Why Is Choice Stochastic? Deliberate Randomization vs. Strength of Preference*

Carlos Alós-Ferrer[†1,2], Johannes Buckenmaier[3], Michele Garagnani[4], and Aldo Rustichini[5]

[1]Lancaster University Management School, Lancaster University. Lancaster LA1 4YD, UK
[2]Department of Economics, University Jaume I, Castellón, Spain
[3]Independent researcher. Zurich (Switzerland)
[4]Brain, Mind & Markets Laboratory, Department of Finance, The University of Melbourne, Melbourne, Australia
[5]Department of Economics, University of Minnesota Twin Cities, Minneapolis, USA

November 16, 2025

Accepted for publication in Management Science.

This is an author-generated version of the accepted manuscript.

Abstract

Choice inconsistency is ubiquitous in economic decision making. Standard random utility models, as well as sequential-sampling models, account for it and predict a relation between inconsistency rates and utility differences (strength-of-preference effects). The latter phenomenon is well-established empirically. However, a recent literature has argued that inconsistency might be the outcome of a deliberate attempt to implement an explicit preference for randomization. We collected new data in three pre-registered experiments and also re-analyzed the previous evidence for this explanation. Our evidence is generally well-explained by strength-of-preference effects. That is, people are less consistent for choices that are closer to indifference. When an option to randomize is made explicit and salient, this option is sometimes chosen. However, in a within-subject design, we statistically reject the hypothesis that explicit mixing in one-shot decisions (revealed preference for randomization) equal choice proportions in repeated decisions (choice inconsistency). The deliberate choice of randomization devices is influenced both by explicit costs (if any) and whether decision makers are close to indifference between the non-randomized alternatives. The latter effect might create a correlation with choice inconsistency at the aggregate level. **JEL Classification:** D87 · D91 · C91

Keywords: Stochastic Choice; Randomization; Strength of Preference; Deliberate Randomization

*We thank Songfa Zhong, three anonymous reviewers and the associate editor, and seminar participants at New York University Abu Dhabi, the Foundations of Utility and Risk conference in Ghent 2022, the IAREP conference in Nice 2023, and the Second Psychoeconomics Workshop in Lancaster 2025 for helpful comments and suggestions. Financial support from the Spanish Ministry of Science and Innovation under grant PID2022-136977NB-I00 is gratefully acknowledged.

†Corresponding author. c.alosferrer@lancaster.ac.uk

1 Introduction

People make inconsistent choices. Even when making the same decision repeatedly, they often choose different options (e.g., Davidson and Marschak, 1959; Tversky, 1969; Camerer, 1989; Hey and Orme, 1994; Hey, 2001; Moffatt, 2005; Alós-Ferrer et al., 2021). This widely-established empirical fact is readily captured in standard random utility models as used in economics and other fields (henceforth RUMs; McFadden, 1974, 2001; Anderson et al., 1992). It is also predicted by sequential-sampling models from psychology and neuroscience (Ratcliff, 1978; Ratcliff and Rouder, 1998; Shadlen and Kiani, 2013; Shadlen and Shohamy, 2016), which are gaining acceptance in economics (Fudenberg et al., 2018; Clithero, 2018; Webb, 2019; Chiong et al., 2024). The latter class of models capture the noisy (neural) computations involved in sorting out the decision values of alternatives, offering a mechanistic explanation for the origins of choice inconsistency. In contrast, a recent series of works has proposed that observed choice inconsistency might rather reflect a deliberate attempt by decision makers to optimally randomize among options (Agranov and Ortoleva, 2017, 2022, 2023; Dwenger et al., 2018; Feldman and Rehbeck, 2022). In this contribution, we examine this alternative explanation and provide new evidence showing that standard accounts involving noisy (unconscious) computations suffice to account for both choice inconsistency and the explicit decision to randomize when this option is made available and salient in a decision problem. We remark right away that we do not contest the existence of preference for randomization, e.g. when people are given the explicit choice to randomize. The point we make is that inconsistency, as typically observed in the empirical literature, is well-explained by standard psychophysical regularities.

The explicit-randomization account postulates that choice inconsistencies arise from an explicit preference for randomization. It is based on a theoretical approach that endows decision makers with convex preferences in a larger space where alternatives are now the possible probabilistic mixtures between the original alternatives (which could themselves be lotteries). By convexity, the decision makers' optimal decisions will often be nondegenerate mixtures, i.e. compound lotteries (if the original alternatives were lotteries themselves). The observed choice stochasticity then reflects the decision makers' attempt to implement a compound lottery over final outcomes, which is not directly available in the actual choice set. The original theoretical approach was put forward by Machina (1985). See Miao and Zhong (2018), Cettolin and Riedl (2019), Cerreia-Vioglio et al. (2015), Cerreia-Vioglio et al. (2019), and Chew et al. (2022) for further work in this direction. Those works provide theoretical models that can be interpreted in an "as if" fashion, i.e. as capturing behavioral patterns without ascribing actual intentions to decision makers. Recent work has gone further to propose that choice inconsistency is a deliberate attempt by decision makers to implement their preferences for randomization in environments where simple choices are repeated (Agranov and Ortoleva, 2017, 2022, 2023; Dwenger et al., 2018; Feldman and Rehbeck, 2022).

In this contribution, we contrast this recent account with the stochastic choice literature, which endorses the view that utility differences are directly related to choice probabilities (e.g., Debreu, 1958; Luce, 1959), moving away from the neoclassical view that utilities reflect preferences of an exclusively ordinal nature (Hicks and Allen, 1934). In this account, choice inconsistencies are linked to positive error probabilities, which in turn are linked to *strength-of-preference effects*. For instance, in the classical Luce model, which is equivalent to logit choice (Luce, 1959; Anderson et al., 1992), choice probabilities are proportional to decision values (utilities), and hence the probability of choosing an alternative which is near optimal but not optimal is larger than the probability of choosing an alternative whose decision value is further away from the optimum. Of course, this approach implies a cardinal relation between utilities and choice probabilities. For example, Debreu (1958) *defines* a utility function as a real-valued function u over alternatives such that the probability of choosing a over b is larger than the probability of choosing c over d if and only if the utility difference $u(a) - u(b)$ is larger that the utility difference $u(c) - u(d)$. Those differences are often informally referred to as "strength of preference" (e.g., Fishburn, 1988, p.4).

Utility-based models of probabilistic choice can of course also be interpreted in an "as if" way, without giving any meaning to the utilities or utility differences. Strength-of-preference effects, however, have been widely documented in psychology and neuroscience (e.g., Cattell, 1902; Dashiell, 1937; Moyer and Landauer, 1967; Laming, 1985; Dehaene et al., 1990; Wichmann and Hill, 2001), supporting a more literal interpretation. Specifically, error rates (or inconsistency rates) in a wide variety of tasks have been shown to be larger whenever the decision values of the options are closer, hence harder to tell apart (harder choices). These observations start with discrimination tasks where decision values are readily observable, e.g. the length of different lines, the brightness of light sources, or the number of dots in a moving cloud. Strength-of-preference effects have also been shown in incentivized versions of discrimination tasks (Duffy and Smith, 2025) and in decisions under risk, starting with an early contribution by Mosteller and Nogee (1951) and following with the recent work of Alós-Ferrer and Garagnani (2021, 2022a,b). Related evidence from neuroscience suggests that this pattern arises because computational processes in the brain are noisy and more likely to produce errors when the differences in decision values are small (Roitman and Shadlen, 2002; Glimcher et al., 2005; Padoa-Schioppa and Assad, 2006; Shadlen and Kiani, 2013; Shadlen and Shohamy, 2016). Under this interpretation, the reason for choice inconsistencies is inherent neuronal noise in the computation of values in the human brain.

It is important to note that strength-of-preference effects are incorporated in the very formulation of (additive) random utility models (RUMs). These models are widely used in economics (McFadden, 1974, 2001; Anderson et al., 1992), marketing (e.g., Baltas and Doyle, 2001; Feng et al., 2022), political science (e.g., Nownes, 1992; Karp, 2009), and many other fields. However, RUMs can be traced back to the probit model of Thurstone (1927), which explicitly aimed to capture strength-of-preference effects, and hence those effects are built in into the formulation of RUMs. To see this, recall that, in RUMs, an option x is chosen over another option y if $u(x) - u(y) + \varepsilon > 0$, where u is an underlying utility function and ε is an error term. Standard RUMs used in applications belong to the class of so-called Fechnerian models, which assumed a fixed error term distribution (this includes both logit and probit models; see, e.g., Moffatt, 2015). Then, if $u(x) > u(y)$, the probability of an error (choosing y) is the probability that $\varepsilon < -(u(x) - u(y))$, which is obviously larger if $u(x) - u(y)$ is small (close to zero) than if it is large (far away from zero). In other words, in the typical RUMs used in the applied microeconomics literature, errors are assumed to be more frequent when utility differences are smaller. Further, note that sequential sampling models, which imply strength of preference effects, can be recast as RUMs (Webb, 2019; Baldassi et al., 2020). The assumption of strength of preference effects in RUMs is also empirically justified. For example, Alós-Ferrer and Garagnani (2021, 2022a) demonstrate strength-of-preference effects in lottery choice by estimating utilities assuming a RUM using only part of a dataset and studying the relation between choice inconsistency and strength of preference using the remaining part of the dataset (i.e., out of sample).

The strength-of-preference account highlights a direct, monotonic relation between the strength of the subjective preferences and the proportion of inconsistent choices. That is, people are not randomly inconsistent. On the contrary, there is a systematic, predictable pattern in their inconsistency which is directly related to the difference in the (potentially subjective) values between the available options. In contrast, Agranov and Ortoleva (2017) (hereafter AO17) conducted an experiment to study preference for randomization and argued that, in their data, "differences in expected utility between the options [had] limited predictive power in determining the stochasticity of choice [...] and [could not] account for the variation in stochastic choice" (AO17, p. 42). Further, A017 viewed randomization as *deliberate* and conscious, while accounts as Machina (1985) can be interpreted in the spirit of rationalization of empirical regularities through "as if" preference models.

In this work, we report on three new experiments allowing to study choice inconsistencies in lottery choice. We show that the patterns in our data are well explained by the strength-of-preference account. Specifically, utility differences explain choice inconsistencies when an option to randomize is not explicit and repetitions are not consecutive. Further, they are also correlated with explicit randomization, i.e. the choice of an additional randomization device. However, in our data, we reject the hypothesis that explicit randomization choices (i.e., a revealed preference for randomization) determine the choice proportions in repeated decisions (i.e., choice inconsistency).

Before discussing our new experiments, we revisit empirical evidence from AO17 (Section 2) and show that the choice inconsistencies reported there (for non-consecutive repetitions) are actually also well-explained by the strength-of-preference account. The reason is a confound in the experimental design of AO17. The choice pairs in AO17's experiment are classified as "HARD" and "EASY," and, in a regression analysis including dummies for those choice pairs, the coefficients for utility differences are found not to be significant. However, AO17's design is such that HARD pairs entail very small utility differences compared to EASY pairs, and there are no pairs with intermediate utility differences. Hence, the effect of utility differences in their design is almost entirely captured by the EASY-HARD dummies. Thus, the evidence in AO17 is compatible with strength-of-preference effects. Our reanalysis of AO17's data shows that choice inconsistencies were directly linked to the differences in the values between the options in line with strength-of-preference effects, which speaks in favor of an account through choice errors.

Our three new experiments were designed to disentangle preference for randomization and strength-of-preference effects (Sections 3, 4, and 5). The first experiment considered a large number of lottery pairs per participant (57). A fixed part of the data (20 lottery pairs) was used to estimate individual utility functions, following an out-of-sample approach. The remaining lottery pairs were used to demonstrate strength-of-preference effects. These 37 pairs included the 7 lottery pairs used in the relevant part of AO17, which we show to be extreme in terms of (estimated) utility differences (HARD and EASY). Most of our lottery pairs, however (30 of the 37) are intermediate in this sense. Our analysis shows that choice inconsistencies are well-explained by utility differences, i.e. a strength of preference effect. An additional dummy capturing the difference between HARD and EASY lottery pairs show effects merely because those pairs are extreme in this dimension. That is, restricting the analysis to the 7 HARD and EASY pairs instead of the 37 in our dataset, the effect of utility differences disappears and is captured by the dummy instead.

Importantly, our Experiment 1 contains an additional task offering participants the explicit and costless choice to randomize. We find that strength-of-preference effects *also* successfully predict the frequency of choices to explicitly randomize when this is a salient option. That is, people are more likely to deliberately delegate their decision to a randomization device when they are closer to indifference. This shows that expected utility differences have predictive power accounting for the variation in deliberate randomization. It also implies that implicit and explicit randomization might share a common cause (strength of preference), which then can create a (possibly spurious) correlation between choice inconsistency and randomization choices at the aggregate level.

For this reason, we sought to test whether there is a direct connection between revealed preference for randomization and choice inconsistency in repeated decisions. The tasks in our Experiment 1 (repeated choices and explicit randomization) were matched in the sense that the choice frequencies in the first could exactly implement the randomization preferences revealed in the second. We can hence explicitly test the hypothesis that the frequencies observed through choice inconsistencies implemented the revealed preferences for randomization, and our test rejects this hypothesis.

In view of the results of Experiment 1, the relation between strength of preference and explicit randomization is crucial to understand the latter. Since the analysis in Experiment 1 is correlational, our second experiment aimed to confirm our findings using a causal manipulation. In the new design, participants were offered the explicit choice to randomize between two available options (a coin toss). Each participant faced the same 37 pairs as in Experiment 1, and 37 additional pairs (intermingled with the previous ones, in random order) derived from the previous ones through a "tax" manipulation which reduced all possible monetary outcomes by half. Thus, expected value and expected utility differences between the two lotteries were greatly reduced in the taxed choice pairs. For each decision, participants could either choose one of the lotteries, or delegate the decision to a coin toss. Hence, we expected participants to explicitly randomize (choose the coin flip) more for taxed pairs, for which strength of preference was reduced, compared to non-taxed ones. Our results show a clear increase in explicit randomization for taxed pairs, confirming our (pre-registered) hypothesis.

In Experiments 1 and 2, randomization devices were costless. It is hence natural to ask whether explicit randomization is still a relevant phenomenon when randomization devices entail an actual, monetary cost. In our third experiment, we causally manipulated the cost of a randomization device (again a coin flip). We aimed to show that people are less likely to explicitly randomize when this option is costly, even when the costs are small or negligible. As in Experiment 1, we included a block of choice pairs allowing to estimate individual-level utility functions (hence strength of preference). The study also included the 7 lottery pairs used in AO17 (and included in our two previous experiments), each repeated four times without a randomization option, and four times with the added option of delegating to a coin flip. The repetitions without added coin-flip options allow us to study implicit randomization, and we replicate the results of our Experiment 1. For the remaining choices, coin flips had varying monetary costs (including zero). We find a sharp decrease in explicit randomization when this option is costly, compared to when it is free. On the one hand, this confirms AO17's demonstration that some decision makers do deliberately randomize when this option is salient, even when it is costly. On the other hand, we find that actual decisions to randomize are empirically infrequent, and more so when there is an associated cost. In our third experiment, a costly coin-flip was only taken 4.5% of the time when available, and 165 participants (out of 221) never chose a costly coin flip.

Our evidence is also relevant for a number of theoretical approaches. First, as commented above, random utility models rely on strength of preference effects. These models are extensively used in applied work to estimate preferences from choice data through maximum likelihood methods. This stands on solid theoretical grounds only if the probability of choice errors is linked to distance to indifference. Second, sequential-sampling models originating in psychology, as the drift-diffusion model (Ratcliff, 1978; Ratcliff and Rouder, 1998) are receiving increased attention in economics (Fudenberg et al., 2018; Webb, 2019; Baldassi et al., 2020; Chiong et al., 2024) as a way to provide a microfoundation to probabilistic choice. Those models, however, predict a relation between utility or option-value differences and error probabilities, and hence are compatible with evidence on strength of preference, but cannot easily accommodate a deliberate randomization account.

A different class of models has proposed that people are inconsistent because of imprecision at the preference level, as in Cohen et al. (e.g., 1987) or Cubitt et al. (2015), where participants are explicitly asked to state how precise is their preference between two lotteries. This literature argues that preferences are less precise for choices closer to indifference (e.g., Cettolin and Riedl, 2019) and hence offer another mechanism compatible with strength of preference effects, but not with deliberate randomization.[1]

Last, a recent literature has proposed that stochastic choice arise due do imperfect perception of objective quantities (e.g., monetary outcomes) (Khaw et al., 2021; Frydman and Jin, 2022; Vieider, 2024; Garagnani and Vieider, 2025). As suggested in early contributions as Dashiell (1937) or Dehaene (1992), humans might perceive the numerical symbols used to represent choice options in an imperfect and stochastic way, which leads them to be inconsistent across repetitions of the same choices. These early discrimination experiments are at the roots of strength of preference effects, since errors are more frequent when the value of the numerical stimuli are closer. Khaw et al. (2021); Frydman and Jin (2022); Vieider (2024) and Garagnani and Vieider (2025) concentrated on these effects (i.e., differences in objective numerical quantities as opposed to differences in subjective utilities) as a source of imprecision. Thus, this literature provides an alternative microfoundation of strength of preference effects, but it is at odds with the deliberate randomization account, as inconsistency is derived from noise terms linked to how close stimuli are to each other, instead of an explicit preference on a space of mixtures.

The remainder of the manuscript is structured as follows. Section 2 reanalyzes the evidence in AO17. Sections 3, 4, and 5 present the design and results for our three experiments. Section 6 revisits the experiments to discuss evidence on dominated choices. Section 7 briefly discusses recent additional contributions to the preference for randomization literature. Section 8 concludes.

2 Revisiting Previous Evidence

In this section we reanalyze the evidence for deliberate randomization presented in AO17, focusing on the main experiment and main results of that paper. We briefly summarize the crucial points of the design below, and refer the reader to AO17 for further details.

2.1 Experimental Design

AO17's main experiment involved $N = 80$ participants and consisted of four parts. In the first and main part, participants faced binary choices including repetitions in random order. Ten pairs of

[1]This literature has also presented evidence for violations of completeness in economic choices, arguing that incompleteness could produce stochastic choices (Cettolin and Riedl, 2019; Nielsen and Rigotti, 2022; Halevy et al., 2023). However, without also incorporating a notion of strength of preference (i.e., without assuming that preferences would be "more incomplete" closer to indifference), violations of completeness alone would not explain our results.

lotteries were presented three times each (3 pairs were dropped in later parts of the experiment), but repetitions were never presented consecutively ("distant repetitions"). The idea is that inconsistencies across choice repetitions reflect an *implicit randomization* on the side of the participants. In addition, each pair was presented a fourth time while giving participants the additional, explicit option to choose a costly flip of a coin, the outcome of which would decide between the two options. This latter mechanism was introduced to allow for *explicit* (and costly) *randomization* between the lotteries.

For our purposes, the main difficulty in AO17's experiment and analysis is that their choice pairs were divided into three categories, called Easy, Hard, and FOSD (First Order Stochastic Dominance). This classification of stimuli is crucial for the interpretation of the results. However, as we will show below, the classification in Easy or Hard is closely linked to strength-of-preference effects.

The experiment of AO17 had three additional parts. The second part elicited the participants' risk attitudes using the procedure of Gneezy and Potters (1997). The third part mirrored the first, but this time participants were explicitly told about the repetitions and these were presented one after the other instead of far apart in the sequence of choices (see Subsection 2.3 below). The last part of the experiment involved measures of violations of expected utility, such as the Allais paradox.

2.2 Results and Reanalysis

The results of AO17 on implicit randomization reproduced the well-established fact that participants were inconsistent in their choices. 90% of participants were inconsistent at least once in the first part of the experiment, although, overall, more than 78% of choices in this part were fully consistent (that is, the decision maker chose the same alternative in the three repetitions of a given choice pair). Regarding explicit randomization, about 29% of participants chose to (costly) flip a coin at least once. In the third part, where the repetitions were presented in a row and hence the possibility to choose inconsistently was made salient, about 71% of participants were inconsistent at least once. The argument for the existence of a preference for deliberate randomization was based on the presence of inconsistent choices in the first and third parts.

The results, however, are very different between Hard and Easy pairs. The vast majority of choice inconsistencies (implicit randomization), and the vast majority of the decisions to delegate the choice to coin flips (explicit randomization), happened for Hard pairs ($> 33\%$) compared to Easy ($< 7\%$) or FOSD pairs ($< 7\%$). Also, inconsistencies in the third part (repetitions presented in a row) occurred exclusively for Hard pairs.

Given these stark differences, it is important to understand the criteria for the classification of choice pairs as Hard or Easy. To differentiate deliberate randomization from strength-of-preference effects, it is crucial that this classification is not confounded with differences in strength of preference. AO17 report that "Hard questions are not necessarily the questions in which the expected values, or utilities, are the closest" (AO17, p. 42) and "questions were designed in a way that allows us to separate these two forces [strength of preference and preference for randomization]" (AO17, p. 56). We contend that these claims are not correct. In actuality, Easy questions have much larger expected value differences than Hard ones, to the point that there is no overlap in the respective ranges. This can be clearly seen in Table 1 of AO17. The difference in expected values between alternatives in Easy pairs ranges from \$ 15 to \$ 60, while the ones for Hard choices are all between \$ 0 and \$ 7.25. AO17 argue that in Easy pairs some alternatives were "clearly better" than the other, while in HARD pairs

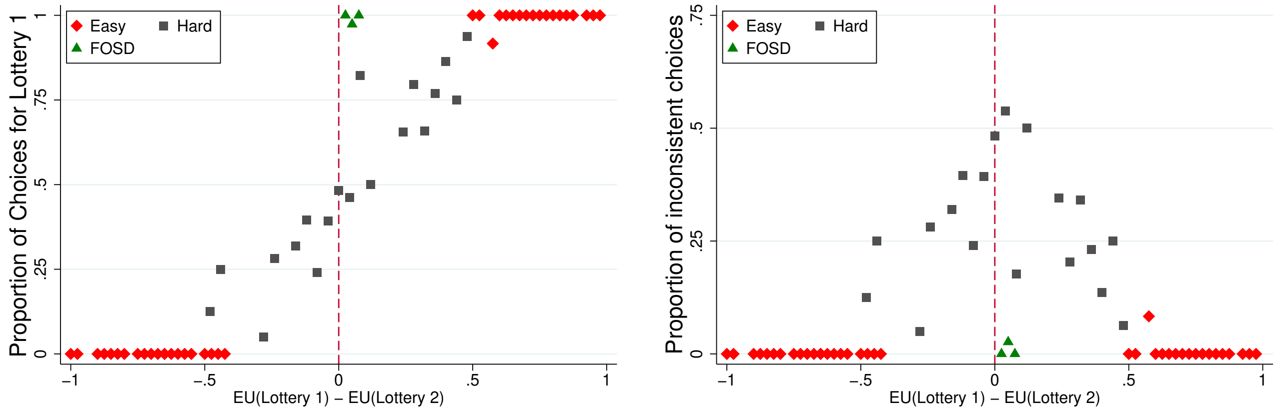

Figure 1: Reanalysis of AO17. Proportion of choices for Lottery 1 (left-hand panel) and proportion of consistent choices (right-hand panel) in part 1 of the experiment, as a function of the difference in expected utilities between Lotteries 1 and 2, using the risk attitudes estimated in part 2. For the graph, data is coded so that Lottery 1 has a (weakly) larger expected value than Lottery 2.

there was no "obvious winner" between two lotteries (AO17, p. 49). However, these are statements on strength of preference, i.e. how close are the utilities of the options.

The reason for this difference is that the classification in Hard and Easy pairs might have been inadvertently based on strength-of-preference effects. AO17 mention that the lottery pairs were selected according to the results of a pretest: "EASY questions were designed so that most participants choose the same option with a short response time; HARD ones instead had a choice distribution close to uniform and a long response time" (AO17, footnote 10, p. 49). However, it is well-known that strength-of-preference effects consistently affect error rates *and* response times. That is, in addition to being less consistent when the option values (or utilities) are closer, people are also slower, a phenomenon called "chronometric effect" (e.g. Dashiell, 1937; Moyer and Landauer, 1967; Moffatt, 2005; Alós-Ferrer et al., 2021; Alós-Ferrer and Garagnani, 2022a,b). Hence, selecting Hard and Easy choice pairs according to response times effectively created two groups of pairs with close and distant utilities, respectively. Thus, the classification in Hard and Easy pairs essentially corresponds to a dichotomization of the absolute value of the differences between the expected utilities of the lotteries in the pair, and the labels "hard" and "easy" could be replaced by "close utilities" and "far apart utilities."

Figure 1 illustrates the relation between the Hard and Easy classification and the differences in option values in the AO17 data. We use the second part of the experiment, which corresponded to the risk elicitation task of Gneezy and Potters (1997), to obtain an estimate of the risk attitude of each participant out of sample, assuming a CARA utility function (assuming CRRA instead does not change the results). The left-hand panel of the figure plots the proportion of choices for the first lottery in a pair as a function of the difference in expected utilities between the first and second lotteries. Data is recoded so that the first lottery in each pair is the one with the (weakly) largest expected value in the pair.[2] For convenience and ease of presentation, the right-hand panel re-plots the data to show the proportion of inconsistent choices (i.e., the frequency of the least-frequent option). The

[2]For the graph, utility differences are normalized to lie in the $[-1, 1]$ interval. For ease of presentation, the plot uses a binning procedure over the x-axis, with bins of width 0.01. That is, the y-value of each point represents an average for all observations with expected utility differences in the same bin (choice frequencies).

figure clearly shows that Easy pairs have both large utility differences (they are further apart from the zero on the horizontal axis) and low inconsistency, while Hard pairs have both small utility differences and high inconsistency. This is hence an illustration of standard strength-of-preference effects.

AO17 state that “[their] finding is not compatible with the many models of stochastic choice that imply a relation between stochastic choice and expected utility difference” (AO17, p. 56). However, the AO17 data display a clear relation between choice inconsistency and the difference in expected utilities between the alternatives. Therefore, the AO17 data can be taken as evidence for standard strength-of-preference effects.

AO17 also present a regression analysis (AO17, Table 4) which includes dummies for Easy and Hard pairs as well as a continuous variable reflecting the difference in expected utilities for each pair. However, given the discussion above, the Hard and Easy dummies are a dichotomization of the differences in expected utilities and hence strongly correlated with those. As a consequence, the fact that the Easy and Hard dummies are significant in AO17’s regressions is evidence for strength-of-preference effects. The observation that, in the presence of those dummies, the differences in expected utilities are often not statistically significant is a mechanical artefact of the correlation between the regressors and cannot be interpreted.

2.3 Explicit, Consecutive Repetitions

In the previous subsection, we have reanalyzed AO17’s data on implicit randomization when repetitions were not presented consecutively and participants were not warned about the existence of repetitions. The central argument and main contribution of AO17, however, is different. In the third part of their experiment, three repetitions of every choice pair were presented consecutively instead of far apart in the sequence of choices. Additionally, participants were explicitly warned about the repetitions. The purpose of this task was to document a preference for randomization when this option is made salient (by making the repetitions consecutive and warning participants), and compare the repetitions to explicit randomization decisions (coin flips). This is convincing evidence of a preference for randomization when this option is given explicitly and made salient.[3] Feldman and Rehbeck (2022) confirm these results in a similar experiment where binary lottery choices were repeated three times consecutively and subjects were explicitly warned about the repetitions. As in AO17, the decisions in consecutive, repeated choices were compared to explicit randomization decisions, in this case convex choice tasks.

AO17 also discuss whether sequential sampling models can explain their evidence or not, especially for the case of consecutive repetitions. A RUM can clearly explain inconsistencies in consecutive repetitions of a decision as long as each repetition involves a different realization of the random utility (the noise term). Analogously, a sequential-sampling model like the Drift-Diffusion Model (DDM) can account for choice inconsistencies in consecutive repetitions of a decision, as long as those allow for (noisy) information accumulation in every trial. In experiments were stimuli are varied in every trial (decision), the DDM typically is assumed to accumulate information starting either from a constant point or an independently-realized one (across trials). That is, information accumulated in one trial does not carry over to the next trial and, crucially, new information accumulation happens in every trial. The question is whether this is a reasonable assumption when the same stimuli are presented

[3]A possible criticism, however, is that making repetitions of a choice consecutive and explicitly warning participants of this fact makes salient that inconsistency is a choice and might create experimenter demand effects (Zizzo, 2010).

several times in a row. A standard interpretation of sequential sampling models (Shadlen and Kiani, 2013; Shadlen and Shohamy, 2016) is that subjective values in the brain are encoded with noise and the accumulation process reflects an internal sampling of information (e.g., from memory). Under this interpretation, new information would be sampled even if stimuli are repeated, which can then result in different decisions.

In contrast, AO17 (p. 42) consider different models where “the stochastic component of the utility does not change for in-a-row repetitions,” or, for the DDM, “that the agent does not collect more information for those repetitions.” This implies that the modified models are unable to explain inconsistencies in consecutively-repeated decisions. The additional assumption is not unreasonable. Originally, the DDM (Ratcliff, 1978; Ratcliff and Rouder, 1998) is a continuous-time implementation of the sequential probability test (Wald, 1945; Wald and Wolfowitz, 1945), which was designed to address the problem of when a decision maker should stop sampling information. Thus, one could consider the DDM as a model for conscious information accumulation, justifying the assumption that no further information is considered if the decision is encountered again.

AO17 suggest that the response times of their choice data might provide evidence in favor of their additional assumption. In their experiment, and especially for consecutive repetitions, participants took longer in the first repetition of a pair compared to later ones. AO17 argue that this happened because participants decided how to randomize when first seeing a choice pair and did not need to collect additional information in later repetitions. For the Drift-Diffusion Model, AO17(p.60–61) argue that the difference in response times between the first repetition and the other ones is “consistent with the assumption that subjects do not gather more information when they observe the same question asked repeatedly in a row.”

There are, however, alternative explanations for this evidence. It is well-known that participants in laboratory experiments very often become faster as the experiment proceeds (e.g., Chabris et al., 2009; Spiliopoulos and Ortmann, 2018). This is usually interpreted as an effect of increased familiarity with the interface. In the case of AO17’s experiment, a decrease in response times for consecutive decisions might also be explained by familiarity with the lotteries. Unfortunately, these two explanations cannot be fully disentangled, as they make the same predictions.

The observation that response times decrease with consecutive repetitions is also consistent with the fact that response times have both decision time components (accruing to the decision process) and other, more mechanical components (perceiving and processing the options, implementing the choice, etc.). For example, to account for these effects, standard implementations of the DDM include a *non-decision time* part of response times, which is usually found to be strictly positive (Ratcliff and Tuerlinckx, 2002). In terms of the DDM, familiarity with the lotteries can be accounted for through a reduction of the non-decision time, which does not affect choice stochasticity.

In any case, our focus is different. As shown in Section 2.2, choice inconsistencies on AO17’s data when repetitions are not consecutive are well-explained by standard strength of preference accounts. This also extends to response times. In AO17’s experiment, participants took longer to make a decision for Hard than for Easy pairs (AO17, Table 6). This is fully in agreement with the chronometric effect discussed above. That is, pairs where utility differences are smaller (Hard pairs) must result in longer response times. Thus, response times in AO17’s data also speak for clear strength-of-preference effects.

Our purpose is to show that choice inconsistency as typically observed in the literature and in economically relevant tasks can be well-explained by well-known psychophysical regularities (strength of preference), which are compatible with the view that the neuronal computation of values is inherently noisy. For this purpose, in the next three sections we present the data from our three new experiments, which focus on choice inconsistency for non-consecutive repetitions.

3 Experiment 1: Comparing Implicit and Explicit Randomization

The objective of our first experiment was twofold. First, we aimed to demonstrate the relation between strength of preference and choice inconsistencies with a large set of lotteries displaying a range of expected utility differences (and not just extreme values). Second, we included an additional task to elicit preferences for explicit randomization in a more fine-grained way than a coin flip. The objective was to test the statement that choice inconsistencies in repeated binary choices reflect explicit mixing choices in one-shot decisions where randomization is available. For this purpose, the explicit randomization task was designed to allow an exact match with every possible choice frequency in the repeated binary choices.

3.1 Experimental Design and Procedures

We collected data from $N = 103$ participants from the participant pool of students of the University of Zurich. The design was approved by the Human Subjects Committee of the Faculty of Economics, Business Administration and Information Technology at the University of Zurich. The experiment had two parts. In Part I, participants faced a series of choices repeated at random (nonconsecutive) trials. We used 37 lottery pairs, each repeated 4 times, presented in random order. 7 of those lottery pairs were identical to the ones used by AO17 in their third part, which were a subset of the 10 used in their first part. The remaining 30 lottery pairs were new lotteries designed to fill the gap between AO17's Easy and Hard pairs. The list of lottery pairs is in Appendix A. We used the same lottery format as in AO17. The experimental instructions and sample screenshots are in the Online Appendix.

An additional 20 lottery pairs, which were not repeated, were also included in this part. The corresponding choices were used to estimate individuals' risk attitudes following a standard microeconometric approach (see Appendix B for details on the estimation procedure). This was done because the task which AO17 used to elicit risk attitudes, taken from Gneezy and Potters (1997), has been shown to have limited predictive power for subsequent choices (Garagnani, 2023). Hence, it is unclear whether the poor predictive power of the expected utility variable in AO17's regressions is simply due to the choice of the elicitation task.

Part II of the experiment implemented a new mechanism to directly elicit the preferred compound lottery between the available options while allowing an exact match with each and every possible randomization implemented through choice inconsistency in Part I. That is, we aimed to test the explicit statement that choice inconsistencies in Part I reflect the attempt made by participants to approximate a compound lottery over final outcomes, which however was not available as such in the actual choice set of Part I. With the new mechanism in Part II, for each of the 37 (repeated) lottery pairs from Part I, participants directly chose a compound lottery by selecting a probability in the set $\left\{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\right\}$ for the first lottery in the pair, with the second lottery selected with the complementary probability. We explained the selection of the compound lottery using a choice among 5 urns, each of them containing four balls corresponding to either one or the other lottery (the labels A and B were used in the choice task), in the appropriate proportions. This was an intuitive, incentive-compatible way to directly elicit a preference over compound lotteries (see Online Appendix for an example of how this part was presented to participants).

For ease of presentation and analysis, in our dataset we recoded the lottery pairs so that the first lottery in every pair is the one with the highest expected value (or a predetermined one in case of ties, which happens for one of the pairs from AO17). Actual screen position was counterbalanced in the experiment. For each fixed pair and each participant, let $k_1 \in \{0, 1, 2, 3, 4\}$ be the number of times that the participant chose the lottery with the (weakly) highest expected value among the 4 repetitions of the choice pair in Part I. Define $\lambda_1 = \frac{k_1}{4}$. If $0 < \lambda_1 < 1$, we say that the participant exhibits stochastic choice for the given pair. Analogously, again for a given pair and participant, let $k_2 \in \{0, 1, 2, 3, 4\}$ be the number of balls corresponding to the lottery with the (weakly) highest expected value that are present in the urn chosen by the participant in Part II for the choice pair. Define $\lambda_2 = \frac{k_2}{4}$. We say that a participant reveals a preference for deliberate randomization for the given pair if $0 < \lambda_2 < 1$. Since λ_1 and λ_2 can take exactly the same values (0, 1/4, 1/2, 3/4, or 1), this allows to directly test for a preference for randomization when randomization is not an explicit option, because every convex combination explicitly chosen in Part II could also be exactly implemented through choice inconsistencies in Part I, and vice versa. This was not possible in the design of AO17, which used coin flips and three repetitions of the choice pairs.

Before the start of the experiment, participants were provided with general instructions and had to answer control questions to ensure their understanding of the concept of a lottery and its representation (see Online Appendix). In particular, we ensured through control questions that participants understood the mechanism in Part II of the experiment. Detailed instructions for each part were presented on-screen before that part started. At the end of the experiment, participants completed a short questionnaire on various demographics (gender, age, field of studies) and a measure of aversion toward compound lotteries (Gneezy and Potters, 1997). The latter was also used in AO17.[4] There was no feedback during the course of the experiment, that is, participants did not receive any information regarding their earnings until the very end of the experiment. All decisions were made independently and at a participant's individual pace.

To determine a participant's payoff, one lottery from each part of the experiment was randomly selected and paid. The total payoff from the experiment was the sum of the amounts received from both parts. In addition, participants received a lab-mandated show-up fee of CHF 10. Payment procedures were explained within the instructions and carried out truthfully. The average total remuneration was CHF 54.91. Sessions lasted about 80 minutes including instructions and payment. 53 (51.46%) of our participants were female and the sample had an average age of 23.94 years (median 24).

3.2 Results

3.2.1 Stochastic Choices

In Part I of the experiment, 97.09% (100 out of 103) of the participants exhibited stochastic choices for at least one pair. Figure 2 (left) displays the distribution of stochastic choices (number of pairs

[4]Participants decide which part of an endowment to invest in a risky asset whose value halves or is multiplied by 2.5 with equal probabilities. The question is presented twice. The first time, the realization of the risky asset's returns is presented as a simple lottery, the second time as a compound lottery. Following AO17, the indicator for aversion to compound lotteries is a dummy taking the value one if the participant invested less in the risky asset in the compound lottery case, and zero otherwise. See Additional Questions 1 and 2 in the Experimental Instructions in the Online Appendix.

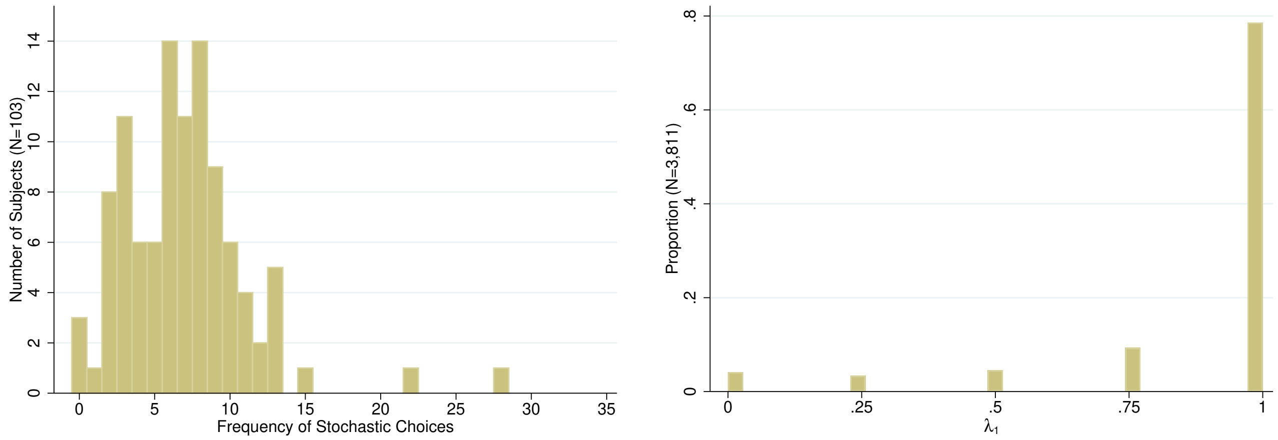

Figure 2: Experiment 1 (Part I). Left-hand panel: Distribution of the frequency of stochastic choices in Part I. Right-hand panel: Distribution of λ_1 in Part I.

for which each participant was inconsistent). Figure 2 (right) plots the raw histogram of λ_1-values in Part I. The overall frequency of stochastic choice is 18.79%, that is, for 81.21% (3095 out of 3811) of the lottery-pair observations in the entire dataset, the same choice was made in the four repetitions of the pair (i.e. λ_1 was either 0 or 1). This is a similar proportion to the one in AO17's experiment (about 78%). In our data, 50% of participants made stochastic choices for 6 or more pairs, and only 3 participants were fully consistent.

We used the 20 additional, non-repeated pairs in Part I to estimate individual risk attitudes through a random utility model (RUM) with CARA utility function (see Appendix B for details). The results are qualitatively unaffected by the particular utility function assumed (CRRA vs. CARA) and by assumptions on the shape of the noise, i.e., RUM vs. random parameter models (RPM; Loomes and Sugden, 1998; Apesteguía and Ballester, 2018). We then computed expected utility differences for each participant and each of the repeated choice pairs.

To quantify the relation between inconsistency and strength of preference, implicit randomization is measured by IR$= |\lambda_1 - 0.5|$, which is 0 for maximal randomization and 0.5 for fully consistent choice. As a first test, we found a strong and highly-significant positive correlation between the average individual IR and the average distance in expected utilities between the lotteries, across pairs (Spearman: $N = 37$, $\rho = 0.707$, $p < 0.001$). This is an indication of strength-of-preference effects.

Figure 3 plots the proportion of choices of the lottery with the highest expected value in a pair (left) and the proportion of inconsistent choices (right) as a function of expected utility differences in the choice pairs.[5] Choices are clearly more inconsistent for options that are closer in terms of expected utility differences (see below for a panel regression), supporting the strength of preference argument. We also replicate AO17's finding that stochastic behavior is more frequent for Hard than for Easy choices according to their classification. The individual averages are 50.49% for Hard and only 1.46% for Easy pairs (WSR: $N = 103$, $z = 8.733$, $p < 0.001$). However, as the figure makes transparent, this is a consequence of the more-general strength-of-preference effect. In particular, EU differences are clearly larger for Easy than for Hard pairs (WSR: $N = 103$, $z = 8.791$, $p < 0.001$). For the 30 new pairs in our design, the individual average stochastic behavior was 16.21%.

[5]As in Figure 1, for graphical purposes, the plot follows a binning procedure and normalizes utility differences to lie in the interval $[-1, 1]$.

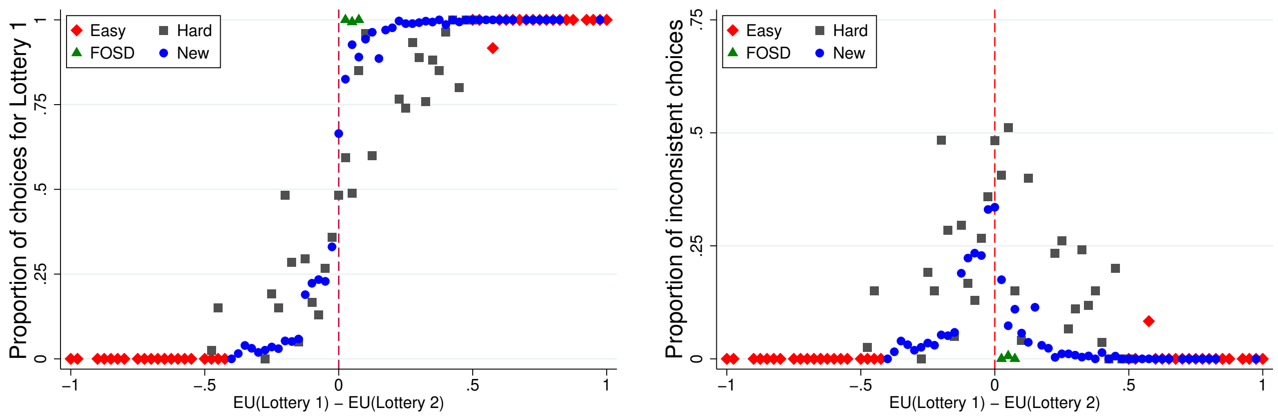

Figure 3: Experiment 1 (Part I). Left-hand panel: Proportion of choices for Lottery 1 in a pair as a function of the difference in expected utilities between the options in Part I of Experiment 1. Right-hand panel: Proportion of inconsistent choices as a function of the same difference. For the graph, data is coded so that Lottery 1 has a (weakly) larger expected value than Lottery 2.

Table 1: Random effects panel regression on $|\lambda_1 - 0.5|$ (Models 1 to 5) and $|\lambda_2 - 0.5|$ (Models 6 to 10).

	IR= $|\lambda_1 - 0.5|$					ER= $|\lambda_2 - 0.5|$				
	All pairs		AO17 pairs			All pairs		AO17 pairs		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
EU-Dist	0.177***	0.184***	0.219***	0.097**	0.111**	0.202***	0.211***	0.215***	0.054	0.108**
	(0.011)	(0.012)	(0.027)	(0.039)	(0.045)	(0.013)	(0.014)	(0.027)	(0.039)	(0.046)
FOSD				0.036	0.039				0.022	0.011
				(0.024)	(0.026)				(0.024)	(0.025)
Hard				−0.134***	−0.132***				−0.152***	−0.158***
				(0.019)	(0.020)				(0.018)	(0.019)
Number of Outcomes		0.002			0.001		0.002*			0.006*
		(0.001)			(0.003)		(0.001)			(0.003)
Aversion Compound Lot.		−0.005			−0.011		0.017			0.015
		(0.010)			(0.017)		(0.018)			(0.024)
Female		−0.024***			−0.020		−0.050***			−0.037**
		(0.007)			(0.012)		(0.013)			(0.018)
Age		0.001			0.003*		0.002			0.001
		(0.001)			(0.002)		(0.002)			(0.003)
Constant	0.410***	0.380***	0.357***	0.452***	0.379***	0.382***	0.338***	0.357***	0.472***	0.416***
	(0.004)	(0.029)	(0.008)	(0.021)	(0.054)	(0.007)	(0.052)	(0.010)	(0.022)	(0.073)
N	3811	3811	721	721	721	3811	3811	721	721	721
χ^2	239.100	255.047	65.523	244.687	251.864	258.559	277.967	64.746	291.550	301.910
R^2_o	0.054	0.064	0.083	0.246	0.255	0.043	0.072	0.061	0.238	0.250

Notes: Robust standard errors in brackets, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. For a given pair of lotteries, EU-Dist denotes the expected utility distance (absolute value of the EU difference) between the lotteries in a pair.

Table 1 (Models 1 to 5) replicates these results using a random effects panel regression, where we can further control for subjects' heterogeneity. The dependent variable is IR= $|\lambda_1 - 0.5|$, that is, the regressions examine the determinants of stochastic choice (or implicit randomization) in our data. Model 1 shows that strength of preference (EU-Dist, the absolute value of expected-utility differences) is a highly-significant predictor of choice consistency, a finding which is robust to including a number of controls (Model 2: gender, age, sum of the number of outcomes in the two lotteries in the pair, and aversion to compound lotteries measured by the task of Gneezy and Potters, 1997). Restricting to the Hard and Easy pairs that we took from AO17 reproduces this effect (Model 3). Adding a Hard (vs. Easy) dummy as a control (Models 4–5) just shows that Hard lottery pairs, which have smaller expected utility differences (in absolute value), are associated with larger inconsistency, hence confounding the measurement of the effect of expected utility differences. This just reflects the fact

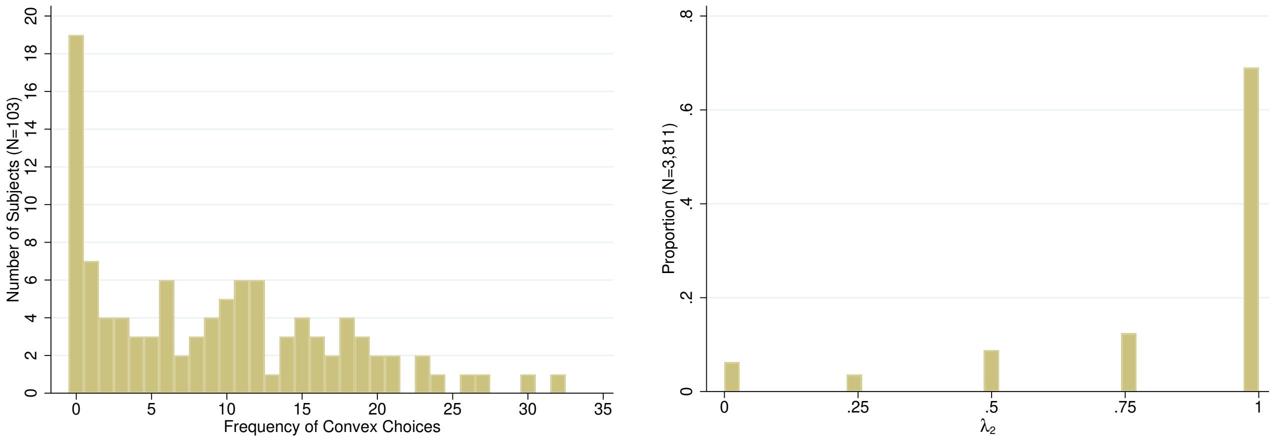

Figure 4: Experiment 1 (Part II). Left-hand panel: Distribution of the frequency of choices with explicit randomization in Part II. Right-hand panel: Distribution of λ_2.

that Hard-Easy dummies and expected utility differences are highly correlated. Models 4–5 also include a dummy for lottery pairs where a lottery was stochastically dominated (FOSD), which was not significant (see Section 6).

Our data hence speaks in favor of a strength of preference argument, and the difference between Easy and Hard pairs just identifies two extreme subsets of lottery pairs according to expected utility differences. AO17's observations are effectively embedded within our results, since our larger design and dataset also includes lottery pairs which are intermediate between Easy and Hard. Our new evidence, together with the reanalysis of the AO17 data in the previous section, is thus compatible with a strength-of-preference explanation of choice inconsistency.

3.2.2 Explicit Randomization

In Part II of the experiment participants revealed a preference for "deliberate randomization" ($0 < \lambda_2 < 1$) when this option was made explicitly available in 24.80% of all cases. That is, the deliberate choice to randomize is relatively infrequent (less than a quarter of the time) even when made explicitly available. As Figure 4 illustrates, 53.40% of participants chose to mix 9 times or less (out of 37 choice pairs), and 18.45% of participants never chose to mix. As was the case in Part I, we find that people chose to (explicitly) randomize more frequently for Hard than for Easy choices. The individual average is 49.27% for Hard pairs vs. only 1.46% for Easy pairs (WSR: $N = 103$, $z = 8.178$, $p < 0.001$). For the 30 new pairs, the individual average is 23.88%.

To study the relation between explicit randomization and strength of preference, and analogously to Part I, we quantify explicit randomization by ER$= |\lambda_2 - 0.5|$, which is 0 for maximal (explicit) randomization, i.e. choosing a 50% probability, and 0.5 for fully consistent choice. Crucially, there is again a strong and highly-significant positive correlation between the average individual ER and the average distance between expected utilities of the lotteries, across pairs ($N = 37$, $\rho = 0.655$, $p < 0.001$). As Figure 5 illustrates, and as in Part I of the experiment, also in Part II there is a clear relation between the proportion of deliberate choices to randomize between the options and their expected utility differences. In particular, this relation between EU differences and preference for randomization (resp. inconsistency) is at least as clear in Part II as is in Part I, as one can see

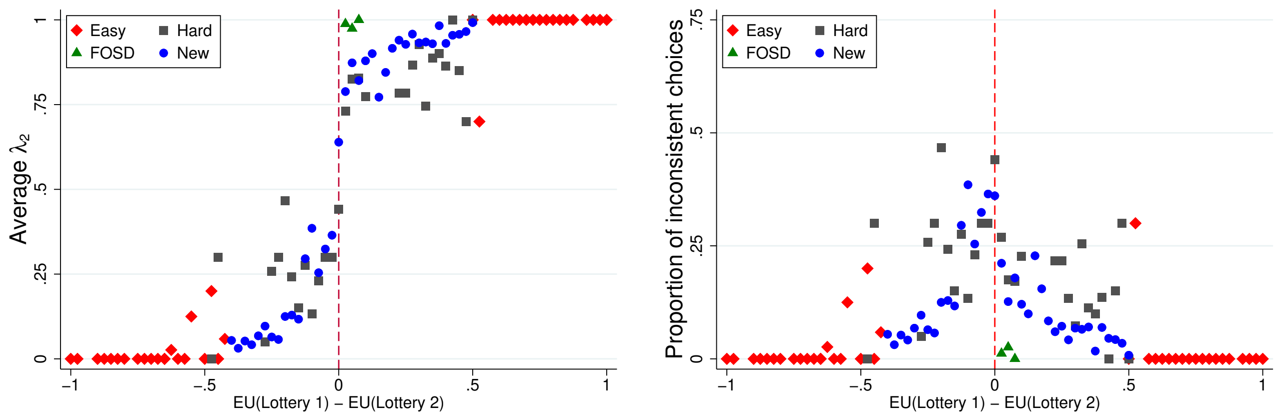

Figure 5: Experiment 1 (Part II). Left-hand panel: Average chosen mixing frequency for Lottery 1 as a function of the difference in expected utilities between the alternatives in Part II of Experiment 1. Right-hand panel: Proportion of inconsistent choices, defined here as the explicitly-chosen probability of the least-likely lottery, as a function of the same difference in expected utilities. For the graph, data is coded so that Lottery 1 has a (weakly) larger expected value than Lottery 2.

by comparing the respective right-hand panels of Figures 3 and 5. That is, in our data, strength of preference predicts the decision to deliberately randomize when this option is made explicitly available.

Table 1 (Models 6 to 10) replicates these results using a random effects panel regression, where we can further control for subjects' heterogeneity. The dependent variable is our measure of explicit randomization, ER$= |\lambda_2 - 0.5|$. Strength of preference (EU-Dist) significantly predicts explicit randomization, that is, explicit randomization is less frequent (larger ER) when the distance between the expected utilities in a pair is larger (Model 6). This result is robust to the inclusion of controls (Model 7) and survives restricting to the Hard-Easy pairs used in AO17 (Model 8). However, the result disappears if one includes a Hard (vs. Easy) dummy, again while restricting to the pairs used in AO17 (Model 9). This should of course not be interpreted, because the latter dummies are highly correlated with expected-utility differences. The main result is recovered even in the presence of the Hard-Easy dummy if one adds controls (Model 10). Models 9–10 also include a dummy for lottery pairs where a lottery was stochastically dominated (FOSD), which was not significant (see Section 6).

3.2.3 Deliberate Randomization?

Given that strength of preference predicts both choice inconsistency in Part I (Figure 3) and explicit randomization when this option is available in Part II (Figure 5), we must ask whether the deliberate choice to randomize when this option is available is predictive of stochastic behavior in repeated one-shot decisions.

Figure 6 indicates that there is indeed a correlation between the two parts of Experiment 1. The left-hand panel shows that there is a correlation between the individual-level average λ_1 and λ_2 (average across lotteries for each given individual; $N = 103, \rho = 0.775, p < 0.001$). That is, decision makers who are more inconsistent also tend to randomize more strongly when this is an explicit option.[6] The right-hand panel shows that the correlation also exists for the lottery-pair-level average λ_1 and λ_2 (average

[6] Feldman and Rehbeck (2022) find correlations of 0.33 and 0.42 between repeated binary choices and convex choice tasks, for two specific lottery pairs.

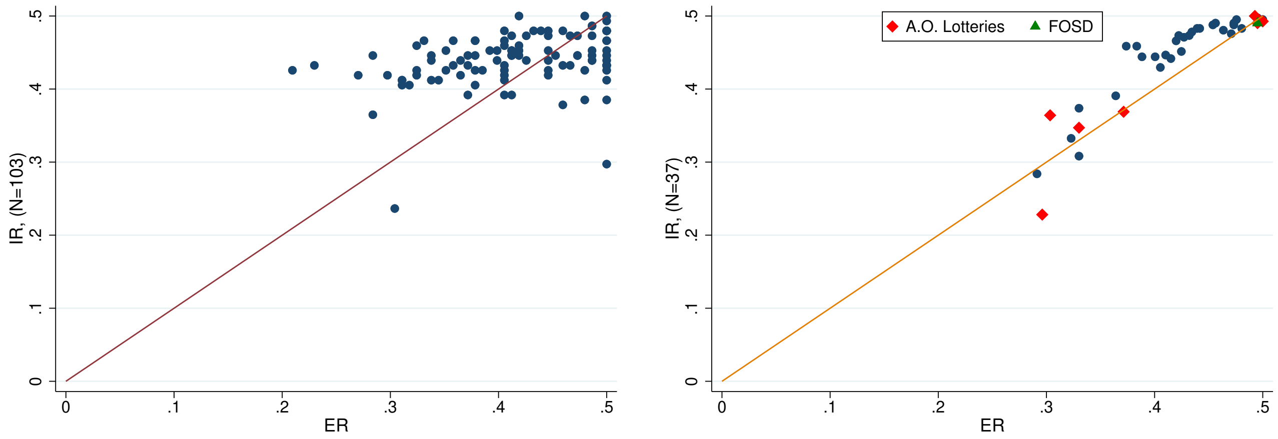

Figure 6: Experiment 1. Correlation between stochastic behavior (λ_1 and λ_2) at the individual (left-hand panel) and lottery (right-hand panel) level.

across individuals for each fixed choice pair; $N = 37, \rho = 0.974, p < 0.001$). That is, pairs eliciting more choice inconsistencies are also the pairs with stronger randomization when this is an explicit option. However, we find that choice inconsistency is systematically smaller than revealed preference for randomization, that is, IR$= |\lambda_1 - 0.5|$ is larger than the corresponding ER$= |\lambda_2 - 0.5|$, both when taking averages at the individual level (WSR: $N = 103$, average IR$= 0.440$, average ER$= 0.416$, $z = 3.421$, $p < 0.001$) and when taking averages at the lottery-pair level (WSR: $N = 37$, $z = 4.105$, $p < 0.001$). These results are reflected in both panels of Figure 6, where most data points are above the diagonal (especially when averaging at the lottery level).

These correlations illustrate that strength of preference successfully predicts both inconsistent choice and the choice of explicit randomization when such option is available. This is of course not enough to show that choice inconsistency reflects the decision makers' deliberate attempt to implement a preference for randomization, because the simple correlations between the two parts of Experiment 1 might be spuriously caused by a common effect, namely strength of preference.

A true preference for randomization in Part I would imply that participants try to implement their optimal compound lottery during the binary choice phase (Part I). This optimal compound lottery must then be revealed by direct choice when given the opportunity (Part II). Since λ_1 and λ_2 can only take (the same) 5 possible values each, the prediction if observed choice inconsistency exactly implements the revealed preference for randomization (i.e., $\lambda_2 \neq 0, 1$) would be that data in the space (λ_1, λ_2), for $\lambda_2 \neq 0, 1$, should be concentrated on the diagonal of the matrix describing all possible choices of a participant for each given choice pair. Experiment 1 was explicitly designed to allow for a statistical test of this hypothesis.

Table 2 displays the raw number of instances for each (λ_1, λ_2) combination. The vast majority of the observations where $\lambda_1 = \lambda_2$ correspond to cases where there is neither choice inconsistency nor explicit randomization (either $\lambda_1 = \lambda_2 = 1$ or $\lambda_1 = \lambda_2 = 0$). To test for the hypothesis that an explicit preference for randomization ($0 < \lambda_2 < 1$) explains observed choice inconsistency, we restrict to the 945 instances of explicit randomization ($0 < \lambda_2 < 1$) in Part II of the experiment. We compare the proportions of observations with $\lambda_1 = \lambda_2$ and with $\lambda_1 \neq \lambda_2$, conditional on $\lambda_2 \in \{0.25, 0.5, 0.75\}$ (i.e., actual explicit randomization). The set of observations where λ_1 agrees with λ_2 conditional on $\lambda_2 \in \{0.25, 0.5, 0.75\}$ makes up just $(23 + 45 + 76)/(136 + 337 + 472) = 15.24\%$ of observations. Thus,

$\lambda_2 \setminus \lambda_1$	0	0.25	0.5	0.75	1	Total
0	86	34	39	36	43	238
0.25	34	23	28	25	26	136
0.5	28	49	45	62	153	337
0.75	6	18	38	76	334	472
1	12	26	54	163	2,373	2,628
Total	166	150	204	362	2,929	3,811

Table 2: All instances of stochastic behavior in Part I (λ_1) and Part II (λ_2) of the experiment. Gray-colored cells indicate those occasions where randomization is chosen in Part II and the two values coincide, hence supporting a preference for randomization hypothesis.

the predicted value of $\lambda_1 = \lambda_2$ was not more frequent than $\lambda_1 \neq \lambda_2$. The effect is significant according to a one-sample test of proportions clustered by subject and adjusted by intraclass correlation of 0.107 ($p < 0.001$),[7] which rejects the null hypothesis that choice inconsistency reflects (revealed) deliberate randomization. Thus, in our data, deliberate mixing choices (revealed preference for randomization) do not explain choice inconsistency in repeated lottery choice.

4 Experiment 2: Manipulating of Explicit Randomization

In Experiment 1, we found that strength of preference predicted deliberate randomization in the sense that pairs with smaller (estimated) expected utility differences are associated with more randomization when this possibility is explicitly available. This result suggests that, as implicit randomization, also explicit randomization is at least partially determined by strength of preference.

Experiment 2 sought to further validate the role of strength of preference as a determinant of explicit randomization. Since the result in Experiment 1 is correlational, we aimed to confirm it using a causal manipulation. In the new experiment, participants were offered the explicit choice to randomize between two available options by tossing a coin. We aimed to show that people are more likely to (explicitly) randomize when they are closer to indifference by manipulating lottery pairs in a straightforward way.

The experimental manipulation was as follows. In a within-design, each participant was presented with the same 37 pairs as in Experiment 1 (but with an exchange rate of 10 points = 1 GBP). They could choose either option or delegate choice to a (costless) coin flip, i.e. a random device. Each participant also faced 37 additional pairs (intermingled with the previous ones, in random order) derived from the previous ones through a "tax" manipulation. Specifically, participants were presented with the exact same 37 pairs and the same (three) choices, but they were also told that for those choices, all possible outcomes from both lotteries would be reduced by 50%, independently of their choices.

The manipulation halves the expected value difference between the two lotteries in any choice pair. Under reasonable assumptions on the shape of utility functions, this also reduces the difference in expected utilities. For example, this is always the case for any risk-averse (or not too risk-seeking) CRRA utility function. To see this, let $u_i(x) = A \cdot x^\alpha$ for arbitrary constants A and α, with $\alpha > 0$. Let (p, q) be a non-taxed pair and (p', q') the corresponding non-taxed pair. The p lottery assigned a

[7] Alternatively, a Wilcoxon signed-rank test shows that, conditional on $\lambda_2 \in \{0.25, 0.5, 0.75\}$, observations with $\lambda_2 \neq \lambda_1$ are more frequent than those with $\lambda_2 = \lambda_1$ (average 1.71 observations with $\lambda_1 = \lambda_2$ vs. 9.54 with $\lambda_1 \neq \lambda_2$; WSR: $N = 84$, $z = 7.944$, $p < 0.001$).

probability of 0.25 each to four possible outcomes x_1, x_2, x_3, x_4. The q lottery assigned a probability of 0.25 each to four possible outcomes y_1, y_2, y_3, y_4. The p' and q' lotteries are identical except that the outcomes are halved. Then,

$$|EU(p') - EU(q')| = \left| \frac{A}{4} \sum_{k=1}^{4} \left[\left(\frac{x_i}{2}\right)^{\alpha} - \left(\frac{y_i}{2}\right)^{\alpha} \right] \right| = \left(\frac{1}{2}\right)^{\alpha} |EU(p) - EU(q)| < |EU(p) - EU(q)|$$

For other functional forms, the difference in expected utilities is also in general smaller for taxed pairs as long as individual utility functions are not too concave for small stakes.[8] Hence, we expected participants to explicitly randomize (choose the coin flip) more for taxed pairs compared to non-taxed ones.

4.1 Experimental Design and Procedures

We collected data from $N = 225$ participants from the participant pool of Prolific (45.78% females; median age 30). The experiment was preregistered at AsPredicted 186120 and obtained ethical approval from the Faculty Research Ethics Committee of the University of Lancaster, FASSLUMS-2024-4717-LeXeL-1. The experiment consisted of two within-subject treatments. That is, participants were presented with the same 37 lottery pairs twice in random order. These are the same lottery pairs we used in Experiment 1, and in particular encompass the 7 pairs used in AO17. For one of the two repetitions of each lottery pair, participants were told that there was a tax on the earnings from the current trial, such that all possible outcomes from both lotteries were reduced by a fixed percentage (50%) independently of their choices. The same lottery pair was never repeated in a row, i.e., there was at least one different lottery pair between repetitions of the same pair.

For each pair, participants could choose one of the two lotteries or select an option, described as a coin toss, which would automatically randomize between the alternatives for them. Selecting this option was costless for participants and there was no mention of any cost for selecting it. Our main dependent variable of interest is how many times participants selected to flip a coin.

There was no feedback during the course of the experiment. That is, participants did not receive any information regarding their earnings until the very end of the experiment. All decisions were made independently and at a participant's individual pace. To determine a participant's payoff, one lottery pair was randomly selected and paid. In addition, participants received a show-up fee of GBP 2.25. Payment procedures were explained within the instructions and carried out truthfully. The average total remuneration was GBP 6.66. The experiment lasted about 12 minutes including (online) instructions.

4.2 Results

We compare the proportion of chosen coin flips, i.e. the proportion of pairs for which participants chose to explicitly randomize, between taxed and non-taxed pairs. Taxed choices, which should correspond to decisions closer to indifference, resulted in more explicit randomization. The individual-average proportion of explicit randomization among the 37 pairs was 4.66% (median 0.00, SD 8.86, min 0.00,

[8] For the monetary amounts involved in our Experiment 2, it is reasonable to expect that utilities are actually close to linear, as the average payoff net of show up fee was 4.41 GBP. This is, however, an assumption, which could be violated for non-CRRA utility functions displaying large small-stakes risk aversion (see Rabin, 2000).

max 45.95) for non-taxed trials and 5.48% (median 0.00, SD 8.85, min 0.00, max 70.27) for taxed trials.

Since individual observations are proportions and they are paired (proportion of coin flips for taxed vs. non-taxed trials for each participant), we rely on a non-parametric Wilcoxon signed-rank test (WSR), which shows a significant effect ($N = 225$, $z = -3.463$, $p < 0.001$). The observed increase in coin flips (from 4.66% to 5.48%) is small, which is not surprising as explicit randomization is infrequent (around one in twenty decisions). In relative terms, however, the increase represents a difference in the predicted direction of about 17.60%.

Since the lottery pairs included the 7 used in AO17, we can also compare how behaviour in those lotteries differed between taxed and non-taxed trials. We see a difference in explicit randomization also within these 7 lottery pairs, with more randomization for taxed trials (6.48%) compared to non-taxed trials (4.25%, WSR test, $N = 225$, $z = 3.876$, $p < 0.001$). This remains true for EASY pairs (taxed trials 8.66%, non-taxed trials 4.22%; WSR test, $N = 225$, $z = 2.795$, $p = 0.005$) and HARD pairs (taxed trials 6.22%, non-taxed trials 4.66%; WSR test, $N = 225$, $z = 2.399$, $p = 0.017$). There are no statistically significant differences for FOSD pairs (taxed trials 3.11%, non-taxed trials 2.67%; WSR test, $N = 225$, $z = 0.277$, $p = 0.782$).

The results of Experiment 2 hence (causally) confirm Experiment 1's result that choices to explicitly randomize are more frequent for choice pairs which are closer to indifference.

5 Experiment 3: Randomization Costs

An important argument for the existence of a deliberate preference for randomization is that decision makers sometimes explicitly choose a randomization device even when doing so is costly. In our Experiment 2, however, coin flip choices were infrequent in absolute terms even though they were costless. This suggests that explicit randomization might be even less frequent if it were costly, which could potentially call into question its practical economic relevance. We hence conducted a third, pre-registered experiment where we causally manipulated the cost of a randomization device (a coin flip). We aimed to gauge the importance of explicit randomization as a behavioral phenomenon and quantify its relation to randomization costs. In particular, the objective was to examine the absolute frequency of costly and costless coin flips.

In Experiment 3, participants repeatedly made choices over the same pairs of lotteries. Half of the time, there was no randomization option. For the remaining repetitions, participants could delegate their choice to a coin flip, for which different cost levels were implemented (including zero). Choices for the coin flip depending on the cost are our measure of explicit randomization. Additionally, as in Experiment 1, we included a number of other lottery pairs with the objective of estimating preferences, hence allowing to relate strength of preference to implicit and explicit randomization.

We pre-registered the natural hypothesis that explicit randomization should decrease with the cost of the randomization device. Additionally, the design allowed us to replicate our results on implicit randomization from Experiment 1 and correlate it with explicit randomization. For this purpose, we also pre-registered the following hypotheses. First, implicit randomization should depend on strength of preference. Second, implicit and explicit randomization should be correlated. Third, if explicit randomization is codetermined by strength of preference, it should depend on the variables reflecting the latter, and be independent of them otherwise.

5.1 Experimental Design and Procedures

We collected data from $N = 221$ participants from the participant pool of Prolific (37.10% females; median age 29). The experiment was preregistered at AsPredicted 192239 and obtained ethical approval from the Faculty Research Ethics Committee of the University of Lancaster, FASSLUMS-2024-4806-LeXeL-3. The experiment consisted of a within-subject design where every participant saw the same set of stimuli. To measure implicit and explicit randomization, participants were presented with 7 lottery pairs, each repeated 8 times in random order. These were the same lottery pairs used in AO17 which were also included in Experiments 1 and 2. Additionally, 20 different lottery pairs, each presented once, were used to estimate risk attitudes out of sample. Those were the same used for that purpose in Experiment 1. Overall, participants were presented with $7 \times 8 + 20 = 76$ lottery pairs.

For 4 of the 8 repetitions of lottery pairs targeting randomization, participants were just asked to make binary choices (as in Experiment 1). The proportion of stochastic choices for each pair out of these 4 repetitions is our measure of implicit randomization. In the remaining 4 repetitions, participants were presented with the additional option to delegate their choice to a coin flip. Choices for the coin flip are our measure of explicit randomization. Selecting the coin flip was costly and 4 different cost levels were implemented (GBP 0, 0.1, 0.3, and 0.5) for each of the lotteries, yielding the 4 repetitions.

At the end of the experiment, we asked participants a number of questions regarding randomization. People were asked whether they recognized that some pairs were repeated; if so, whether they explicitly made different choices among repetitions; and, if so, why they did this. These questions were aimed at giving us further anecdotal insights on the reasons behind inconsistent behavior.

There was no feedback during the course of the experiment. That is, participants did not receive any information regarding their earnings until the very end of the experiment. All decisions were made independently and at a participant's individual pace. To determine a participant's payoff, one lottery pair was randomly selected and paid. In addition, participants received a show-up fee of GBP 3.00. Payment procedures were explained within the instructions and carried out truthfully. The average total remuneration was GBP 11.70. The experiment lasted about 18 minutes including (online) instructions.

5.2 Results

First, we study implicit randomization. As in Experiment 1, for each choice pair and each participant, let λ_1 be the proportion of times that each participant chose the alternative with the highest expected value (or a predetermined one in case of ties, which happens for one of the pairs) among the 4 repetitions with binary choice only (i.e., without a coin flip available). Implicit randomization is then defined as IR$= |\lambda_1 - 0.5|$, which is 0 for maximal randomization and 0.5 for fully consistent choice.

We observe that the frequency of implicit randomization is comparable with our first experiment and AO17. Out of the 1547 choice pairs (221 participants times 7 pairs), 67.74% (1048) were deterministic and 32.26% reflected inconsistent behavior. In particular, implicit randomization happened frequently. Using the same estimation method reported in Experiment 1, on the same lottery pairs, assuming CARA utility functions, we obtain individual risk attitudes for the new participants (the results are qualitatively unchanged with other utility specifications). As shown in Table 3, implicit randomization is predicted by distance from indifference (EU-Dist; Model 1). As explained in the

Table 3: Experiment 3. Random effects panel regression on implicit randomization (λ_1).

$\lvert\lambda_1 - 0.5\rvert$	Model 1	Model 2
EU-Dist	0.117***	0.063***
	(0.011)	(0.015)
FOSD		0.028**
		(0.012)
Hard		-0.084***
		(0.012)
Female		0.027**
		(0.013)
Age		-0.000
		(0.001)
Constant	0.364***	0.414***
	(0.008)	(0.023)
N	1547	1547
χ^2	118.85	197.83
R^2_o	0.043	0.111

Notes: Robust standard errors in brackets, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. For a given pair of lotteries, EU-Diff denotes the expected utility distance between lottery A (higher EV) and B (lower EV).

discussion of Experiment 1, the classification in HARD and EASY when only these particular lotteries are used just partitions the lottery pairs according to strength of preference, and indeed Model 2 shows that this classification captures the same effect, with the coefficient for HARD (vs. EASY) being significantly negative. We hence replicate the results reported in Table 1 for Experiment 1 (Models 3–5). Model 2 also includes a dummy for lottery pairs where a lottery was stochastically dominated (FOSD), which was significantly positive (see Section 6).

In contrast, explicit randomization is infrequent. Out of the 6188 choices (221 participants times 7 lotteries repeated 4 times), we only observe 392 explicit randomizations, which is just 6.33% of the total. That is, on average, participants chose 1.77 coin flips out of the 28 possible ones. The median number of coin flips per individual is 0, and 125 individuals never chose even a single coin flip (and 165 never chose a single costly coin flip).

The individual average proportion of coin flips (6.33%) is comparable with the proportion in Experiment 2 (4.66% for non taxed and 5.48% for taxed trials). In agreement with our hypothesis, we observe more explicit randomization when it was free (11.89%) compared to when a cost was implemented (average 4.48%; 0.1 GBP, 5.88%; 0.3 GBP, 3.75%; 0.5 GBP, 3.81%). Moreover, we also observe more coin flips for HARD (7.04%) than for EASY lottery pairs (5.49%).

Table 4 reports random-effects panel probit regressions on the dummy that indicates whether in a given trial a participant chose a coin flip (explicit randomization). The results show that costs are a good predictor of the probability of choosing to flip a coin (Model 1), and this conclusion remains unchanged when controlling for strength of preference, implicit randomization, or demographic variables (Models 5–8). This is the main result from Experiment 3.

Compared to Experiment 1, the main focus of Experiment 3 was on costly, explicit randomization decisions. As a consequence, in order to make randomization a simple, explicit decision, we used binary

Table 4: Experiment 3. Random effects panel probit regressions on whether a participant chose to flip a coin.

Con Flip	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Cost	-0.180***				-0.180***	-0.180***	-0.181***	-0.182***
	(0.026)				(0.026)	(0.026)	(0.026)	(0.026)
IR		-0.373*			-0.396		-0.335	-0.108
		(0.226)			(0.244)		(0.258)	(0.262)
EU-Dist			-0.176	-0.086		-0.188	-0.142	-0.069
			(0.125)	(0.175)		(0.132)	(0.140)	(0.200)
FOSD				-0.105				-0.088
				(0.181)				(0.200)
Hard				0.161				0.151
				(0.105)				(0.111)
Female								0.317*
								(0.170)
Age								-0.008
								(0.009)
Constant	-1.913***	-2.054***	-2.152***	-2.263***	-1.766***	-1.870***	-1.756***	-1.845***
	(0.110)	(0.130)	(0.105)	(0.159)	(0.141)	(0.114)	(0.141)	(0.357)
N	6188	6188	6188	6188	6188	6188	6188	6188
χ^2	49.063	2.719	1.985	10.548	50.356	51.951	52.993	56.715
LL	-1104.725	-1156.683	-1157.161	-1152.293	-1103.181	-1103.660	-1102.617	-1076.254

Notes: Robust standard errors in brackets, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. For a given pair of lotteries, EU-Dist denotes the expected utility distance (absolute value of the difference) between the two lotteries in the pair.

yes/no choices (coin flips) and a limited set of lotteries, while Experiment 1 allowed for (costless) urn choices which could fully reproduce the exact proportions of consistent choices under implicit randomization. That is, in Experiment 3, participants could only choose a coin flip (a 50-50 explicit randomization), which does not match the possible values generated by implicit randomization (0.25, 0.5, or 0.75). This makes the study of the effect of costs on explicit randomization more transparent, but it limits the design's ability to investigate the relation between implicit and explicit randomization and the dependence of randomization on strength of preference. Also, recall that, in Experiment 3, and unlike in Experiment 1, there are no intermediate pairs in the design in addition to the Hard and Easy ones. Indeed, in Table 4, expected utility distance (Model 3) does not reveal an effect of strength of preference on explicit randomization ($p = 0.138$). Adding a dummy for HARD (vs. EASY) pairs also shows no significant effect of strength of preference for explicit randomization (Models 4 and 8). A dummy for pairs where one lottery was stochastically dominated (FOSD) was also non-significant (see Section 6). The correlation between explicit and implicit randomization (Model 2) is also only marginally significant ($p = 0.099$) and becomes non-significant when controlling for cost (Model 5, $p = 0.159$). A Spearman's test also shows limited correlation between the explicit and implicit randomization at the individual level (correlation of individual-level average IR and individual proportion of coin flips, $N = 221$, $\rho = 0.074$, $p = 0.276$).

Finally, at the end of the experiment, we asked participants a series of (exploratory) questions on their behavior. The first question was whether they were aware that some of the lottery pairs were repeated. 89.59% of participants answered yes to this question. These participants were then asked whether they did try to be consistent between the repetitions of the questions. 84.34% of them answered yes to this question. The remaining 31 participants (15.66%) answering "no" were further asked why they gave different answers to the same questions (open-ended). While the specific wording varied, anecdotally, 12 of the 31 participants answered in line with a change of heart, e.g. "I changed my mind on which I deemed suitable." 10 participants explicitly mentioned hedging or equivalent strategies, e.g. "I was unsure of which answer to choose on some of the repeated questions and figured if I chose a different answer for some the chance might balance out," or "Hedging bets - if I know one of the questions is randomly selected at the end, I don't want a disproportionate number of 'risky' bets, but some are okay." Others reported trying to follow specific strategies (which should have made them consistent: "I wanted big money," " I tried to do quick math," "I tried to find the option with the most points") and only one specifically reported flipping a coin in his/her head when the differences between the options were small. Taken together, these answers do suggest that, for some individuals, inconsistency in repeated choices has some deliberate components. However, as far as we can tell from our exploratory questions, this appears to affect only a small part of the decisions in our experiments.

Overall, Experiment 3 confirms that some decision makers choose explicit randomization options when those are made available and salient, but this is a relatively infrequent phenomenon compared to implicit randomization (i.e. inconsistency in the face of repeated choices) and it is further reduced when the randomization option is costly.

6 FOSD

Our experiments included one lottery pair, taken from the 7 used in AO17, where one of the lotteries was stochastically dominated by the other one in the pair. It is well-known that decisions involving dominance typically show less inconsistency than other decisions with comparable utility differences (e.g., Harless and Camerer, 1994; Loomes et al., 2002). In particular, this poses a problem for a strength of preference argument as embodied in additive RUMs, which imply that inconsistency should be larger for FOSD pairs with small utility differences than for other pairs with large utility differences. This is not the case in the data we report. On the contrary, choices involving FOSD correspond to very small expected utility differences, but repeated choices are quite consistent (see Figure 1 for the reanalysis of AO17's data, and Figure 3 for our Experiment 1). For example, in Experiment 1, only 4 out of 103 participants (3.88%) made inconsistent choices in the repetitions of the FOSD pair.

We also find that explicit randomization is rare when one lottery is dominated. In Experiment 1, only one participant (0.97%) chose to explicitly randomize in an FOSD pair. In Experiment 3, the average proportion of coin flips for FOSD pairs was 5.20% (median 0.00%), compared to 7.04% for Hard pairs and 5.49% for Easy ones. That is, choices for FOSD pairs resemble those for Easy pairs even though they are more comparable to Hard ones in terms of expected utility differences (see Figure 5). For Experiment 2, the proportion of coin flips was also quite small for FOSD pairs (taxed trials 3.11%, non-taxed trials 2.67%).

In our regressions analyses, whenever we look at the contrast between Hard and Easy lottery pairs (through a dummy for Hard pairs), we also include a dummy for FOSD lottery pairs. This dummy is not significant for either implicit or explicit randomization in Experiment 1 (Table 1, Models 4–5 and 9–10). In Experiment 3, this dummy is also non-significant for the probability of a coin flip (Table 4, Models 4 and 8). In all these regressions, the reference category are the Easy pairs. Thus, consistency (or lack of explicit randomization) is comparable to that of Easy pairs, even though the expected utility differences in FOSD pairs are comparable to those of Hard pairs. In Experiment 3, the FOSD dummy is significantly positive for implicit randomization (Table 3, Model 2). Since the dependent variable is IR, which is minimal for complete inconsistency and maximal for full consistency, this means that, in this experiment, participants were even more consistent for FOSD pairs than for Easy pairs.

Our data thus reproduces the observation that when choices involve dominance, decision makers tend to be very consistent (and overwhelmingly choose the dominating alternative) even if the expected utility differences are small. We extend this observation to explicit randomization, be it in the form of coin flips (Experiment 3) or explicitly selected frequencies (Experiment 1).

From the theoretical point of view, the consistency of decisions (lack of implicit randomization) for FOSD pairs is problematic for RUMs where the noise is formulated as a perturbation of the utilities themselves (additive RUMs). Other models capturing strength of preference arguments do not have this problem. In particular, random parameter models (Loomes and Sugden, 1998; Apesteguía and Ballester, 2018) postulate a pair-specific (as opposed to option-specific) random perturbation of a parameter in a functional utility form (e.g., the risk aversion parameter in a CARA or CRRA utility function). While random parameter models still capture strength of preference arguments in general (for a discussion, see Alós-Ferrer and Garagnani, 2024), they predict no inconsistency in the presence of an FOSD relation, as the expected utility of a first-order stochastically dominated lottery is always smaller independently of the realized parameter of the Bernoulli utility function. This adds to other theoretical arguments in favor of random parameter models (Apesteguía and Ballester, 2018).

The evidence for the lack of explicit randomization in FOSD pairs also agrees with the predictions of the Cautious Expected Utility model of Cerreia-Vioglio et al. (2015). This model postulates that decision makers evaluate opportunities for explicit randomization by minimizing certainty equivalents across a given set of utility functions. Consequently, it predicts that first-order stochastically dominated options should be avoided (Cerreia-Vioglio et al., 2015, Proposition 6), which agrees with our data.

7 Further Related Literature

Dwenger et al. (2018) study applications to medical schools in Germany, where applicants must submit multiple rank-order lists of universities. They document that 14% of applicants exhibit preference inconsistencies in their applications, which introduces randomness in the school allocation. They then carry out several classroom experiments where participants choose among lotteries, but their analysis does not compare inconsistencies and explicit decisions to randomize, as this was not the objective of the contribution. In one experiment, 28% of participants were inconsistent when binary choices were repeated. In another experiment, an explicit option to randomize was made available and chosen by 53% of participants. The latter shows that an option to randomize is often chosen when made available and salient, but the difference in the results of the experiments does not speak in favor of a (deliberate) link between preference for randomization and inconsistencies in repeated binary choices.

Other recent contributions have argued that people sometimes attempt to deliberately implement a randomization between alternatives even when this is not an explicit option. Agranov et al. (2022) documents preferences for randomization in games and individual choices and argue that this seems to be a fundamental trait of individual preferences and not just a pattern of errors. However, the evidence reported in Agranov et al. (2022) can be accounted for by standard strength-of-preference arguments. In particular, that work defines a participant as a “mixer” if she or he chose the modal option less than 90% in a decision problem, and show that levels of mixing in any given domain “are obviously responsive to this cutoff.” However, this is confounded with a strength-of-preference argument if preference errors have a neural (hence individual) origin, i.e. some decision makers are noisier than others. It can also be explained if preferences across domains are correlated. That is, some participants might be closer to indifference than others, hence be more inconsistent and be classified as mixers.

Interestingly, several findings in Agranov et al. (2022) support strength-of-preference effects instead of preferences for deliberate randomization. The authors report that (i) “[a]s the risky bet becomes less risky, participants become more likely to mix by adding in the risky bet”; (ii) “[r]andomization [...] responds monotonically to parameter changes in the environment,” (iii) “we find that our mixing participants put more weight on a lottery or strategy when its expected value increases, consistent with the common assumption that better responses are played more frequently”; and (iv) “we see that mixing responds sensibly to changes in payoff parameters, with participants mixing less in each domain as the difference in expected values is increased.” These results are well-established in the literature, and are also direct implications of standard strength-of-preference effects.

Agranov et al. (2022) also report that mixing is related to individual risk preferences, a result which again is easily explained by strength-of-preference effects. However, the analysis relating mixing behavior to individual risk preferences is conducted by reducing the latter to a binary variable, which prevents a more detailed investigation of strength-of-preference effects. We also remark that the results regarding mixing behavior in games are predicted and well-known in the literature on Quantal Response Equilibria (McKelvey and Palfrey, 1995, 1998), which is the game-theoretical counterpart of strength-of-preference effects and prescribes that larger payoff differences monotonically translate into more consistent choices. This is in agreement with the results of Agranov et al. (2022) and does not require a preference for explicit randomization.

Relatedly, Agranov and Ortoleva (2023) aimed to quantify the prevalence of preferences for randomization. They allowed participants to randomize between options in a series of questions in which one of the alternatives was fixed and the other varied, capturing the range of values for which participants preferred to randomize. However, as Agranov and Ortoleva (2023) acknowledged, instead of attesting for the existence of a preference for randomization, these ranges can be explained by preference imprecision, which can be related to strength of preference phenomena (see, e.g. Butler and Loomes, 2007, 2011). This parallels the experiment on “cognitive uncertainty” by Enke and Graeber (2023), where participants were allowed to indicate the bounds for which they were “certain” about their choices. Hence, the presence of ranges where participants decide to randomize is not necessarily evidence for the existence of a preference for randomization.

Feldman and Rehbeck (2022) also explore the link between risk preferences and a preference for randomization. They compare behavior in budget choice tasks with consecutive binary choices over lotteries, where participants are explicitly warned that they will face the exact same choice several times in a row. Many subjects exactly match the choice from the convex choice task with the repeated discrete choice task, creating a positive correlation among them (between 0.3 and 0.4).[9] Crucially, Feldman and Rehbeck (2022) report that in their data, this is mostly because “many individuals who chose either the numeraire or extreme lottery in the convex choice task also repeatedly choose the same lottery in the repeated choice task.” That is, the correlation between the two measures is

[9]We remark that Garagnani (2023) showed that most people tend to consistently choose dominated allocations in budget choice tasks with sliders, casting doubts on whether all participants understand how to use these tasks.

mostly due to deterministic behavior of participants choosing the same option consistently in both tasks (see Feldman and Rehbeck, 2022, Figure 8). In this sense, the results of Feldman and Rehbeck (2022) are analogous to our Table 2 and do not provide strong evidence for a literal preference for deliberate randomization. Only correlation of non-deterministic behavior between the tasks would provide sufficient evidence for the existence of a deliberate preference for randomization, but there are very few data points revealing non-deterministic and equivalent behavior across tasks in either our Table 2 or Figure 8 of Feldman and Rehbeck (2022). Thus, the correlation between tasks might reflect deterministic choices and the effect of aggregating choices among participants, and not an explicit preference for randomization.

Some additional, recent studies have investigated the effect of adding explicit, salient options for randomizations in specific paradigms. For example, Ong and Qiu (2023) investigate the presence of a preference for randomization in the ultimatum game. Using the strategy method, responders were given the explicit option to actively randomize between accepting and rejecting a given offer (by assigning a specific acceptance probability). Most receivers (86.7%) chose to randomize for at least some decision. An additional stage showed that many participants were willing to pay for randomization. The stated acceptance probability increased with the responder's share, implying less randomization for larger amounts of money, as predicted by strength-of-preference effects. Analogously, Shi et al. (2024) recently studied preference reversals (e.g., Lichtenstein and Slovic, 1971; Grether and Plott, 1979; Tversky and Thaler, 1990; Butler and Loomes, 2007) with a lottery-choice design where participants could either pay a small cost to select a specific option or opt for a free randomization choice where a computer randomly selects an option (a coin flip). In a sense, this design is the opposite of our third experiment. They show that the majority of subjects (77.7%) chose the randomization option compared to costly selecting their preferred alternative at least once during the experiment. They interpret this as subjects lacking a clear preference. In this study, lottery pairs were constructed using a bisection process through a series of binary decisions designed to elicit a certainty equivalent (Shi et al., 2024, Fig. 2). Specifically, one of the options was always the mid-point between 0 and the winning amount of money of the previous chosen option (with additional filler trials). This procedure mechanically constructs new pairs that are closer to indifference for each participant. Hence, strength-of-preference effects predict the high degree of randomization observed in the data. Note that Alós-Ferrer et al. (2016) already studied the relation between strength of preference and the preference reversal phenomenon.

The contributions discussed above are compatible with our results and further suggest that explicit, salient options to randomize are taken at least part of the time by many decision makers, but that the overall patterns can be explained through strength-of-preference effects.

8 Discussion

Our results suggest that both choice inconsistencies and deliberate decisions to randomize can be well-explained by strength-of-preference effects. As frequently shown in the literature, people make inconsistent choices when confronted with the same decision multiple times (e.g., Mosteller and Nogee, 1951; Davidson and Marschak, 1959; Hey and Orme, 1994; Ballinger and Wilcox, 1997). As shown here and elsewhere (Alós-Ferrer and Garagnani, 2021, 2022a,b; Duffy and Smith, 2025), the extent to which people behave inconsistently is strongly correlated with strength of preference, i.e. the difference in underlying utilities between the options. This relation is implied by random utility models (McFadden, 1974, 2001), which are the basis for standard microeconometric estimation methods, and is also an integral prediction of sequential sampling models (Ratcliff, 1978; Fudenberg et al., 2018; Webb, 2019).

Perhaps more surprisingly, our data shows that a similar relation exists between strength of preference and the choice to delegate the decision to a randomization device if available. That is, there is also a relation between strength of preference and deliberate choices to randomize. However, in our first experiment, we do reject the hypothesis that choice inconsistency in the face of (nonconsecutive) repetitions of a given choice pair reflects a deliberate attempt to exactly implement a preference for explicit randomization, as revealed in the corresponding task.

Our data does confirm that, if given the explicit and salient option to delegate a choice to a randomizing device, people sometimes do so (e.g., Cohen et al., 1987; Cubitt et al., 2015; Agranov and Ortoleva, 2017; Cettolin and Riedl, 2019), although this happens infrequently (that is, the average percentage of explicit delegations to a randomization device is low). However, our results suggest that choice inconsistency in repeated decisions and deliberate choices to randomize when the option is salient might both have a common cause: strength-of-preference effects. That is, in our data, as well as in previous evidence, people are indeed inconsistent in their choices, or actively choose to mix between the available options when randomization is explicitly added as a possibility, but both kinds of behavior are directly linked to the strength of their preference rather than to an intrinsic preference for randomization. This might create (spurious) correlations between both phenomena, but it does not imply that choice inconsistency follows because people deliberately attempt to implement a preference for randomization.

In our first experiment, we collect both choice inconsistencies and explicit randomization decisions making sure that both can generate the exact same range of probabilities. However, the data rejects the hypothesis that the choice proportions derived from inconsistencies coincide with the preferences for randomization as revealed by the explicit randomization decisions. Our results also show that both implicit and explicit randomization can be explained by strength of preference. Our second experiment uses a causal manipulation showing that the explicit decision to randomize is more frequent when the utility differences among the alternatives are reduced. Our third experiment adds monetary costs to explicit randomization options and showed that the frequency of explicit randomization decreased when costs increased, with a sharp decrease compared to zero costs.

Our data is compatible with the view that the monotonic relation between the strength of the subjective preferences and the proportion of consistent choices is mainly due to strength-of-preference effects, for which there is widespread evidence in psychology, neuroscience, and (more recently) also economics. This is, for example, an explicit prediction of standard sequential-sampling models (e.g., Ratcliff, 1978; Ratcliff and Rouder, 1998; Shadlen and Kiani, 2013; Shadlen and Shohamy, 2016), which can be interpreted as modelling the process by which the human brain evaluates options in a noisy way (Fudenberg et al., 2018; Baldassi et al., 2020). In such processes, noise naturally results in more errors when underlying values are closer than when they are further apart. In other words, in those models choice inconsistency appears because the choice process is noisy, but noise results in more errors if utility differences are small. This creates psychometric effects (strength of preference).

Why, then, is there also a strength-of-preference effect for the explicit decision to deliberately randomize? On the one hand, psychometric effects need not be in any way conscious. It is well-known in psychology that people very often fail to have conscious access to the actual reasons for their actions (or their mistakes), especially when those are spontaneous (e.g., Nisbett and Wilson, 1977; Greenwald and Banaji, 1995), and that, if asked, they often provide *ex post* rationalizations instead. Thus, the decision to randomize when this is an explicit option might arise from strength of preference in an unconscious way, hence creating the same kind of strength-of-preference effect as in the case of choice inconsistency. That the decision is deliberate does not mean that the underlying reasons are conscious. On the other hand, the relation between strength of preference and deliberate randomization might arise through a different but closely-related channel. Sequential sampling models also imply that harder decisions take longer, reflecting the difficulty of sorting apart closer decision values. Hence, decision makers take longer for harder decisions, which might make an explicit option to delegate the decision to a randomizing device attractive (e.g., Krajbich et al., 2014), and in particular more attractive the longer the decision maker struggles, which in turn correlates with strength of preference. In both cases, the decision to take an explicit randomization option will correlate with strength of preference, exactly as choice inconsistency.

References

Agranov, M., P. J. Healy, and K. Nielsen (2022): "Stable Randomization," *Economic Journal*, 133, 2553–2579.

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," in *AEA Papers and Proceedings*, vol. 112, 426–30.

——— (2023): "Ranges of Randomization," *Review of Economics and Statistics*, 1–44.

Alós-Ferrer, C., E. Fehr, and N. Netzer (2021): "Time Will Tell: Recovering Preferences when Choices are Noisy," *Journal of Political Economy*, 129, 1828–1877.

Alós-Ferrer, C. and M. Garagnani (2021): "Choice Consistency and Strength of Preference," *Economics Letters*, 198, 109672.

——— (2022a): "Strength of Preference and Decisions Under Risk," *Journal of Risk and Uncertainty*, 64, 309–329.

——— (2022b): "The Gradual Nature of Economic Errors," *Journal of Economic Behavior and Organization*, 200, 55–66.

——— (2024): "Improving Risky-Choice Predictions Using Response Times," *Journal of Political Economy: Microeconomics*, 2(2), 335–354.

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

Anderson, S. P., J.-F. Thisse, and A. De Palma (1992): *Discrete Choice Theory of Product Differentiation*, Cambridge, MA: MIT Press.

Apesteguía, J. and M. A. Ballester (2018): "Monotone Stochastic Choice Models: The Case of Risk and Time Preferences," *Journal of Political Economy*, 126, 74–106.

Baldassi, C., S. Cerreia-Vioglio, F. Maccheroni, and M. Marinacci (2020): "A Behavioral Characterization of the Drift Diffusion Model and its Multi-Alternative Extension to Choice under Time Pressure," *Management Science*, 66, 5075–5093.

Ballinger, T. P. and N. T. Wilcox (1997): "Decisions, Error and Heterogeneity," *Economic Journal*, 107, 1090–1105.

Baltas, G. and P. Doyle (2001): "Random Utility Models in Marketing Research: A Survey," *Journal of Business Research*, 51, 115–125.

Butler, D. J. and G. Loomes (2007): "Imprecision as an Account of the Preference Reversal Phenomenon," *American Economic Review*, 97, 277–297.

——— (2011): "Imprecision as an Account of Violations of Independence and Betweenness," *Journal of Economic Behavior and Organization*, 80, 511–522.

Camerer, C. F. (1989): "Does the Basketball Market Believe in the 'Hot Hand'?" *American Economic Review*, 79, 1257–1261.

Cappellari, L. and S. P. Jenkins (2003): "Multivariate Probit Regression Using Simulated Maximum Likelihood," *The Stata Journal*, 3, 278–294.

Cattell, J. M. (1902): "The Time of Perception as a Measure of Differences in Intensity," *Philosophische Studien*, 19, 63–68.

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

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.

Chabris, C. F., C. L. Morris, D. Taubinsky, D. Laibson, and J. P. Schuldt (2009): "The Allocation of Time in Decision-Making," *Journal of the European Economic Association*, 7, 628–637.

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.

Chiong, K. X., M. Shum, R. Webb, and R. Chen (2024): "Combining Choice and Response Time Data: A Drift-Diffusion Model of Mobile Advertisements," *Management Science*, 70, 1238–1257.

Clithero, J. A. (2018): "Response Times in Economics: Looking Through the Lens of Sequential Sampling Models," *Journal of Economic Psychology*, 69, 61–86.

Cohen, M., J.-Y. Jaffray, and T. Said (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.

Cubitt, R. P., D. Navarro-Martinez, and C. Starmer (2015): “On Preference Imprecision,” *Journal of Risk and Uncertainty*, 50, 1–34.

Dashiell, J. F. (1937): “Affective Value-Distances as a Determinant of Aesthetic Judgment-Times,” *American Journal of Psychology*, 50, 57–67.

Davidson, D. and J. Marschak (1959): “Experimental Tests of a Stochastic Decision Theory,” in *Measurement: Definitions and Theories*, ed. by W. Churchman and P. Ratoosh, New York: Wiley, vol. I, Part I, 233–269.

Debreu, G. (1958): “Stochastic Choice and Cardinal Utility,” *Econometrica*, 26, 440–444.

Dehaene, S. (1992): “Varieties of Numerical Abilities,” *Cognition*, 44, 1–42.

Dehaene, S., E. Dupoux, and J. Mehler (1990): “Is Numerical Comparison Digital? Analogical and Symbolic Effects in Two-Digit Number Comparison,” *Journal of Experimental Psychology: Human Perception and Performance*, 16, 626–641.

Duffy, S. and J. Smith (2025): “Stochastic Choice and Imperfect Judgments of Line Lengths: What is Hiding in the Noise?” *Journal of Economic Psychology*, 106, 102787.

Dwenger, N., D. Kübler, and G. Weizsäcker (2018): “Flipping a Coin: Evidence from University Applications,” *Journal of Public Economics*, 167, 240–250.

Enke, B. and T. Graeber (2023): “Cognitive Uncertainty,” *Quarterly Journal of Economics*, 138, 2021–2067.

Feldman, P. and J. Rehbeck (2022): “Revealing a Preference for Mixtures: An Experimental Study of Risk,” *Quantitative Economics*, 13, 761–786.

Feng, Y., R. Caldentey, and C. T. Ryan (2022): “Robust Learning of Consumer Preferences,” *Operations Research*, 70, 918–962.

Fishburn, P. C. (1988): *Nonlinear Preference and Utility Theory*, Baltimore, Maryland: Johns Hopkins University Press.

Frydman, C. and L. J. Jin (2022): “Efficient Coding and Risky Choice,” *Quarterly Journal of Economics*, 137, 161–213.

Fudenberg, D., P. Strack, and T. Strzalecki (2018): “Speed, Accuracy, and the Optimal Timing of Choices,” *American Economic Review*, 108, 3651–3684.

Garagnani, M. (2023): “The Predictive Power of Risk Elicitation Tasks,” *Journal of Risk and Uncertainty*, 67, 165–192.

Garagnani, M. and F. M. Vieider (2025): “Economic Consequences of Numerical Adaptation,” *Psychological Science*, 36, 407–420.

Glimcher, P. W., M. C. Dorris, and H. M. Bayer (2005): “Physiological Utility Theory and the Neuroeconomics of Choice,” *Games and Economic Behavior*, 52, 213–256.

Gneezy, U. and J. Potters (1997): "An Experiment on Risk Taking and Evaluation Periods," *Quarterly Journal of Economics*, 112, 631–645.

Greenwald, A. G. and M. R. Banaji (1995): "Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes," *Psychological Review*, 102, 4–27.

Grether, D. M. and C. R. Plott (1979): "Economic Theory of Choice and the Preference Reversal Phenomenon," *American Economic Review*, 69, 623–638.

Halevy, Y., D. Walker-Jones, and L. Zrill (2023): "Difficult Decisions," Working Paper, University of Toronto.

Halton, J. H. (1960): "On the Efficiency of Certain Quasi-Random Sequences of Points in Evaluating Multi-Dimensional Integrals," *Numerische Mathematik*, 2, 84–90.

Harless, D. W. and C. F. Camerer (1994): "The Predictive Utility of Generalized Expected Utility Theories," *Econometrica*, 62, 1251–1289.

Hey, J. D. (2001): "Does Repetition Improve Consistency?" *Experimental Economics*, 4, 5–54.

Hey, J. D. and C. Orme (1994): "Investigating Generalizations of Expected Utility Theory Using Experimental Data," *Econometrica*, 62, 1291–1326.

Hicks, J. R. and R. G. Allen (1934): "A Reconsideration of the Theory of Value. Part I," *Economica*, 1, 52–76.

Karp, J. A. (2009): "Candidate Effects and Spill-over in Mixed Systems: Evidence from New Zealand," *Electoral Studies*, 28, 41–50.

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

Krajbich, I., B. Oud, and E. Fehr (2014): "Benefits of Neuroeconomic Modeling: New Policy Interventions and Predictors of Preference," *American Economic Review (Papers and Proceedings)*, 104, 501–506.

Laming, D. (1985): "Some Principles of Sensory Analysis," *Psychological Review*, 92, 462–485.

Lichtenstein, S. and P. Slovic (1971): "Reversals of Preference Between Bids and Choices in Gambling Decisions," *Journal of Experimental Psychology*, 89, 46–55.

Loomes, G., P. G. Moffatt, and R. Sugden (2002): "A Microeconometric Test of Alternative Stochastic Theories of Risky Choice," *Journal of Risk and Uncertainty*, 24, 103–130.

Loomes, G. and R. Sugden (1998): "Testing Different Stochastic Specifications of Risky Choice," *Economica*, 65, 581–598.

Luce, R. D. (1959): *Individual Choice Behavior: A Theoretical Analysis*, New York: Wiley.

Machina, M. J. (1985): "Stochastic Choice Functions Generated from Deterministic Preferences over Lotteries," *Economic Journal*, 95, 575–594.

McFadden, D. L. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, ed. by P. Zarembka, New York: Academic Press, 105–142.

——— (2001): "Economic Choices," *American Economic Review*, 91, 351–378.

McKelvey, R. D. and T. R. Palfrey (1995): "Quantal Response Equilibria for Normal Form Games," *Games and Economic Behavior*, 10, 6–38.

——— (1998): "Quantal Response Equilibria for Extensive Form Games," *Experimental Economics*, 1, 9–41.

Miao, B. and S. Zhong (2018): "Probabilistic Social Preferences: How Machina's Mom Randomizes Her Choice," *Economic Theory*, 65, 1–24.

Moffatt, P. G. (2005): "Stochastic Choice and the Allocation of Cognitive Effort," *Experimental Economics*, 8, 369–388.

——— (2015): *Experimetrics: Econometrics for Experimental Economics*, London: Palgrave Macmillan.

Mosteller, F. and P. Nogee (1951): "An Experimental Measurement of Utility," *Journal of Political Economy*, 59, 371–404.

Moyer, R. S. and T. K. Landauer (1967): "Time Required for Judgements of Numerical Inequality," *Nature*, 215, 1519–1520.

Nielsen, K. and L. Rigotti (2022): "Revealed Incomplete Preferences," ArXiv preprint, arXiv:2205.08584.

Nisbett, R. E. and T. D. Wilson (1977): "Telling More Than We Can Know: Verbal Reports on Mental Processes," *Psychological Review*, 84, 231–259.

Nownes, A. J. (1992): "Primaries, General Elections, and Voter Turnout: A Multinomial Logit Model of the Decision to Vote," *American Politics Quarterly*, 20, 205–226.

Ong, Q. and J. Qiu (2023): "Paying for Randomization and Indecisiveness," *Journal of Risk and Uncertainty*, 67, 45–72.

Padoa-Schioppa, C. and J. A. Assad (2006): "Neurons in the Orbitofrontal Cortex Encode Economic Value," *Nature*, 441, 223–226.

Rabin, M. (2000): "Risk Aversion and Expected-Utility Theory: A Calibration Theorem," *Econometrica*, 68, 1281–1292.

Ratcliff, R. (1978): "A Theory of Memory Retrieval," *Psychological Review*, 85, 59–108.

Ratcliff, R. and J. N. Rouder (1998): "Modeling Response Times for Two-Choice Decisions," *Psychological Science*, 9, 347–356.

Ratcliff, R. and F. Tuerlinckx (2002): "Estimating Parameters of the Diffusion Model: Approaches to Dealing with Contaminant Reaction Times and Parameter Variability," *Psychonomic Bulletin & Review*, 9, 438–481.

Roitman, J. D. and M. N. Shadlen (2002): "Response of Neurons in the Lateral Intraparietal Area During a Combined Visual Discrimination Reaction Time Task," *Journal of Neuroscience*, 22, 9475–9489.

Shadlen, M. N. and R. Kiani (2013): "Decision Making as a Window on Cognition," *Neuron*, 80, 791–806.

Shadlen, M. N. and D. Shohamy (2016): "Decision Making and Sequential Sampling from Memory," *Neuron*, 90, 927–939.

Shi, L., J. Qiu, J. Li, and F. Bohn (2024): "Consciously Stochastic in Preference Reversals," *Journal of Risk and Uncertainty*, 68, 255–297.

Spiliopoulos, L. and A. Ortmann (2018): "The BCD of Response Time Analysis in Experimental Economics," *Experimental Economics*, 21, 383–433.

Thurstone, L. L. (1927): "A Law of Comparative Judgement," *Psychological Review*, 34, 273–286.

Train, K. E. (2009): *Discrete Choice Methods with Simulation*, New York: Cambridge University Press, 2nd ed.

Tversky, A. (1969): "Intransitivity of Preferences," *Psychological Review*, 76, 31–48.

Tversky, A. and R. H. Thaler (1990): "Anomalies: Preference Reversals," *Journal of Economic Perspectives*, 4, 201–211.

Vieider, F. M. (2024): "Decisions Under Uncertainty as Bayesian Inference on Choice Options," *Management Science*, 70, 9014–9030.

Wald, A. (1945): "Sequential Tests of Statistical Hypotheses," *Annals of Mathematical Statistics*, 16, 117–186.

Wald, A. and J. Wolfowitz (1945): "Optimum Character of the Sequential Probability Ratio Test," *Annals of Mathematical Statistics*, 19, 326–339.

Webb, R. (2019): "The (Neural) Dynamics of Stochastic Choice," *Management Science*, 64, 230–255.

Wichmann, A. F. and N. J. Hill (2001): "The Psychometric Function: I. Fitting, Sampling, and Goodness of Fit," *Attention, Perception, & Psychophysics*, 63, 1293–1313.

Zizzo, D. J. (2010): "Experimenter Demand Effects in Economic Experiments," *Experimental Economics*, 13, 75–98.

A List of Lotteries

Table 5: List of Lotteries.

ID Pair	A.O. Lotteries	Lottery A				Lottery B				EV LA	EV LB
1	FOSD	103	103	103	103	98	98	98	98	103.00	98.00
2	Hard	13	30	51	81	19	32	38	86	43.75	43.75
3		2	32	38	92	2	28	60	70	41.00	40.00
4		5	5	70	98	5	5	75	77	44.50	40.50
5		27	27	45	86	22	34	34	83	46.25	43.25
6		10	10	75	75	27	27	27	69	42.50	37.50
7		28	35	63	76	26	33	53	85	50.50	49.25
8	Hard	10	10	90	90	32	45	45	56	50.00	44.50
9	Hard	16	16	94	94	38	38	38	77	55.00	47.75
10	Hard	6	84	105	200	54	60	117	135	98.75	91.50
11		5	25	55	96	5	30	46	75	45.25	39.00
12		3	3	88	90	3	3	63	95	46.00	41.00
13		24	24	55	96	26	26	38	75	49.75	41.25
14		28	56	56	68	26	26	60	60	52.00	43.00
15		36	44	60	73	12	38	55	82	53.25	46.75
16		25	25	76	76	27	36	36	46	50.50	36.25
17		12	12	91	91	23	23	23	85	51.50	38.5
18		23	52	55	95	25	35	57	58	56.25	43.75
19		6	33	50	98	6	34	52	53	46.75	36.25
20		27	30	30	97	20	34	34	56	46.00	36.00
21	Easy	23	23	30	30	5	5	5	31	26.50	11.50
22		23	23	86	86	25	35	35	55	54.50	37.50
23		7	40	48	98	9	13	50	56	48.25	32.00
24		4	49	52	97	4	18	53	55	50.50	32.50
25		36	36	67	99	38	38	43	43	59.50	40.50
26		30	35	35	99	17	37	37	38	49.75	32.25
27		17	17	86	86	21	24	29	40	51.50	28.50
28		14	14	89	89	16	16	16	75	51.50	30.75
29		20	25	63	98	26	30	31	33	51.50	30.00
30		7	23	85	99	7	35	36	47	53.50	31.25
31		10	10	90	99	19	19	37	37	52.25	28.00
32		22	27	70	95	24	28	30	32	53.50	28.50
33		27	27	89	90	28	28	28	30	58.25	28.50
34		18	68	85	85	24	28	32	32	64.00	29.00
35		60	60	60	65	6	6	6	68	61.25	21.50
36		65	69	69	77	5	5	5	85	70.00	25.00
37	Easy	85	85	85	85	12	14	16	96	85.00	34.50

Table 6: List of lotteries for the utility estimation.

ID Pair	Lottery A				Lottery B			
1	42	54	54	55	44	44	44	45
2	5	53	75	95	5	55	55	55
3	32	44	44	47	34	34	36	49
4	10	100	100	100	13	13	13	20
5	21	60	60	67	25	25	25	70
6	30	60	65	65	35	35	35	70
7	5	55	65	85	5	60	60	60
8	30	30	60	95	30	35	35	40
9	20	50	55	95	25	25	55	55
10	20	25	65	100	25	30	30	35
11	25	40	50	80	35	35	35	85
12	20	35	75	75	25	40	40	80
13	5	5	90	90	5	50	50	50
14	25	25	45	85	20	35	35	80
15	50	50	50	60	20	20	20	95
16	45	45	50	50	20	20	20	75
17	55	55	60	60	20	20	20	75
18	30	30	70	70	30	35	35	75
19	25	60	70	70	30	35	35	75
20	25	65	65	65	30	30	30	70

B Description of the Risk Attitude Estimation

To estimate individual utility functions from the binary lottery choices in part I of the experiment we follow the approach described in Moffatt (2015, Chapter 13) and in Alós-Ferrer and Garagnani (2021, 2022a). All $T = 20$ trials used for the utility estimation involved binary choices between lotteries of the form $A = (p, x)$ and $B = (q, y)$, where A pays x with probability p and B pays y with probability q. We index the trials in the experiment by $t = 1, ..., 20$, that is, in trial t participants face the choice between $A_t = (p_t, x_t)$ and $B_t = (q_t, y_t)$. Further, we index the N participants by $i = 1, ..., N$.

The results do not qualitatively change depending on the utility function which is assumed (CARA or CRRA) or on the shape of the noise, RUM or RPM. In the main analysis we assume a constant absolute risk aversion (CARA) utility function, which is given by

$$u(x \mid r) = \begin{cases} \frac{1-e^{-rx}}{1-e^{-rx_{\max}}}, & \text{if } r \neq 0 \\ \frac{x}{x_{\max}}, & \text{if } r = 0, \end{cases} \tag{1}$$

where $x_{\max} = \max\{x_1, \ldots, x_T, y_1, \ldots, y_T\}$ is the maximum outcome across all T lottery pairs (trials). The normalization ensures that $u(x \mid r)$ is increasing also for negative values of r (indicating risk-seeking). Under the assumption of Expected Utility maximization, participant i with utility function $u(x \mid r_i)$ chooses A_t over B_t if the difference in expected utilities is positive, that is,

$$\nabla_t(r_i) := p_t u(x_t \mid r_i) - q_t u(y_t \mid r_i) = \frac{p_t(1 - e^{-r_i x_t}) - q_t(1 - e^{-r_i y_t})}{1 - e^{-r_i x_{\max}}} > 0. \tag{2}$$

We then add an error term $\varepsilon_{it} \sim N(0, \sigma^2)$ with $\sigma^2 > 0$ to (2). That is, the lottery A_t is chosen if the following condition holds:

$$\nabla_t(r_i) + \varepsilon_{it} > 0 \tag{3}$$

Define the binary choice indicator for trial t

$$\gamma_{it} = \begin{cases} 1 & \text{if } A_t \text{ chosen by participant } i \\ -1 & \text{if } B_t \text{ chosen by participant } i. \end{cases}$$

Then the probability of a choice conditional on the risk-parameter r_i is given by

$$P(\gamma_{it} \mid r_i) = P\left(\gamma_{it}\nabla_t(r_i) > \gamma_{it}(-\varepsilon_{it})\right) = P\left(\gamma_{it}\frac{\nabla_t(r_i)}{\sigma} > \gamma_{it}\frac{-\varepsilon_{it}}{\sigma}\right) = \Phi\left(\gamma_{it}\frac{\nabla_t(r_i)}{\sigma}\right) \tag{4}$$

where Φ is the standard normal cumulative distribution function.

The conditional probabilities above were derived conditional on a participant's risk parameter r_i. In other words, estimating this model over the entire population would imply homogeneity in risk attitude across participants. In order to allow for between-participant heterogeneity, we let the risk attitudes vary across the population. In particular, we assume that the individual risk attitudes in the population are distributed normally in our participant pool according to

$$r \sim N(\mu, \eta^2).$$

Hence, the log-likelihood of a sample given by the matrix $\Gamma = (\gamma_{it})$ consisting of T trials and N participants is

$$\log L = \sum_{i=1}^{N} \ln \int_{-\infty}^{\infty} \prod_{t=1}^{T} \Phi\left(\gamma_{it} \frac{\nabla_t(r)}{\sigma}\right) f(r \mid \mu, \eta) dr \tag{5}$$

where $f(r \mid \mu, \eta) = \frac{1}{\sqrt{2\rho\eta^2}} e^{-\frac{1}{2}\left(\frac{r-\mu}{\eta}\right)^2}$ is the density function of the risk parameter r.

In order to evaluate the integral in (5) we use the method of maximum simulated likelihood (MSL) (see Train, 2009, for details). Specifically, we will approximate this integral by the following average

$$\frac{1}{H} \sum_{h=1}^{H} \left(\prod_{t=1}^{T} \Phi\left(\gamma_{it} \frac{\nabla_t(r_{ih})}{\sigma}\right)\right) \tag{6}$$

using a sequence of H (transformed) Halton draws $(r_{i1}, \dots, r_{iH})$ from $N(\mu, \eta^2)$ for each participant i (fixed over trials t). For the estimation, we use the Stata implementation "mdraws" of this procedure (Cappellari and Jenkins, 2003). Halton draws, a by-now-standard procedure, simulate random draws that ensure even coverage of the parameter space (e.g. avoiding clustering) using Halton sequences (Halton, 1960; Moffatt, 2015). Specifically, a Halton sequence is defined for a given prime number p, for example $p = 2$, is $(\frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{5}{8}, \frac{3}{8}, \frac{7}{8}, \frac{1}{16}, \frac{9}{16}, \dots)$. Such a sequence $(h_1, h_2, \dots)$ provide pseudo-random draws from the uniform distribution $U(0,1)$. To obtain draws from $N(\mu, \eta^2)$ we apply the following transformation $r_{ij} = \mu + \eta\Phi^{-1}(h_j)$ where Φ^{-1} is the inverse of the normal cumulative distribution function.

The MSL approach amounts to replacing the integral in (5) by (6) and then maximize the resulting function

$$\log \hat{L} = \sum_{i=1}^{N} \ln \frac{1}{H} \sum_{h=1}^{H} \left(\prod_{t=1}^{T} \Phi\left(\gamma_{it} \frac{\nabla_t(r_{ih})}{\sigma}\right)\right). \tag{7}$$

Maximization of (7) is carried out using standard MLE routines in Stata to obtain the estimates $(\hat{\mu}, \hat{\eta}, \hat{\sigma})$. Given those estimates we obtain the posterior expectation of each participant's risk attitude $\hat{r}_i$ conditional on their T choices applying Bayes' rule as follows

$$\hat{r}_i = E(r_i | \gamma_{i1}, ..., \gamma_{iT}) \approx \frac{\frac{1}{H} \sum_{h=1}^{H} r_{ih} \left(\prod_{t=1}^{T} \Phi\left(\gamma_{it} \frac{\nabla_t(r_{ih})}{\hat{\sigma}}\right)\right)}{\frac{1}{H} \sum_{h=1}^{H} \left(\prod_{t=1}^{T} \Phi\left(\gamma_{it} \frac{\nabla_t(r_{ih})}{\hat{\sigma}}\right)\right)}$$

for a sequence of Halton draws $(r_{i1}, \dots, r_{iH})$ from $N(\hat{\mu}, \hat{\eta}^2)$.

Given the estimated individual mean risk parameter $\hat{r}_i$, we obtain

$$\hat{u}_i(x) = \frac{1 - e^{-\hat{r}_i x}}{1 - e^{-\hat{r}_i x_{\max}}} \text{ for } \hat{r}_i \neq 0$$

as the estimated utility function of participant i.

Why Is Choice Stochastic?
Deliberate Randomization vs. Strength of Preference

ONLINE APPENDIX: Experimental Instructions for Experiment 1

[These are the written instructions given to participants. Text in brackets [...] was not displayed to participants.]

Welcome to this experiment. Thank you for supporting our research.

Please note the following rules:

1. Please do not talk to other participants.
2. If you have a question, please raise your hand. An experimenter will come to answer your question privately.
3. Please do not use any feature of the computer that is not part of the experiment.

General instructions

The experiment consists of two main parts, two additional questions, and a questionnaire. You will receive specific instructions for each part or question before the beginning of that part or question. In Part 1 and Part 2 you have to make decisions involving lotteries (on the next screen we will explain in detail what a lottery is). At the end of the experiment your payoff for Part 1 and Part 2 will be determined as follows:

- Part 2 consists of 168 rounds. The computer will choose one of these rounds at random. Your choice in this randomly selected round determines your payoff for Part 1.
- Part 2 consists of 37 rounds. The computer will choose one of these rounds at random. Your choice in this randomly selected round determines your payoff for Part 2.

In Part 1 and 2 you will earn points, which will be converted to Swiss Francs (CHF) using the following exchange rate: 10 points = 2.5 Swiss Francs (CHF). After Part 1 and 2, there will be two additional questions in which you have to choose how much to invest in a risky investment. You will be paid for both these additional questions. In these additional questions you will also earn points, which will be converted to Swiss Francs (CHF) using the following exchange rate: 10 points = 0.30 Swiss Francs (CHF). Independently of your decisions, you will receive an additional CHF 10 for your participation in the experiment. That is, your total earnings will consist of the sum of your earnings in Part 1 and Part 2 determined as explained above, and your earnings from the two additional questions plus a participation fee of CHF 10.

Instructions: Lotteries

In this experiment you will be asked to make decisions that involve lotteries. Therefore, we will now explain in detail what a lottery is. A lotteries pays a certain amount of points depending on chance. We present lotteries in the form of a table with two rows and four columns. The second row of each column shows a possible amount of points that the lottery can pay, whereas the first row shows the probability (in percentages) that this amount is paid. Below you see an example of a lottery.

Each lottery has four possible outcomes and each of those occurs with the same probability (25%). The second row shows all possible outcomes of the lottery in points. Note that some lotteries may show the same number of points more than once. Summarizing, this means that the lottery above pays either 40 points, 55 points, 70 points, or 80 points, each with a probability of 1/4 (25%).

25%	25%	25%	25%
40	55	70	80

Figure 7 **[Example of how lotteries were represented in the experiment.]**

Depending on chance, a lottery returns one of the four possible outcomes. The amount the lottery will actually pay is determined as follows: The computer will simulated the roll of a four-sided dice with sides numbered 1, 2, 3, and 4. The amount the lottery will pay depends on the result of the simulated dice role. If the result is 1, then the lottery returns the number of points in column 1. If the result is 2, then the lottery returns the number of points in column 2, and so on. For the lottery in the example above this means: If the result of the dice roll is 1, the lottery pays 40 points; if the result is 2, the lottery pays 55 points; if the result is 3, the lottery pays 70 points; and if the result is 4, the lottery pays 80 points.

Instructions for Part 1

This part consists of 168 rounds. In each round you will see two lotteries on screen, Option A and Option B. Here is an example:

	25%	25%	25%	25%
Option A	40	55	70	80
	25%	25%	25%	25%
Option B	30	50	80	85

Figure 8 **[Example of how the task was represented in Part 1 of the experiment.]**

Your task is to choose one of the two lotteries, Option A or Option B, by clicking on the corresponding button next to that lottery. In each round, right after you have made your choice both lotteries will be played out independently. That is, the computer will simulate independently the role of two dice, one for Option A and one for Option B. However, only the lottery you have chosen will be relevant for your payoff for Part 1. If this round is randomly selected by the computer, then you obtain the outcome of the lottery you have chosen. At the end of the experiment, you will be informed about the result of the dice roll for the lottery that determines your payoff. Remember that points will be converted to Swiss Francs (CHF) using the following exchange rate: 10 points = 2.5 Swiss Francs (CHF).

Instructions for Part 2

This part consists of 37 rounds. In each round you will see two lotteries on screen, Option A and Option B. In each round, your task is to select one out of five urns. Each urn contains four balls that can be either black or white. Here is how the five urns look like:

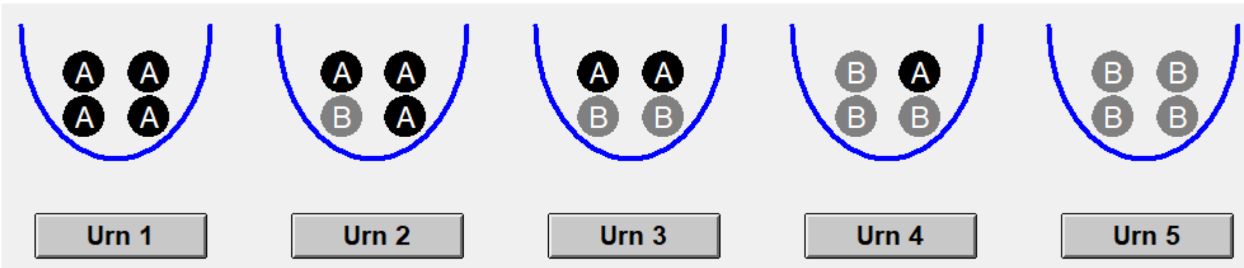

Figure 9 **[Example of how the task was represented in Part 2 of the experiment.]**

You can select an urn by pressing the corresponding button below that urn. The urn you have chosen determines the probability that you receive Option A and the probability that you receive Option B in the following way: The computer will draw one ball from the urn you have chosen. The color of that ball determines whether you receive Option A or Option B. If the ball is black, you will receive Option A, and if the ball is white you will receive Option B.

Comprehension Questions for Part 2

- If you choose Urn 1, what is the probability (in %) that you will receive Option A? [Urn 1 contained only black/A balls.]
- Suppose a ball is randomly drawn from Urn 4, what is the probability (in %) that it is grey? [Urn 4 had only grey balls]
- Suppose that a grey ball has been drawn from an urn. Which of the two lotteries, Option A or Option B, will you receive in that case? [Grey balls corresponded to Option B]

Instructions for additional questions

You will now be asked to answer two additional questions. In those questions you have to choose how much to invest in a risky investment. You will be paid for both these additional questions. In these additional questions you will also earn points, which will be converted to Swiss Francs (CHF) using the following exchange rate: 10 points = 0.30 Swiss Francs (CHF).

Additional question 1

You are endowed with 100 points and asked to choose the portion of this amount (between 0 and 100 points, inclusive) that you wish to invest in a risky option. Those points not invested are yours to keep. There is a 1/2 probability (50%) that the investment will fail and a 1/2 probability (50%) that it will succeed. If the investment fails, you lose the amount you invested. If the investment succeeds, you receive 2.5 (two and one-half) times the amount invested. To determine if the investment is successful or not, the computer will simulate the roll of a four-sided dice with sides marked 1, 2, 3, and 4. Your investment succeeds if the result of the dice roll is 1 or 4. If the result of the dice roll is instead 2 or 3 the investment fails.

Additional question 2

You are endowed with 100 points and asked to choose the portion of this amount (between 0 and 100 points, inclusive) that you wish to invest in a risky option. Those points not invested are yours to keep. There is a 1/2 probability (50%) that the investment will fail and a 1/2 probability (50%) that it will succeed. If the investment fails, you lose the amount you invested. If the investment succeeds, you receive 2.5 (two and one-half) times the amount invested. To determine if the investment is successful or not, the computer will simulate the roll of a four-sided dice with sides marked 1, 2, 3, and 4. Your investment succeeds if the result of the dice roll is 1. If instead the result of the dice roll is 4 the investment fails. If the result of the dice roll is 2 or 3, then the computer will simulate the roll of a second dice. Your investment succeeds if the result of the second dice roll is 1 or 4. If instead the result of the second dice roll is 2 or 3 the investment fails.

ONLINE APPENDIX: Experimental Instructions for Experiment 2

[These are the on-screen instructions given to participants. Text in brackets [...] was not displayed to participants.]

[General instructions]

Thank you for participating. This study investigates decision-making under risk.

On top of your fixed earnings, you will earn a bonus payment which will depend on your decisions in the study.

Please read all questions carefully. Completing the survey will take about 15 minutes.

[Consent]

If you have any queries or if you are unhappy with anything that happens concerning your participation in the study, please contact the experimenter:

(Name and E-Mail of corresponding author)

If you have any concerns or complaints that you wish to discuss with a person who is not directly involved in the research, you can also contact:

(Name, E-mail, and address of Department Head)

When you are ready, please press the button below to start the experiment. By continuing you explicitly give us your consent that:

- We can collect your anonymous, non-sensitive personal data (like age, gender, etc.).
- We can use this personal data for academic research purposes.
- We can store your personal data on our safe-guarded university servers for up to 10 years.
- We can make anonymized data available to other researchers online.

For further information about how Lancaster University processes personal data for research purposes and your data rights please visit our webpage: www.lancaster.ac.uk/research/dataprotection

[Payment]

Your bonus payment depends on your decisions and chance. The study involves choices between risky options.
An Option pays one of four potential monetary outcomes, each occurring with a 25% chance.

Here is an example: Option A: 5 ECU with 25% chance, 10 ECU with 25% chance, 50 ECU with 25% chance, and 100 ECU with 25% chance.

ECUs are Experimental Currency Units, 10 ECUs are 1 Pound.

After the study one of the questions will be randomly selected. The Option that you chose for this question will be played out, and you will be paid according to the resulting outcome.

Each decision could be the one that counts for your payment. It is therefore in your best interest to consider all your answers carefully.

[Explicit Randomization]

You will be asked a number of simple questions.

For each question, you can either choose one of the two options or delegate your choice to a coin flip (a random device which will select one of the two options with 50% chance).

[Tax]

In some of the questions a "tax" will be introduced.

The tax is a reduction on the earnings from that question. That is, the monetary amounts of the options in that question will be reduced by 50

For example, suppose that you chose this Option:
Option A: 5 ECU with 25% chance, 10 ECU with 25% chance, 50 ECU with 25% chance, and 100 ECU with 25% chance.

Suppose that this question is selected for bonus payment and then the computer selects the "50 ECU" outcome. If the question did contain a tax, you will be paid 25 ECU instead. If the question did not contain a tax, you will be paid the full 50 ECU.

[Choice Screen]

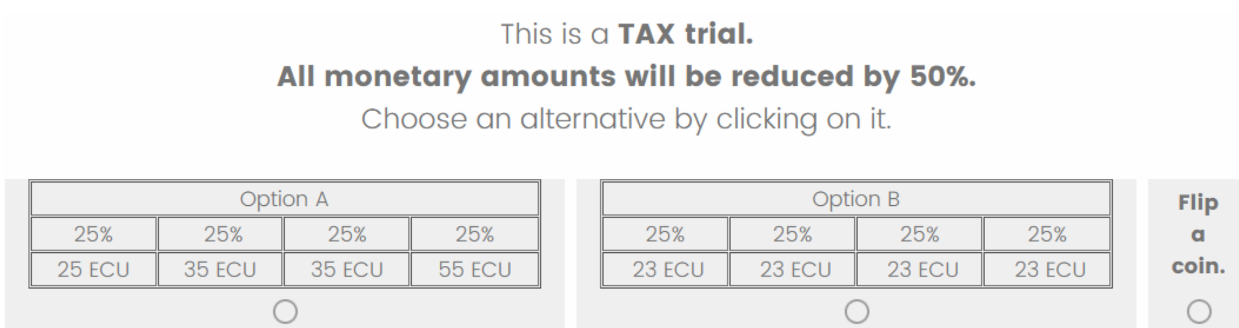

Figure 10 **[Example of how the task was represented for tax trials in the experiment.]**

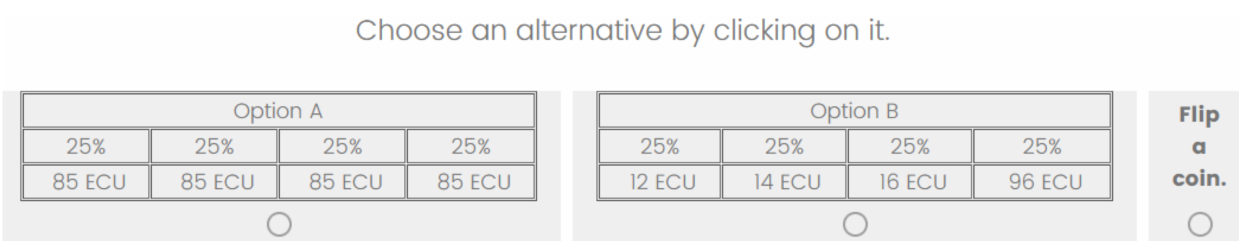

Figure 11 **[Example of how the task was represented for non-tax trials in the experiment.]**

ONLINE APPENDIX: Experimental Instructions for Experiment 3

[These are the on-screen instructions given to participants. Text in brackets [...] was not displayed to participants.]

[General instructions]

Thank you for participating. This study investigates decision-making under risk.

On top of your fixed earnings, you will earn a bonus payment which will depend on your decisions in the study.

Please read all questions carefully. Completing the survey will take about 30 minutes.

[Consent]

If you have any queries or if you are unhappy with anything that happens concerning your participation in the study, please contact the experimenter:

(Name and E-Mail of corresponding author)

If you have any concerns or complaints that you wish to discuss with a person who is not directly involved in the research, you can also contact:

(Name, E-mail, and address of Department Head)

When you are ready, please press the button below to start the experiment. By continuing you explicitly give us your consent that:

- We can collect your anonymous, non-sensitive personal data (like age, gender, etc.).
- We can use this personal data for academic research purposes.
- We can store your personal data on our safe-guarded university servers for up to 10 years.
- We can make anonymized data available to other researchers online.

For further information about how Lancaster University processes personal data for research purposes and your data rights please visit our webpage: www.lancaster.ac.uk/research/dataprotection.

[Payment]

Your bonus payment depends on your decisions and chance. The study involves choices between risky options.
An Option pays one of four potential monetary outcomes, each occurring with a 25% chance.

Here is an example: Option A: 5 ECU with 25% chance, 10 ECU with 25% chance, 50 ECU with 25% chance, and 100 ECU with 25% chance.

ECUs are Experimental Currency Units, 10 ECUs are 1 Pound.

After the study one of the questions will be randomly selected. The Option that you chose for this question will be played out, and you will be paid according to the resulting outcome. Each decision could be the one that counts for your payment. It is therefore in your best interest to consider all your answers carefully.

[Explicit Randomization]

You will be asked a number of simple questions. For some questions you just have to choose one of the two options. For other questions, you can either choose one of the two options or delegate your choice to a coin flip (a random device which will select one of the two options with 50% chance). Choosing the coin flip is costly. We will subtract this cost from your potential earnings. The cost of choosing the coin flip changes from question to question. You will be told the cost for each question.

[Choice Screen]

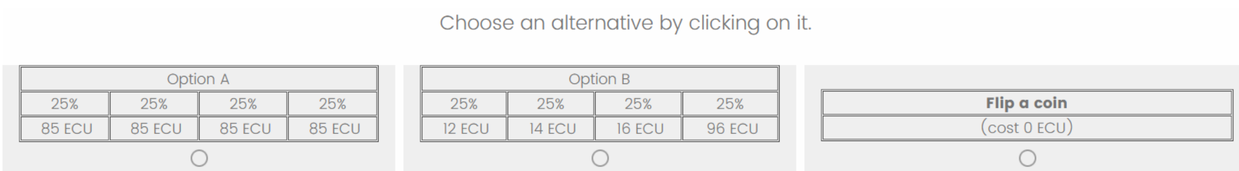

Figure 12 [Example of how the task was represented in the experiment.]

[Additional Questions]

Are you aware that some of the questions where repeated?

[If they replied no, the program skipped the following two questions. Each question was reported on a different screen.]

Did you try to be consistent between the repetitions of the questions?

[If they replied yes, the program skipped the following question.]

Could you tell us why you gave different answers to the same questions?