify subjects into types based on 1) choice monotonicity in MPLs, and 2) the existence of ranges91 in r-MPLs. This leaves us with four types:

1. Monotone and no ranges. These are the types predicted by standard theory. In our sample, only 17% (25 subjects) are monotone in all MPLs and never report ranges.

2. Monotone and ranges. About half of our subjects (49% or 72 subjects) belong to this category.

3. Non-monotone and ranges. A third of subjects (30% or 45 subjects) display non-monotone behavior in MPLs and exhibit ranges r-MPLs.

4. Non-monotone and no ranges. This category is essentially non-existent in our sample. We observe only 6 subjects (4%) of this type.

This classification shows two key aspects of our data. First, 'standard' subjects—monotone and no ranges—are a real minority in our sample. The majority has ranges and tends to have monotone choices. Importantly, of the subjects that ever violate monotonicity, essentially all of them also exhibit ranges, showing a connection between these tendencies. These results confirm and elaborate on the ones in Chew et al. (2022).

These results also have implications for the role of external randomization devices. While subjects may have a preference for randomization, some may need an external device, while others may be able to do so in their heads. From this point of view, subjects in group “monotone and ranges,” randomize with an external device but don’t do so within a given standard MPL; subjects in group “non-monotone and ranges,” may instead be randomizing internally even in standard MPLs.

4.5 Concerns

Noise. Could our findings be due to noise? Subjects may report ‘5’ in a row of our r-MPL not because they genuinely want to randomize but because they are close to indifferent and give a ‘noisy’ answer; in our design, this can imply randomization. (Such would be the implication, for example, of a probit or logit model.) Our data suggests this is unlikely to be driving our results. First, subjects’ answers are generally coherent within r-MPLs, suggesting that if noise is at play, it is at play at the whole r-MPL level, not on individual rows. (Otherwise, we would not observe coherent descending numbers but a much more ‘noisy’ sequence.) Second, in all standard models, including logit and probit, noise is most prevalent close to indifference and becomes uncommon as utility difference increases. Yet, we observe randomization in large ranges involving wide spans of dollar values, where utility differences become more considerable. Moreover, we observe them for several rows. For these reasons, we do not believe our findings are likely to be due (only) to noise. At the same, one may want to account for noise in elicitation: even assuming that subjects do want to randomize, our data is likely reflecting some noise. Several approaches have been developed to account for noise in regular MPLs; we are not aware of suitable methods applicable to our setup and believe this to be an interesting question for future research.

Experimental Demand. Could our results be due to experimental demand rather than genuine preferences for randomization? The collective evidence in our experiment suggests this is not the case. First and foremost, we have seen that when subjects randomize, they choose extremely large ranges, with a median of 5 or 6 lines per question and a very large span of dollar values. If they had no real desire to randomize, they would be giving up a non-trivial payoff by choosing ranges of this magnitude; if they only wanted to satisfy the experimenters, they could have chosen much smaller and sporadic ranges. Second, ranges appear only when 'reasonable' to do so (e.g., not with dominance) and with regularity. Third, we can look at the answers given in our final questionnaire, presented in Appendix A.8. We observe a strong and significant correlation between the choices in the experiment and the answers subjects provide in the questionnaire, where two main motives for randomization reported by our participants are 'desire to randomize' and 'desire to delegate the decision to the computer.' While the latter may be an ex-post rationalization, in combination with the other observations, we believe this limits the concern of experimental demand.

Beyond Lotteries? We use objective lotteries as they provide a natural example of complex choice for which existing evidence points to the desire to randomize, for which several theories make predictions, and as they allow us to discuss risk attitudes. However, lotteries add a layer of complexity as randomizing between lotteries generates a compound lottery, towards which different subjects may have complicated attitudes. It would be important to understand the extent to which our results replicate in other domains, e.g., the choice between consumption streams or interpersonal allocations, for which existing evidence already documents preferences for randomization.

5 Discussion

This paper introduces a simple method—a modified Multiple-Price-List—to capture the range of values for which subjects want to randomize between two options. In an experiment, we find that decision-makers express this desire 1) very frequently and 2) for very wide ranges of values, that 3) spill into the risk-seeking realm; moreover, 4) choices in standard questions tend to fall either in the middle or in the bottom of such ranges and, 5) ranges are related to non-monotonic choices in standard questions.

Overall, our results show that preferences for randomization and the corresponding violations of Expected Utility not only exist but are, in fact, prevalent for wide ranges encompassing values typically used in economic problems. These phenomena are therefore far from being restricted to smartly-constructed, knife-edge cases; instead, they apply broadly. Ignoring them may lead to the wrong choices in modeling or in deriving conclusions from existing data.

Implications for Theoretical Models. As we discussed in Section 2, preferences for randomization are mainly explained using models of complete preferences that deviate from Expected Utility or using models of incompleteness, imprecision, and cognitive uncertainty.

If we understand preferences for randomization as emerging from complete non-Expected Utility preferences, our results are compatible with some versions of Rank Dependent Expected Utility in which probability weighting is optimistic at least for some probabilities; it must also satisfy specific restrictions to guarantee that preferences for randomization emerge to the extent we observe in our experiment. Preferences for randomization are also allowed by Cautious Expected Utility, but not for randomization with sure amounts of money extensively documented in Q1r. Overall, our results point towards general models of convex preferences, as studied in Cerreia-Vioglio (2010) and applied to stochastic choice in Cerreia-Vioglio et al. (2019). Additionally, because randomization creates a compound lottery, if we follow the standard approaches and further assume that subjects are averse to such compounding, we need complete preferences over lotteries that exhibit very substantial convexity, as it needs to compensate for the aversion towards compound lotteries and also give preference for randomization for the wide spans of values we document.

Alternatively, we can understand preferences for randomization as a sign of preference incompleteness, imprecision, or cognitive uncertainty: subjects may choose to randomize as a reaction to not knowing what to choose. Indeed, this approach is supported by the answers to our final questionnaire. In this light, our results would point to incompleteness/imprecision/cognitive uncertainty that is both pervasive and holding for wide spans of value.[19] From this point of view, randomization would be a way to resolve a difficulty in choosing, and, as we mentioned above, not be perceived as creating a compound lottery; therefore, it may not stand in contrast with the standard aversion to the latter.

Furthermore, recall that choices in regular MPLs tend to fall at the bottom (more risk-averse) of range64. If we understand this range as indicating areas of more extreme uncertainty where the agent picks randomizations closer to 50/50 because they do not really know what to choose, then our results suggest that *forced choices tend to fall in the most risk-averse part of the range of more extreme uncertainty*. In turn, this points to a specific way of resolving incompleteness: picking the most risk-averse option. As we discussed in Section 2, this aligns with the cautious completion rule suggested in Cerreia-Vioglio et al. (2015), which leads to the Cautious Expected Utility representation for forced choices and Certainty Bias. This has a further testable implication: we should observe a relationship between forced choices lying close at the bottom of range64 (low λ_{64}) and Certainty Bias; as discussed in Section 4.3, we have suggestive evidence consistent with this.

Therefore, one possible explanation of the behavior in our experiment may be: subjects have incomplete preferences and randomize when unsure of what to choose; when randomization is not available, as in regular MPLs, they select the option corresponding to the lowest value for the area of extreme uncertainty, which points to behavior in line with Cautious Expected Utility for forced choices.

[19] This is particularly relevant as several papers linked them to many other phenomena like status quo bias, endowment effect, certainty bias, ambiguity aversion, stochastic choice, present bias, attraction effect, and other aspects: see Masatlioglu and Ok (2005, 2014); Ok and Masatlioglu (2007); Gilboa et al. (2010); Ortoleva (2010); Cerreia-Vioglio et al. (2015, 2022); Ok et al. (2015); Agranov and Ortoleva (2017); Ok and Nishimura (2019); Butler and Loomes (2007, 2011); Khaw et al. (2021); Enke and Graeber (2022, 2023); Gabaix (2019).

Finally, our results on randomization and risk attitude show subjects violating the basic notion of risk aversion even facing simple, 50/50 lotteries, demonstrating a more complex relationship towards risk than typically assumed.

References

Agranov, M., P. J. Healy, and K. Nielsen (2022): "Stable Randomization," *The Economic Journal*, forthcoming.

Agranov, M. and P. Ortoleva (2017): "Stochastic Choice and Preferences for Randomization," *Journal of Political Economy*, 125, 40–68.

——— (2022): "Revealed Preferences for Randomization: An Overview," *AEA Papers and Proceedings*, 112, 426–430.

Allais, M. (1953): "Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école Américaine," *Econometrica*, 21, 503–546.

Anderson, M. (2008): "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association*, 103, 1481–1495.

Arts, S., Q. Ong, and J. Qiu (2021): "Measuring subjective decision confidence," Mimeo.

Becker, G. M., M. H. DeGroot, and J. Marschak (1964): "Measuring utility by a single-response sequential method," *Behavioral Science*, 9, 226–232.

Benjamini, Y. and Y. Hochberg (1995): "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society*, 57, 289–300.

Bewley, T. F. (1986): "Knightian Decision Theory: Part I," *Cowles Foundation discussion paper*, 807.

Butler, D. and G. Loomes (2007): “Imprecision as an account of the preference reversal phenomenon,” *The American Economic Review*, 277–297.

——— (2011): “Imprecision as an account of violations of independence and betweenness,” *Journal of Economic Behavior & Organization*, 80, 511–522.

Cerreia-Vioglio, S. (2010): “Maxmin Expected Utility on a Subjective State Space: Convex Preferences under Risk,” Mimeo Bocconi University.

Cerreia-Vioglio, S., D. Dillenberger, and P. Ortoleva (2015): “Cautious Expected Utility and the Certainty Effect,” *Econometrica*, 83, 693–728.

——— (2020): “An Explicit Representation for Disappointment Aversion and Other Betweenness Preferences,” *Theoretical Economics*, 15, 1509–1509.

——— (2022): “Caution and Reference Effects,” Mimeo Princeton University.

Cerreia-Vioglio, S., D. Dillenberger, P. Ortoleva, and G. Riella (2019): “Deliberately Stochastic,” *American Economic Review*, 109, 2425–45.

Cettolin, E. and A. Riedl (2019): “Revealed preferences under uncertainty: Incomplete preferences and preferences for randomization,” *Journal of Economic Theory*, 181, 547–585.

Chapman, J., M. Dean, P. Ortoleva, E. Snowberg, and C. Camerer (2022): “Econographics,” *Journal of Political Economy Microeconomics*, forthcoming.

——— (2023): “Willingness to Accept, Willingness to Pay, and Loss Aversion,” NBER Working Paper 30836.

Chew, S. H., B. Miao, Q. Shen, and S. Zhong (2022): "Multiple-Switching Behavior in Choice-List Elicitation of Risk Preference," *Journal of Economic Theory*, 204, 105510.

Cohen, M., J.-Y. Jaffray, and T. Said (1985): "Individual behavior under risk and under uncertainty: An experimental study," *Theory and Decision*, 18, 203–228.

——— (1987): "Experimental comparison of individual behavior under risk and under uncertainty for gains and for losses," *Organizational Behavior and Human Decision Processes*, 39, 1–22.

Condon, D. M. and W. Revelle (2014): "The International Cognitive Ability Resource: Development and initial validation of a public-domain measure," *Intelligence*, 43, 52–64.

Cubitt, R., D. Navarro-Martinez, and C. Starmer (2015): "On preference imprecision," *Journal of Risk and Uncertainty*, 50, 1–34.

Dekel, E. (1986): "An axiomatic characterization of preferences under uncertainty: Weakening the independence axiom," *Journal of Economic Theory*, 40, 304–318.

Dillenberger, D. and C. Raymond (2019): "On the consensus effect," *Journal of Economic Theory*, 183, 384–416.

Dost, F. and R. Wilken (2012): "Measuring willingness to pay as a range, revisited: When should we care?" *International Journal of Research in Marketing*, 148–166.

Dubourg, W. R., M. W. Jones-Lee, and G. Loomes (1994): "Imprecise preferences and the WTP-WTA disparity," *Journal of Risk and Uncertainty*, 9, 115–133.

——— (1997): "Imprecise preferences and survey design in contingent valuation," *Economica*, 681–702.

Dubra, J., F. Maccheroni, and E. A. Ok (2004): "Expected utility theory without the completeness axiom," *Journal of Economic Theory*, 115, 118–133.

Dwenger, N., D. Kübler, and G. Weizsäcker (2018): "Flipping a Coin: Evidence from Laboratory and Field," *Journal of Public Economics*, 167, 240–250.

Enke, B. and T. Graeber (2022): "Noisy Cognition and Intertemporal Choice," Mimeo, Harvard University.

——— (2023): "Cognitive Uncertainty," Mimeo, Harvard University.

Feldman, P. and J. Rehbeck (2022): "Measuring willingness to pay as a range, revisited: When should we care?" *Quantitative Economics*, 13, 761–786.

Gabaix, X. (2019): "Behavioral inattention," in *Handbook of Behavioral Economics: Foundations and Applications*, ed. by D. Bernheim, S. DellaVigna, and D. Laibson, Elsevier, vol. 2, 261–343.

Ghirardato, P., F. Maccheroni, and M. Marinacci (2004): "Differentiating ambiguity and ambiguity attitude," *Journal of Economic Theory*, 118, 133–173.

Gilboa, I., D. Schmeidler, F. Maccheroni, and M. Marinacci (2010): "Objective and subjective rationality in a multi-prior model," *Econometrica*, 78, 755–770.

Gneezy, U. and J. Potters (1997): "An experiment on risk taking and evaluation periods," *The Quarterly Journal of Economics*, 112, 631–645.

Kahneman, D., J. Knetsch, and R. Thaler (1991): “Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias,” *Journal of Economic Perspectives*, 5, 193–206.

Khaw, M. W., Z. Li, and M. Woodford (2021): “Cognitive Imprecision and Small-Stakes Risk Aversion,” *Review of Economic Studies*, 88, 1979–2013.

——— (2022): “Cognitive Imprecision and Stake-Dependent Risk Attitudes,” NBER Working Paper no. 30417.

Kircher, P., S. Ludwig, and A. Sandroni (2013): “On the Difference between Social and Private Goods,” *The BE Journal of Theoretical Economics*, 13, 1–27.

Machina, M. J. (1985): “Stochastic choice functions generated from deterministic preferences over lotteries,” *The Economic Journal*, 95, 575–594.

Masatlioglu, Y. and E. A. Ok (2005): “Rational Choice with status quo bias,” *Journal of Economic Theory*, 121, 1–29.

——— (2014): “A Canonical Choice Model with Initial Endowment,” *Review of Economic Studies*, 81, 851–883.

Miao, B. and S. Zhong (2018): “Probabilistic social preference: how Machina’s Mom randomizes her choice,” *Economic Theory*, 65, 1–24.

Ok, E. and H. Nishimura (2019): “Preference Structures,” Mimeo, New York University.

Ok, E., P. Ortoleva, and G. Riella (2012): “Incomplete Preferences under Uncertainty: Indecisiveness in Beliefs vs. Tastes,” *Econometrica*, 80, 1791–1808.

——— (2015): "Revealed (P)Reference Theory," *American Economic Review*, 105, 299–321.

Ok, E. A. and Y. Masatlioglu (2007): "A theory of (relative) discounting," *Journal of Economic Theory*, 137, 214–245.

Ortoleva, P. (2010): "Status Quo Bias, Multiple Priors and Uncertainty Aversion," *Games and Economic Behavior*, 69, 411–424.

Ortoleva, P. and E. Snowberg (2015): "Overconfidence in Political Behavior," *American Economic Review*, 105, 504–535.

Permana, Y. (2020): "Why do people prefer randomisation? An experimental investigation," *Theory and Decision*, 88, 73–96.

Quiggin, J. (1982): "A theory of anticipated utility," *Journal of Economic Behavior & Organization*, 3, 323–343.

Rubinstein, A. (2002): "Irrational diversification in multiple decision problems," *European Economic Review*, 46, 1369–1378.

Sopher, B. and J. M. Narramore (2000): "Stochastic choice and consistency in decision making under risk: An experimental study," *Theory and Decision*, 48, 323–350.

Wang, T., R. Venkatesh, and R. Chatterjee (2007): "Reservation price as a range: An incentive-compatible measurement approach," *Journal of Marketing Research*, 44, 200–213.

Zhang, X. and S. Zhong (2020): "Putting Preference for Randomization to Work," Mimeo.

Table 1: Main Questions

	Question type	Left option	Right option	Values
Q1	MPL	$0.5\ \$20, 0.5\ \$0$	$\$x$	$x \in [0, 20]$
Q2	MPL	$\$18$	$0.5\ \$x, 0.5\ \$0$	$x \in [18, 54]$
Q3	MPL	$0.5\ \$22, 0.5\ \$0$	$0.5\ \$x, 0.5\ \$4$	$x \in [14, 22]$
Q1r	r-MPL	$0.5\ \$20, 0.5\ \$0$	$\$x$	$x \in [0, 20]$
Q2r	r-MPL	$\$18$	$0.5\ \$x, 0.5\ \$0$	$x \in [18, 54]$
Q3r	r-MPL	$0.5\ \$22, 0.5\ \$0$	$0.5\ \$x, 0.5\ \$4$	$x \in [14, 22]$
Q4	MPL	$\$16$	$0.8\ \$x, 0.2\ \$0$	$x \in [16, 27]$
Q5	MPL	$0.25\ \$16, 0.75\ \$0$	$0.2\ \$x, 0.8\ \$0$	$x \in [16, 27]$
Q6	MPL	$\$14$	$0.8\ \$x, 0.2\ \$0$	$x \in [14, 25]$
Q7	MPL	$0.25\ \$14, 0.75\ \$0$	$0.2\ \$x, 0.8\ \$0$	$x \in [14, 25]$

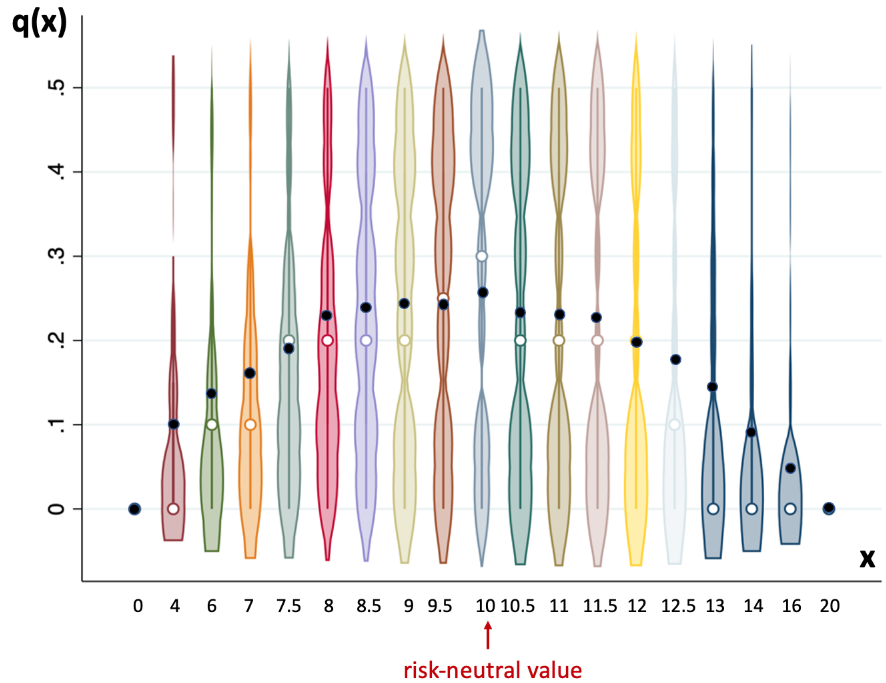
Notes: We denote by $p\ \$x, (1-p)\ \$y$ the lottery that pays $\$x$ with probability p and $\$y$ with probability $(1-p)$. In all questions, the Left option stays the same in all rows, while the Right option changes, with values of x increasing from the top row to the bottom. The last column indicates the range.

Table 2: Summary Statistics about Ranges and q–Measure

	Q1r ($\$20, \$0; 50\%$) vs $\$x$			**Q2r** $\$18$ vs ($\$x, \$0; 50\%$)			**Q3r** ($\$22, \$0; 50\%$) vs ($\$x, \$4; 50\%$)		
	range91	range64	q	range91	range64	q	range91	range64	q
% of Subjects with non-zero ranges	72% ($n = 106$)	53% ($n = 79$)	74% ($n = 109$)	74% ($n = 109$)	56% ($n = 83$)	76% ($n = 112$)	64% ($n = 94$)	51% ($n = 75$)	64% ($n = 95$)
Non-zero q-measure									
Average			0.35			0.38			0.35
Median			0.34			0.35			0.34
Non-zero ranges									
Av. size in $ (s.e.)	6.6 (0.34)	2.7 (0.20)		16.3 (0.88)	8.5 (0.79)		4.3 (0.22)	2.1 (0.19)	
Av. bottom $ (s.e.)	5.3 (0.30)	8.3 (0.29)		28.4 (0.57)	31.6 (0.55)		15.1 (0.15)	16.5 (0.19)	
Av. top $ (s.e.)	11.9 (0.23)	10.9 (0.24)		44.7 (0.74)	40.1 (0.78)		19.4 (0.20)	18.6 (0.21)	
Av. # rows (s.e.)	9.3 (0.43)	4.4 (0.28)		12.7 (0.58)	7.6 (0.57)		8.6 (0.44)	4.1 (0.39)	
Median/ Total # rows	9 / 19	4 / 19		14 / 23	6 / 23		9 / 17	3 / 17	

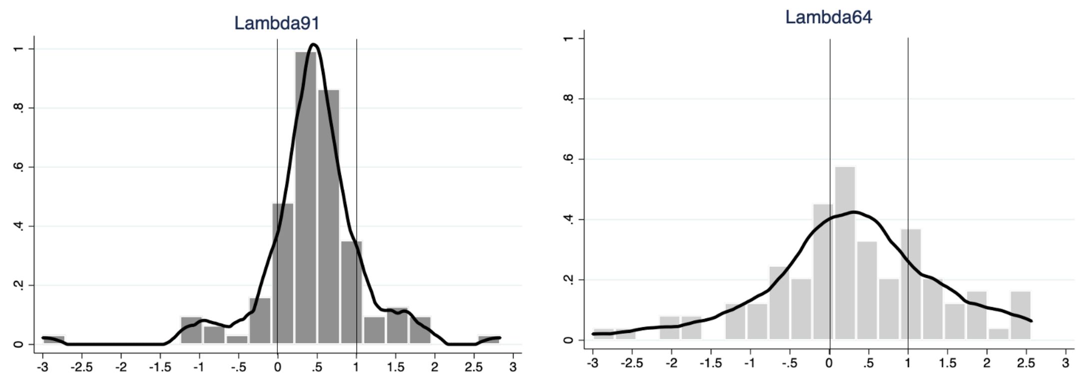
Notes: Participants who report intermediate numbers in the r-MPL question only in one row will have *range91*= 0 and *range64*= 0 but positive q. The third and fourth lines report the average and median values of the q-measure for participants with a positive q. The last five lines report the average values conditional on exhibiting ranges (standard errors in parenthesis). The last line also includes the total number of rows in each question.

Figure 1: Distribution $q(x)$ in each row of Q1r conditional on having a range91



Notes: These are violin graphs depicting the distribution of $q(x) = \min\{p(x), 1 - p(x)\}$ for each row in question Q1r. Black dots depict the mean values of q for each x, while hollow dots depict the median values.

Figure 2: Distribution of Average Values of λ_{91} and λ_{64}



Notes: Distribution of λ_{91} (left) and λ_{64} (right), focusing on values between -3 and 3 (this removes 2 outliers). The bottom of the range is the 'more risk-averse' behavior. For both panels, we first estimate λ for each question for each subject; then, we average across questions (for subjects with more than one estimated λ).

Table 3: Ranges, Individual Characteristics, and Non-Monotone Behavior

	Range 91			Range 64			q-measure
	Ind. Range	Range Freq.	Range Size	Ind. Range	Range Freq.	Range Size	average
Risk Q1	0.15 (0.21)	0.18 (0.11)	0.08 (0.57)	0.20 (0.10)	0.22 (0.08)	0.07 (0.65)	0.18 (0.11)
Risk Q2	-0.22 (0.08)	-0.22 (0.08)	-0.25 (0.06)	-0.14 (0.21)	-0.19 (0.11)	-0.21 (0.08)	-0.27 (0.06)
Risk Q3	0.08 (0.57)	0.09 (0.54)	0.02 (0.93)	0.14 (0.21)	0.16 (0.15)	0.01 (0.99)	0.06 (0.66)
Inv. Task 1	0.01 (0.93)	-0.02 (0.93)	-0.02 (0.93)	-0.001 (0.99)	-0.05 (0.72)	-0.01 (0.99)	-0.06 (0.65)
Inv. Task 2	-0.03 (0.89)	-0.02 (0.93)	-0.04 (0.86)	-0.01 (0.93)	-0.02 (0.93)	-0.01 (0.93)	-0.09 (0.51)
C-bias	-0.18 (0.15)	-0.17 (0.15)	-0.21 (0.10)	-0.23 (0.08)	-0.27 (0.06)	-0.28 (0.06)	-0.28 (0.06)
Non-Mon. Q1-Q7	0.16 (0.13)	0.14 (0.20)	0.25 (0.06)	0.14 (0.18)	0.15 (0.15)	0.34 (0.06)	0.23 (0.06)

Notes: Pairwise correlations are reported in parenthesis, with p-values adjusted for multiple hypotheses using Benjamini and Hochberg (1995) (see Anderson 2008). For C-Bias we have $n = 108$ observations (the number of subjects that report monotonic choices and have a switching point in Q4-Q7). Inv. Task 1 (2) is the number of tokens (out of 100) that a subject chose to keep and not invest in the risky project in investment task 1 (2), with higher numbers indicating a higher degree of risk aversion. The last column presents individual average q-measure across three questions Q1r, Q2r, and Q3r.

Online Appendix

A Additional Analysis

A.1 Steps in MPLs and r-MPLs

The list below lists the steps used in each of our MPL and r-MPL questions (all amounts are in US dollars):

- Q1 and Q1r:
  0,4,6,7,7.5,8,8.5,9,9.5,10,10.5,11,11.5,12,12.5,13,14,16,20
- Q2 and Q2r:
  18,21,24,27,30,33,33.50,34,34.50,35,35.50,36,36.50,37,37,50,38,38.50,39,42,45,48,51,54
- Q3 and Q3r:
  14,14.50,15,15.50,16,16.50,17,17.50,18,18.50,19,19.50,20,20.50,21,21.50,22
- Q4 and Q5:
  16,18,18.50,19,19.50,20,20.50,21,21.50,22,22.50,23,23.50,24,24.50,25,27
- Q6 and Q7:
  14,16,16.50,17,17.50,18,18.50,19,19.50,20,20.50,21,21.50,22,22.50,23,25

A.2 Other Types of Ranges

In the main body of the paper we focused on ranges '9-1' and '6-4.' Table A.1 reports the relevant measures for ranges '8-2', '7-3', and '5-5,' defined analogously. We find that 18% of subjects never report range82, 20% never report range73, and 29% never report range55. The remaining subjects report these ranges at least once. Specifically, 49% (42%, 14%) of subjects report range82 (range73, range55) in all three questions, 23% (28%, 26%) do so for two out of three questions, and 10% (11%, 31%) do so in one question.

Table A.1: Summary Statistics about Other Types of Ranges

	Q1r ($20,$0;50%) vs $x			**Q2r** $18 vs ($x,$0;50%)		
	range82	range73	range55	range82	range73	range55
% of Subjects with non-zero ranges	65% ($n = 97$)	61% ($n = 90$)	37% ($n = 55$)	71% ($n = 105$)	64% ($n = 95$)	39% ($n = 57$)
For non-Zero Ranges						
Av. size in $ (s.e.)	5.0 (0.28)	3.4 (0.21)	2.4 (0.28)	13.3 (0.83)	11.0 (0.82)	7.4 (0.99)
Av. bottom $ (s.e.)	6.6 (0.30)	7.7 (0.27)	8.5 (0.38)	29.6 (0.55)	31.0 (0.58)	32.0 (0.70)
Av. top $ (s.e.)	11.6 (0.23)	11.1 (0.23)	10.9 (0.25)	42.9 (0.72)	42.1 (0.81)	39.4 (0.93)
Av. # rows (s.e.)	7.7 (0.38)	5.7 (0.32)	3.8 (0.34)	11.0 (0.57)	9.5 (0.62)	6.6 (0.75)
Median/ Total # rows	7 / 19	5 / 19	3/ 19	12 / 23	9 / 23	4 / 23

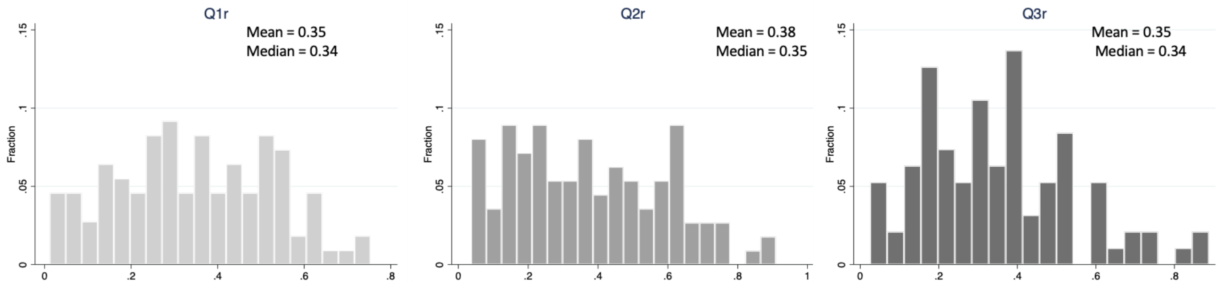
	Q3r ($22,$0;50%) vs ($x,$4;50%)		
	range82	range73	range55
% of Subjects with non-zero ranges	59% ($n = 87$)	54% ($n = 80$)	28% ($n = 42$)
For non-Zero Ranges			
Av. size in $ (s.e.)	3.7 (0.21)	2.6 (0.19)	2.0 (0.29)
Av. bottom $ (s.e.)	15.5 (0.15)	16.2 (0.17)	16.5 (0.28)
Av. top $ (s.e.)	19.2 (0.19)	18.8 (0.20)	18.5 (0.29)
Av. # rows (s.e.)	7.4 (0.42)	5.3 (0.39)	4.0 (0.58)
Median/ Total # rows	7/ 17	4.5 / 17	2/ 17

Notes: In each question, the last five lines report average values conditional on exhibiting ranges (standard errors in parenthesis). The last line also includes the total number of rows in each question.

A.3 Q-measures

Figure A.1 plots the distribution q-measures for each question.

FIGURE A.1: Distribution of q-measures for each question



Notes: We include only participants with positive q-measure.

A.4 Subjects with No Ranges and Monotone Behavior

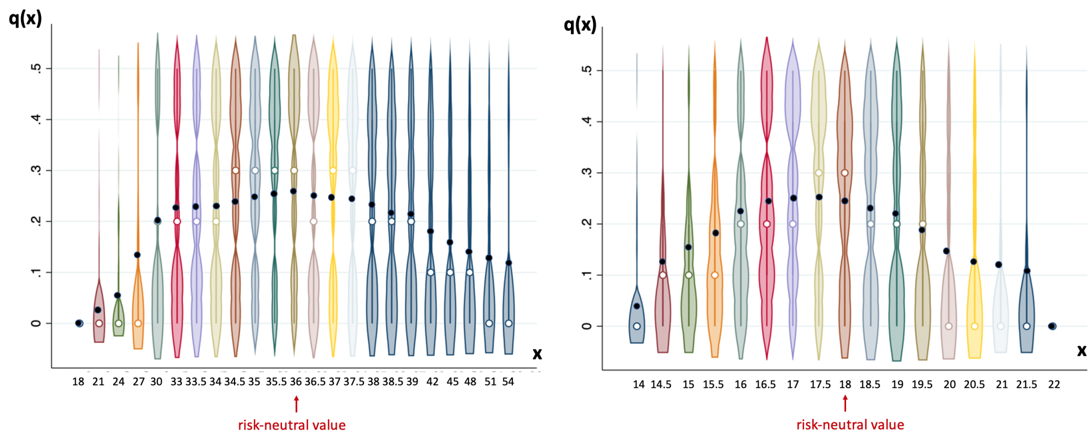
We expand our analysis on the comparison between choices in standard MPLs and range MPLs for subjects who did not exhibit ranges and are monotone in the standard MPL. Figure A.3 shows the kernel distributions of the differences between switching points in the two formats for each question. The distributions appear quite symmetric around zero; the value zero is in the 95% confidence interval of the estimated mean (for each question separately). For all three questions, we cannot reject the null that this difference is equal to zero at the standard 5% level. We obtain this result using both parametric two-sided T-test and non-parametric Wilcoxon matched-pairs signed-rank test (p-values for the two-sided t-test are $p = 0.09$, $p = 0.36$, and $p = 0.51$ for Q1, Q2, and Q3, respectively; for the Wilcoxon test, they are $p = 0.17$, $p = 0.11$, and $p = 0.83$ for Q1, Q2, and Q3, respectively). It is worth noting, however, that we only have a smaller number of subjects in these subgroups (36 for Q1-Q1r, 28 for Q2-Q2r, and 50 for Q3-Q3r), which means that these tests have relatively low power.

Figure A.4 shows the distribution of answers in regular MPLs for questions where subjects who do not exhibit ranges. The key observation is that these answers are not concentrated around salient values (e.g., the expected value) but rather are spread throughout.

A.5 Individual Estimates of Lambda

Figure A.5 depicts the distributions of λ_{91} and λ_{64} in each question.

Figure A.2: Distribution $q(x)$ in each row of Q2r and Q3r conditional on having a range91



Q2r question Q3r question

<u>Notes:</u> These are violin graphs depicting the distribution of $q(x) = \min\{p(x), 1-p(x)\}$ for each row in question Q2r and Q3r. Black and hollow dots denote the mean and median of q, respectively.

Figure A.3: Difference between choices in Qx and Qxr for subjects who are monotone in Qx and do not exhibit a range in Qxr

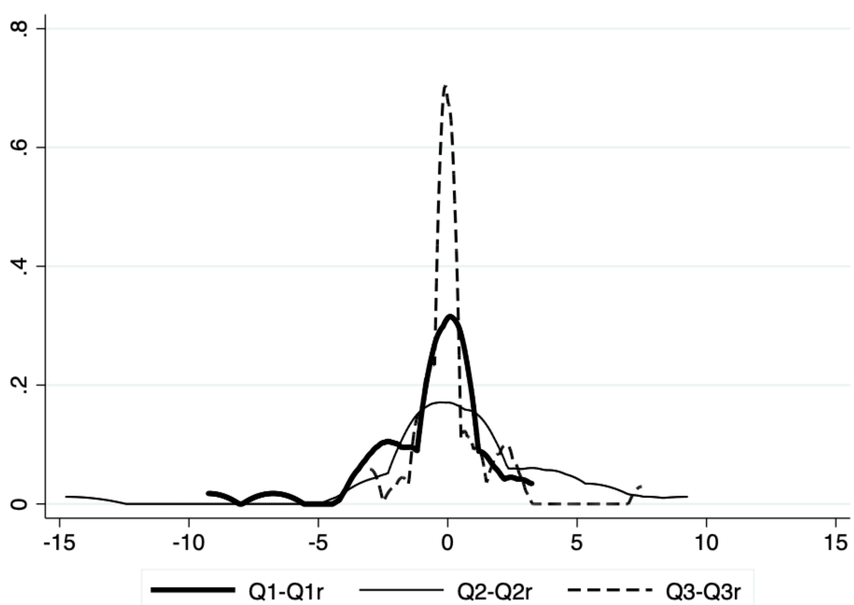


Figure A.4: Switching point in regular MPLs for monotone subjects who do not exhibit ranges in that question

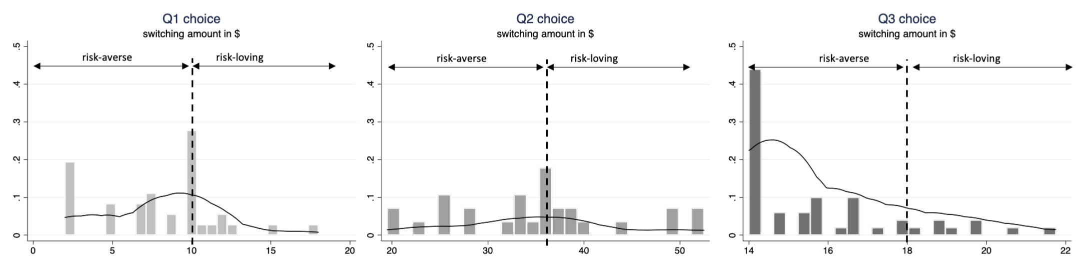
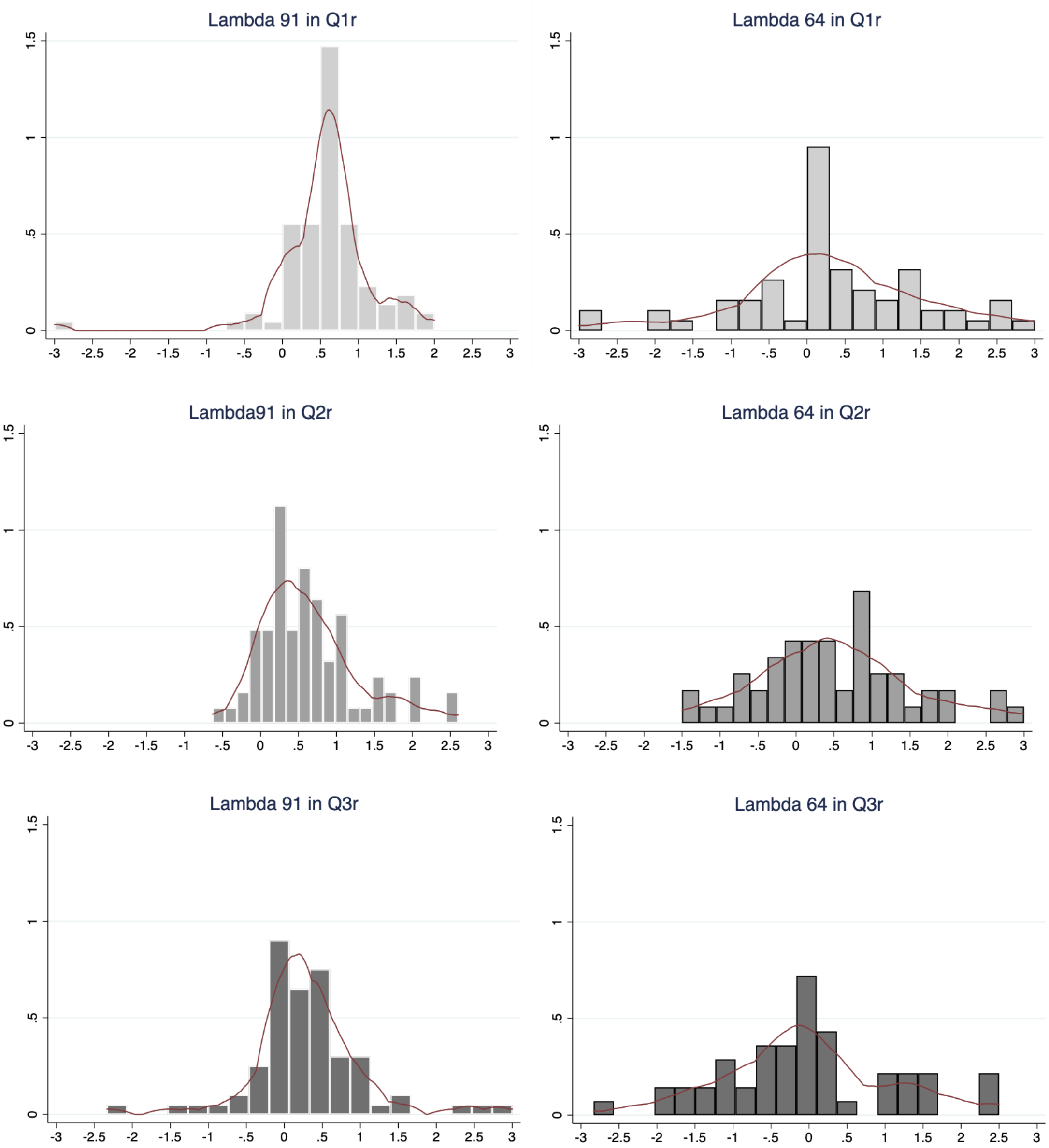


Figure A.5: Distribution of λ_{91} and λ_{64} for Q1r, Q2r, and Q3r



A.6 Alternative Measure of Range Size and Individual Characteristics

Table A.2 presents the equivalent of Table A.2 without the Benjamini and Hochberg (1995) corrections for multiple hypotheses.

Table A.3 repeats the relevant part of Table A.2 for an alternative definition of range size, where the normalization is made using the maximum possible range size.

A.7 Relation between Ranges, IQ and Overconfidence

We now explore more in detail the relation between the tendency to exhibit ranges and measures of IQ and overconfidence. Table A.4 presents pairwise correlations; all are uncorrected for multiple hypotheses testing (such corrections would only further reduce the significance).

We have two different measures of IQ: six matrices from the ICAR database and the three CRT questions. The total IQ is the average score across the two, measured by $\frac{\#\text{correct ICAR}}{6} + \frac{\#\text{correct CRT}}{3}$.

The overconfidence measures reported in the last four rows are computed following standard practice. Overestimation is the difference between how many ICAR questions a subject thinks she solved correctly minus how many she actually solved correctly. Overplacement is the reported rank minus actual rank in a sample of the 100 randomly selected adults in the US (obtained from Chapman et al. 2022). Overprecision 1 and Overprecision 2 are calculated based on the answers subjects give to the trivia question ("when was the land phone invented?") and the confidence that they have in their answer being correct. We follow the approach in Ortoleva and Snowberg (2015): first, we construct a measure of accuracy by taking the absolute value between the reported year and the actual year; then, we run a regression in which we try to predict confidence with a 4th-degree polynomial of accuracy. The residual is our measure of overprecision. Overconfidence 1 uses the confidence measure constructed from the qualitative question "how confident you are in your answer?" (admitting four possible answers), while overconfidence 2 uses the confidence measure constructed from the question "what is the probability that you answered correctly?".

Table A.4 shows no systematic relationship between IQ or overconfidence and the tendency to exhibit ranges, even without multiple hypotheses corrections.

A.8 Relation between Ranges and Questionnaire Answers

At the end of the experiment, we asked subjects the following question:

"In one of the Parts of the experiment you were asked to specify a number between 0 and 10 that determined the probability of receiving the left or the right option. Did you ever choose a number that was different from 0 or 10? If so, can you tell us why? Please elaborate if you can."

To interpret these answers and to correlate them with behavior in the experiment, we have hired a research assistant, who read through all the answers and classified them into the following categories:

- Category 0: No Response

- Category 1: NO
- Category 2: YES
  - Category 20: no explanation
  - Category 21: YES, desire to randomize
    - yes I did simply due to not wanting to place all my chances into one option
    - yes because I wanted to have a chance at both options
  - Category 22: YES, delegation
    - yes, I was unsure of my decision so I decided to let the computer choose
    - yes, because the probability allowed for the computer to take some responsibility for the outcome
  - Category 23: YES, for fun
  - Category 24: YES, indifference
    - the differences were not large enough

Table A.5 displays the coded answers and the individual-level types of participants that we recover in our analysis.

In our main sample of subjects, 127 out of 148 subjects (86%) answered the question and the remaining 14% did not. Among those who answered the question: 72% indicated that they used numbers others than 0 and 10 and gave more or less elaborate reasons for doing so, while the remaining 28% said that they did not use numbers others than 0 or 10.

Despite the fact that the questionnaire is not incentivized, we find that answers written by subjects are meaningful, consistent with their choices in the experiment, and informative about the mechanism underlying their decision. In particular, the vast majority of subjects that reported that they randomized between the Left and the Right options in some of the questions indeed did so (91%), while all subjects who reported not choosing an intermediate number indeed only used 0 and 10. The main motives for randomization are the 'desire to randomize' (category 21) and, to a much smaller extent, the 'desire to delegate' (category 22). Subjects in our sample almost never indicate that they were randomizing for fun or because of indifference between alternatives.

Table A.6 shows that there is a strong and significant correlation between the answers subjects provide in the questionnaire and their actual behavior in the experiment. In this table, we focus on subjects who answered the questionnaire (127 subjects). In particular, subjects who report in the questionnaire that they used ranges are very likely to use them, have higher average range sizes, and have higher average q-measure across all three questions. This holds true irrespective of which range measure we use (range91 or range64).

Overall, we find that subjects provide coherent reasons for reporting ranges, many of which are reminiscent of the mechanisms described by non-Expected Utility frameworks. Importantly, subjects' answers are consistent with their choices in the incentivized part of the experiment, which suggests that subjects are aware of their preferences and make these choices consciously.

B Structure of the Experiment and Order Effects

To investigate the possibility of order effects, we used two different orders across subjects with randomization at a session level; moreover, some parts had questions ordered randomly (at an individual level). Table B.7 illustrates the structure.

Table B.8 replicates our key summary statistics of ranges (similar to Table 2 in the main body of the paper), for Order A and Order B separately.

Results are broadly consistent and the main message of the paper holds in both cases. For example, among subjects facing Order A, we have 14% of subjects who never report range91 and 21% who never report range64. The remaining subjects report ranges at least once. Specifically, 74% (52%) of subjects report range91 (range64) in all three questions, 21% (33%) do so for two out of three questions, and 5% (16%) do so in one question. The distribution of types is similar for subjects facing Order B. Specifically, we have 28% of subjects who never report range91 and 36% who never report range64. Among the remaining subjects, 65% (31%) of subjects report range91 (range64) in all three questions, 28% (46%) do so for two out of three questions, and 7% (23%) do so in one question. Conditional on having ranges, the characteristics of the ranges (size, what is the bottom and what is the top, etc.) are very similar in the two orders.

Table A.2: Ranges, Individual Characteristics, and Non-Monotone Behavior

	Range 91			Range 64			q-measure
	Ind. Range	Range Freq.	Range Size	Ind. Range	Range Freq.	Range Size	average
Risk Q1	0.15 (0.11)	0.18 (0.04)	0.08 (0.37)	0.20 (0.03)	0.22 (0.02)	0.07 (0.45)	0.18 (0.04)
Risk Q2	-0.22 (0.02)	-0.22 (0.02)	-0.25 (0.01)	-0.14 (0.12)	-0.19 (0.04)	-0.21 (0.02)	-0.27 (0.00)
Risk Q3	0.08 (0.36)	0.09 (0.33)	0.02 (0.83)	0.14 (0.12)	0.16 (0.07)	0.01 (0.95)	0.06 (0.47)
Inv. Task 1	0.01 (0.86)	-0.02 (0.81)	-0.02 (0.82)	-0.001 (0.99)	-0.05 (0.53)	-0.01 (0.97)	-0.06 (0.45)
Inv. Task 2	-0.03 (0.69)	-0.02 (0.82)	-0.04 (0.66)	-0.01 (0.87)	-0.02 (0.78)	-0.01 (0.87)	-0.09 (0.30)
C-bias	-0.18 (0.06)	-0.17 (0.07)	-0.21 (0.03)	-0.23 (0.02)	-0.27 (0.00)	-0.28 (0.00)	-0.28 (0.00)
Non-Mon. Q1-Q7	0.16 (0.05)	0.14 (0.10)	0.25 (0.00)	0.14 (0.09)	0.15 (0.07)	0.34 (0.00)	0.23 (0.00)

Notes: Pearson pairwise correlations with significance level in the parenthesis (uncorrected for multiple hypotheses). For C-Bias we have $n = 108$ observations (the number of subjects that report monotonic choices and have a switching point in Q4-Q7). Inv. Task 1 (2) is the number of tokens (out of 100) that a subject chose to keep and not invest in the risky project in investment task 1 (2), with higher numbers indicating a higher degree of risk aversion. The last column presents individual average q-measure across three questions Q1r, Q2r, and Q3r.

Table A.3: Size of Ranges, Individual Characteristics, and non-monotone Behavior

	Size of Range 91	Size of Range 64
Risk Q1	0.13 (0.17)	0.11 (0.21)
Risk Q2	-0.28 (0.00)	-0.24 (0.01)
Risk Q3	0.02 (0.86)	0.01 (0.91)
C-bias	-0.21 (0.03)	-0.27 (0.01)
Inv. Task 1	-0.03 (0.69)	-0.03 (0.75)
Inv. Task 2	-0.06 (0.47)	-0.04 (0.60)
Non-Mon. Q1-Q7	0.23 (0.00)	0.36 (0.00)

Notes: Pearson pairwise correlations with significance level in parenthesis (uncorrected for multiple hypotheses). For C-Bias we have $n = 80$ observations (the number of subjects that report monotone choices and have a switching point in Q4-Q7). Inv. Task 1 (2) is the number of tokens (out of 100) that a subject chose to keep and not invest in the risky project in investment task 1 (2), with higher numbers indicating a higher degree of risk aversion.

Table A.4: Relation between Ranges, IQ and Overconfidence

	Range 91			Range 64		
	Ind. Range	Range Freq.	Range Size	Ind. Range	Range Freq.	Range Size
# correct ICAR	0.20 (0.02)	0.21 (0.01)	0.11 (0.17)	0.21 (0.01)	0.16 (0.05)	0.07 (0.41)
# correct CRT questions	-0.04 (0.65)	-0.07 (0.40)	-0.18 (0.03)	-0.06 (0.50)	-0.11 (0.18)	-0.19 (0.02)
total IQ	0.07 (0.41)	0.05 (0.54)	-0.08 (0.36)	0.06 (0.46)	-0.01 (0.99)	-0.10 (0.21)
Overestimation	-0.16 (0.05)	-0.18 (0.03)	-0.16 (0.06)	-0.11 (0.20)	-0.12 (0.14)	-0.10 (0.24)
Overplacement	0.14 (0.10)	0.17 (0.04)	0.20 (0.01)	0.12 (0.15)	0.12 (0.16)	0.14 (0.09)
Overprecision 1	0.13 (0.23)	0.22 (0.04)	0.13 (0.25)	0.16 (0.14)	0.19 (0.08)	0.01 (0.98)
Overprecision 2	0.10 (0.23)	0.11 (0.18)	0.12 (0.13)	0.13 (0.13)	0.11 (0.18)	0.08 (0.32)

Notes: Pearson pairwise correlations are reported alongside with p-values (uncorrected for multiple hypotheses). The average size of the range is computed as the weighted average of the size of the range in dollars weighted by the expected value of the left option in the question.

Table A.5: Coded Answers to the Questionnaire

	freq	no response	NO	YES	If YES, why? 20	21	22	23	24
Monotone MPL & No Ranges	$n = 25$	4	21						
Monotone MPL & Ranges	$n = 72$	10	5	57	11	31	9	1	5
Non-Monotone MPL & No Ranges	$n = 6$		5						
Non-Monotone MPL & Ranges	$n = 45$	6	4	35	5	25	2	1	2

Table A.6: Relation between Ranges and Answers in the Questionnaire

	Range 91		Range 64		Q-measure
	Ind. Range	Range Size	Ind. Range	Range Size	in at least one question
Answer ‘yes’ in the questionnaire	0.82 (0.00)	0.50 (0.00)	0.70 (0.00)	0.38 (0.00)	0.54 (0.00)

Notes: Pearson pairwise correlations are reported alongside with p-values. We focus on subjects who answered the questionnaire. The average size of the range is computed as the weighted average of the size of the range in dollars weighted by the expected value of the left option in the question. The last column correlates the answers in the questionnaire with the average q-measure across three questions Q1r, Q2r, and Q3r.

Table B.7: Two Orders of Questions

	Order A		Order b	
	questions	order	questions	order
Part I	Q1, Q2, Q3	random	Q6, Q7, Q4, Q5	fixed
Part II	Q4, Q5, Q6, Q7	fixed	Q1r, Q2r, Q3r	random
Part III	Risk1 and Risk2	random	Risk1 and Risk2	random
Part IV	Q1r, Q2r, Q3r	random	Q1, Q2, Q3	random
Part V	IQ + overconfidence	fixed	IQ + overconfidence	fixed
Part VI	Questionnaire		Questionnaire	

Notes: Random order indicates that the order of questions in this part of the experiment was randomized across subjects. Otherwise, fixed order was implemented for all subjects.

Table B.8: Summary Statistics about Ranges of Subjects in Two Orders

	Q1r ($20, $0; 50%) vs $x		**Q2r** $18 vs ($x, $0; 50%)		**Q3r** ($22, $0; 50%) vs ($x, $4; 50%)	
	range91	range64	range91	range64	range91	range64
ORDER A (73 subjects)						
% of Subjects	81%	67%	84%	64%	68%	56%
with non-zero ranges	($n = 59$)	($n = 49$)	($n = 61$)	($n = 47$)	($n = 50$)	($n = 41$)
For non-Zero Ranges						
Av. size in $ (s.e.)	6.73 (0.44)	2.70 (0.23)	16.06 (1.18)	8.13 (1.15)	4.80 (0.29)	2.48 (0.30)
Av. bottom $ (s.e.)	4.99 (0.40)	8.07 (0.35)	23.58 (0.73)	32.07 (0.69)	15.20 (0.20)	16.80 (0.28)
Av. top $ (s.e.)	11.72 (0.32)	10.77 (0.31)	44.64 (1.03)	40.20 (0.99)	20.00 (0.23)	19.29 (0.27)
Av. # rows (s.e.)	9.3 (0.58)	4.5 (0.35)	12.5 (0.77)	7.7 (0.77)	9.7 (0.57)	5.0 (0.61)
Median/Total # rows	9/19	5/19	13/23	6/23	11/17	4/17
ORDER B (75 subjects)						
% of Subjects	63%	40%	64%	48%	59%	45%
with non-zero ranges	($n = 47$)	($n = 30$)	($n = 48$)	($n = 36$)	($n = 44$)	($n = 34$)
For non-Zero Ranges						
Av. size in $ (s.e.)	6.46 (0.53)	2.59 (0.36)	16.54 (1.34)	8.91 (1.05)	3.65 (0.33)	1.55 (0.19)
Av. bottom $ (s.e.)	5.64 (0.47)	8.62 (0.50)	28.27 (0.90)	30.99 (0.90)	15.07 (0.23)	16.23 (0.26)
Av. top $ (s.e.)	12.10 (0.33)	11.21 (0.36)	44.80 (1.08)	39.90 (1.26)	18.73 (0.30)	17.78 (0.28)
Av. # rows (s.e.)	9.2 (0.66)	4.1 (0.44)	12.8 (0.88)	7.3 (0.86)	7.4 (0.65)	3.1 (0.38)
Median/Total # rows	10/19	4/19	14.5/23	5.5/23	7.5/17	3/17

Notes: The last five lines for each order report average or median values conditional on exhibiting ranges (standard errors in parenthesis). The last line for each order also includes the total number of rows in each question.

C Instructions

General Instructions. Welcome! This is an experiment designed to study decision-making. The instructions are simple, and if you follow them you may earn a considerable amount of money.

Please **turn off your cell phones** and **do not use them during the experiment**. Please do not talk with others. Also, please do not open any other applications or internet windows on the computer.

Structure of the Experiment. The main section of the experiment consists of **4 parts** with a total of **12 questions**. Once you are finished, we will ask you a few additional short questions. The experiment is thus very short. Please think carefully about each choice.

At the end of each Part, the computer will tell you to wait to proceed until prompted: please do so.

Let us highlight from the start that in the main part of the experiment **there are no right or wrong answers**. We are only interested in studying **your preferences**.

Lotteries. In many questions, we will ask you to choose between lotteries. Here is an example of a lottery:

50% chance of $10
50% chance of $5

This lottery pays either $10, with probability 50%, or $5, with probability 50%. To determine which, the computer will randomly draw an (integer) number between 1 and 100, where each number is equally likely to be drawn. If the drawn number is less or equal to 50, the lottery will pay $10. If the drawn number is above 50, the lottery will pay $5. Thus, it pays either $5 or $10 with equal probability.

Depending on the questions, the probabilities involved could be different: for example, they could be 25%, 75%, etc. In some cases, the lottery will involve no chance at all: for example, the option may just pay $12. In all cases, the outcome of lotteries will be determined by the computer using the probabilities specified.

Your Payment. Your payment consists of three components:

- First, the computer randomly chooses one of the 12 questions from the main part of the experiment. Each question is equally likely to be selected. Some questions will have several rows, in each of which you will be prompted to make a choice. If the selected question has more than one row in it, then computer also randomly chooses one of the rows in the selected question. Each row is equally likely to be selected. Your choice in the selected row of the selected question will be the first component of your final payment in this experiment.

- Second, you will receive additional payment for short questions that you will answer at the end of the experiment (after completing the main part of the experiment). You will see the exact instructions on how the short tasks will be paid on your screen.

- Third, you will receive $12 for showing up and completing the experiment

PART I. There are 3 questions in this part. Each question consists of several rows. In each row, there are two options: the Left Option and the Right Option. Here is an example of a question with 5 rows:

50% chance of $8 50% chance of $5	●	○	$5
50% chance of $8 50% chance of $5	○	○	$6
50% chance of $8 50% chance of $5	○	○	$6.50
50% chance of $8 50% chance of $5	○	○	$7
50% chance of $8 50% chance of $5	○	●	$8

For each row, you must **select one of the two available options**: the Left or the Right one.

Note: **the option on the Left is always the same**. The option on the Right instead changes: it **pays more money as we go down the rows**. This will be the case in **all** questions.

Also, note that in some of the rows, one of the two options pays as much, or more, than the other. This is the case in the first and in the last rows above:

- In the **first row**, the Left option pays either $8 or $5, while the Right option pays $5 for sure. Therefore, the Left Option pays at least as much as the Right Option.

- In the **last row**, the Left option pays either $8 or $5, while the Right option pays $8 for sure. Therefore, the Right Option pays at least as much as the Left Option.

In cases like these, the option that yields higher payoff will be **preselected** for you (indicated by the filled orange circle). You can change this if you wish.

Recall that each question is equally likely to be selected for payment, and that each row within a question is equally likely to be selected for payment.

Also recall that there are **no right or wrong answers**: we are only interested in studying **your preferences**. Finally, there are only **3 questions** in this part, so please think carefully about your answers.

Please raise your hand if you have questions.

Part II. [EXPERIMENTER SAYS IT OUT LOUD]. There are 4 questions in this part and each question consists of several rows. The instructions for the Part II are the same as the instructions for Part I of the experiment. You may proceed and answer the questions in Part II.

Part III. [EXPERIMENTER SAYS IT OUT LOUD]. There are 2 questions in this part. As you will see, these questions differ from the ones you have answered before. The instructions will appear on your computer screens. Please read those instructions carefully and answer the questions. You may proceed.

Risk 1. You are endowed with 100 points. Each point is worth 10 cents, so you are endowed with $10. You can choose to invest any amount between 0 and 100 points in a risky project. The remaining amount (points not invested in the risky project) is yours to keep. The risky project has a 50% chance of success:

- If the project is successful, you will receive 2.5 times the amount you chose to invest.
- If the project is unsuccessful, you will lose the amount invested.

Please choose the amount you want to invest in the risky project. Note that you can pick any amount between 0 and 100 points, including 0 or 100.

Risk 2. You are endowed with 100 points. Each point is worth 10 cents, so you are endowed with $10. You can choose to invest any amount between 0 and 100 points in a risky project. The remaining amount (points not invested in the risky project) is yours to keep. The risky project has a 40% chance of success:

- If the project is successful, you will receive 3 times the amount you chose to invest.
- If the project is unsuccessful, you will lose the amount invested.

Please choose the amount you want to invest in the risky project. Note that you can pick any amount between 0 and 100 points, including 0 or 100.

Part IV. There are **3 questions** in this part. Each question consists of several rows. In each row, there are two options: the Left Option and the Right Option. Here is an example of a question with 5 rows:

Just like before, the **Left option is always the same** in every row, while the **Right option changes**, getting better and better as you go down the rows.

In each row, your task is to indicate in the box an (integer) number between 0 and 10. This number determines the probability with which you get the Left and the Right options:

- If you select 10, you get the Left option **for sure** (probability 100% on Left option)
- If you select 9, you get the Left option with probability 90%, the Right option with probability 10%

50% chance of $8 50% chance of $5	$5	10
50% chance of $8 50% chance of $5	$6	
50% chance of $8 50% chance of $5	$6.50	
50% chance of $8 50% chance of $5	$7	
50% chance of $8 50% chance of $5	$8	0

- If you select 8, you get the Left option with prob. 80%, the Right option with probability 20%
- If you select 7, you get the Left option with probability 70%, the Right option with probability 30%
- ...
- If you select 2, you get the Left option with probability 20%, the Right option with probability 80%
- If you select 1, you get the Left option with probability 10%, the Right option with probability 90%
- If you select 0, you get the Right option **for sure** (probability 0% on Left option)

In general, the higher the number, **the higher the probability you receive the option on the Left**. Which option you receive will be determined by the computer following the number you specified.

Like in Part I, in some of rows one of the two options pays at least as much as the other. This is the case in the **first** and the **last row** of our example above. In these cases, the numbers 10 or 0 will be pre-entered for you. You can change these numbers if you like.

Recall that there are **no right or wrong answers**, we are only interested in studying **your preferences**; and that there are only **3 questions** in this part, so please think carefully about your answers. Please raise your hand if you have questions.

Part V. [EXPERIMENTER SAYS IT OUT LOUD]. This part of the experiment consists of a series of short questions. The instructions for each question are on your computer screens. Please read those instructions carefully and answer the questions. You may proceed.

IQ 1 - IQ 6. Subjects were asked to answer 6 IQ questions from the ICAR database (Condon and Revelle, 2014), three Matrix reasoning ones (reminiscent of Raven tests) and three Three-dimensional rotation; these are the same tasks used in (Chapman et al., 2022). In each of these questions, subjects are presented with the visual geometric design with a missing piece. The task is to find the missing piece.

Overconfidence 1. Think about the last 6 puzzles you solved. How many of them do you think you answered correctly?

Overconfidence 2. Think again about the last 6 puzzles. Now think about 100 typical people in the United States. Where do you think you rank in terms of how many correct answers you got? For example,

- if you think you got the most correct, you should answer 1;
- if you think you got the least correct, you should answer 100.

Overconfidence 3. Think about the wired telephone (landline). What year was the telephone invented? We are interested in your best guess, so please do not look this up if you do not know. Please type the year in which the wired telephone was invented.
How confident are you of your answer to the previous question, in which we asked you to specify the year in which the wired telephone was invented?

- No confidence at all
- Not very confident
- Somewhat unconfident
- Very confident
- Certain

What do you think the probability is (from 0%, or no chance, to 100%, or certainty) that your answer to the question in what year the wired telephone was invented is within 25 years of the correct answer?
Please type the number between 0 and 100 indicating the percentage chance that your answer is within 25 years of the correct answer.

CRT 1. A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball casts in cents? If you answer correctly, you receive 10 cents. Please enter your answer in cents.
CRT 2. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? If you answer correctly, you receive 10 cents. Please enter your answer in minutes
CRT 3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? If you answer correctly, you receive 10 cents. Please enter your answer in days