Stochastic Choice and Preferences for Randomization — Source link

Marina Agranov, Pietro Ortoleva

Institutions: California Institute of Technology, Columbia University

Published on: 01 Feb 2017 - Journal of Political Economy (University of Chicago PressChicago, IL)

Topics: Choice set

Stochastic Choice and Preferences for Randomization*

Marina Agranov[†] Pietro Ortoleva[‡]

January 2015

Abstract

We conduct an experiment to investigate the origin of stochastic choice and to differentiate between the three main classes of models that account for it: Random Expected Utility; Mistakes; and Deliberate Randomization. Subjects face the same questions multiple times in two ways: 1) following the literature, with repetitions distant from each other; 2) in a novel treatment, with repetitions *in a row*, telling subjects that questions will be repeated. A large majority of subjects exhibited stochastic choice in *both* cases, and stochasticity is strongly correlated in the two cases. Our results support the class of models of Deliberate Randomization.

JEL: *C91, D11, D81.*

Keywords: *Stochastic Choice, Random Utility, Mistakes, Hedging*

*A previous version of this manuscript circulated under the title "Stochastic Choice and Hedging." We thank Mark Dean, David Dillenberger, Federico Echenique, Kfir Eliaz, Guillaume Frechette, Yoram Halevy, Mark Machina, Efe Ok, Gil Riella, Kota Saito, Chris Shannon, Leeat Yariv, and participants at various seminars and conferences for helpful comments and suggestions. Ortoleva gratefully acknowledges the financial support of NSF grant SES-1156091 and of that of the Caltech's Gordon and Betty Moore Foundation through Grant GBMF1158.

†Division of the Humanities and Social Sciences, California Institute of Technology, 1200 E. California Blvd, Pasadena, CA 91125 USA. Email: magranov@hss.caltech.edu.

‡Department of Economics, Columbia University, 420 West 118th St., New York NY 10027 USA. Email: pietro.ortoleva@columbia.edu.

1. Introduction

A consistent finding of individual decision making is that when subjects are asked to choose from the same set of options many times, they often make different choices. This phenomenon is often called stochastic, or random choice.[1] Extensive research has documented how widespread this pattern is, in different settings and populations, both in the lab and in field data. Moreover, it has been shown to involve a significant fraction of choices, even within the course of a single experimental session.[2] Importantly, stochastic choice is documented also in environments in which subjects have no value for experimentation (e.g., when there is no feedback), or when there are no bundle or portfolio effects (e.g., when only one choice is paid). For these reasons, stochastic choice is often seen as incompatible with the typical assumption in economics that subjects have a well-defined, complete, and stable preference ranking (or utility) and consistently choose the option that maximizes it.

This robust evidence has led to the development of a very large body of theoretical models designed to capture this behavior. These models belong to three 'classes': 1) *Mistakes*, according to which subjects have stable preferences but exhibit stochastic choice as they make mistakes when choosing;[3] 2) *Random Utility/Preferences*, according to which subjects' answers change because their preferences change stochastically;[4] 3) *Deliberate Randomization*, according to which subjects deliberately choose to report different or stochastic answers, following a specific preference originating from the desire to reduce regret, incomplete preferences, difficulty to judge one's true risk aversion, or other forms of Non-Expected Utility.[5] Amongst them, models in the first

[1]To avoid confusion, note that in economics these terms are used to denote two very different phenomena: 1) the situation in which *one* person faces the same question *multiple times* and gives different answers; 2) the situation in which *different* subjects answer the same question only *once*, but subjects who appear similar given the available data make different choices. In this paper we focus on the first phenomenon.

[2]The pattern of stochastic choice was first reported in Tversky (1969), in which subjects were presented with 10 pairs of monetary gambles 20 times, separated by decoys. A very large number of experimental studies have then replicated these results. Focusing on choices between risky gambles (as in our experiment), see, amongst many, Camerer (1989a), Starmer and Sugden (1989), Hey and Orme (1994), Ballinger and Wilcox (1997), Hey (2001), Regenwetter et al. (2011) and Regenwetter and Davis-Stober (2012). Dwenger et al. (2013) show similar patterns in real-world data on university choices in Germany, and suggest other environments where it is common.

[3]This is one possible interpretation of the well-known model of Luce (1959), and models of this kind appear, amongst many, in Harless and Camerer (1994), Hey and Orme (1994), Camerer and Ho (1994), Wu and Gonzalez (1996), and Fudenberg and Strzalecki (2014).

[4]Models of this kind are discussed in Thurstone (1927), Harsanyi (1973), Falmagne (1978), Cohen (1980), Barberá and Pattanaik (1986), McFadden and Richter (1991), Loomes and Sugden (1995), Clark (1996), McFadden (2006), Gul and Pesendorfer (2006), Ahn and Sarver (2013), and Fudenberg and Strzalecki (2014).

[5]Models of this kind are suggested by Machina (1985), Marley (1997), Cerreia-Vioglio et al. (2014b), Dwenger et al. (2013), Fudenberg et al. (2014), Fudenberg and Strzalecki (2014), Swait and

two classes are by far the most popular: to our knowledge, models in the third class have not yet been used to analyze experimental or field data. We refer to Section 4 for an in-depth discussion.

The goal of this paper is to shed light on the origin of stochastic choice and in particular to distinguish between the three broad classes of theories described above. The main question we are interested in is: Can stochastic choice be seen as a deliberate choice of the individual, as in the third class of theories above, or is it something that subjects just happen to express without this being their preference, as in the first two more popular classes of theories? We design a controlled laboratory experiment to answer this question.

We believe that understanding the origin of stochastic choice is relevant for three reasons. First, stochastic choice is such a widespread phenomenon that for economics it is important to understand the underlying dynamics in order to use the appropriate model to study it (where, as we shall see, different models have different empirical predictions). Second, this understanding is relevant to determine the correct welfare implications and the need for corrective measures: if stochastic choice is, for example, a mistake, then it is welfare-reducing, and it may be possible to 'correct' it; but if it is a deliberate choice of the agent, then it may be optimal not to prevent it. Third, understanding the origin of stochastic choice will help us correctly interpret data from the field or from the lab: if stochastic choice is not a decision of the agent, but, say, a mistake, then it is not an expression of preferences and could be 'disregarded.' If instead it is a deliberate choice, then it can be seen as an explicit expression of preferences and it should be used as a potential source of information.

We use the following design. We have one main experiment and one robustness test. We begin by describing the design and results of the main experiment. Subjects are asked to make several choices between objective lotteries.[6] In the first part of the experiment we replicate standard designs that establish the presence of stochastic choice at an individual level: subjects are asked a set of questions repeated several times, with each repetition separated by other questions; moreover, subjects are not told in advance of these repetitions. To shed light on the underlying mechanism that leads to stochastic choice, we then introduce the following novel experimental feature: we include a new part of the experiment in which subjects face the same question three times *in a row*, and, importantly, are explicitly told that each question would be repeated three times. The key intuition is that observing different answers in this case suggests that this is a deliberate choice of the subjects. As an additional test, in a separate part subjects were allowed to choose either one of two lotteries, or to choose to flip a (computer-simulated) coin to determine which lottery they would be

Marley (2013)

[6]We focus on choice between objective lotteries as this appears to be one of the typical environments in which stochastic choice has been documented empirically (see footnote 2). At the same time, our design could be easily generalized to different frameworks.

assigned to them. Selecting the coin had a small fee. The choice to flip a costly coin would again show a strict preference for stochastic answers. We also elicit subjects' attitudes towards risk, compound lotteries, and their proneness to violate Expected Utility as captured by the Allais paradox. Finally, at the end of the experiment subjects are given un-incentivized questionnaire in which they are asked to motivate some of their choices.

Our main findings are the following. First, in line with previous experiments, the vast majority of participants (90%) exhibit stochastic choice by selecting different lotteries in the repetition of the same question when they are not explicitly told about the repetition. The subjects who exhibit stochastic choice tend to do so more than once, and for more than half of the questions in the class of questions where stochastic choice is prevalent (see below).

Second, we find that a large majority of subjects (71%) select different lotteries also when questions are repeated three times *in a row* and subjects are explicitly told that the same question will be repeated three times. That is, stochastic choice is far from being eliminated in this case. We believe that this is the main finding of our experiment, and it suggests that subjects are choosing stochastic answers on purpose. Also in this case, those who exhibit stochastic choice do so multiple times. Importantly, stochastic choice behavior is strongly correlated in the two cases (when subjects are aware of the repetition, and when they are not). It is also significantly correlated with violations of EU *à la* Allais (which will be relevant for the link with theoretical models), but not with risk attitudes. An analysis of the response times is compatible with subjects who, knowing that repetitions are coming, spend a relatively long time the first time they see each question, and then make their choice in the repetitions of that question very rapidly, as if implementing a plan.

Third, a strong additional pattern emerges across the experiment: stochastic choice is present almost exclusively in questions in which none of the available option is 'clearly' better than the other – what we call 'hard' questions.[7] Stochastic choice is extremely frequent for these questions, and virtually absent for others. Importantly, 'hard' questions are not necessarily the ones where the expected value (or utility) of the options are similar. In particular, once we control for the 'difficulty' of the question, the difference in Expected Value (or Expected Utilities as measured using CRRA functional form given risk elicitation task) is not a significant predictor of stochastic choice in our data. This is in contrast with the typical finding that stochastic choice behavior is highly correlated with how similar are the Expected Values/Utilities of the lotteries involved.[8]

[7] We will later describe in detail how 'hard' questions were identified.

[8] As we shall discuss, our results can be easily reconciled with the previous experimental literature once we note that in most existing experiments 'hard' questions are also those with very similar expected values (or utilities). By contrast, our experiment allows us to separate these two forces and we find the main predictor of stochastic choice to be the difficulty of the questions and not the

Fourth, we find that about 29% of the subjects choose the option to flip a costly coin, again virtually only for 'hard' questions. Like before, those who choose this option tend to do so multiple times. This suggests that a substantial fraction, albeit a minority, of the subjects, is willing to explicitly pay a cost to have the computer randomize on their behalf.

Finally, when asked in the questionnaire why did they choose different answers when the same question was asked three times in a row, subjects' typical answer was that they did so because they didn't know which option was best, and thus didn't want to commit to a specific choice.[9]

To test the robustness of our main result we have conducted another experiment at the California Institute of Technology. We designed it to address the concern that observing stochastic choice in the part of our experiment where repetitions are explicit is due to specific features of our design and not to a desire to report stochastic answers. For example, one may worry that switching behavior could be due to the fact that, since only one question is paid, stakes are effectively low and subjects may have no incentives to think about what to choose. Or that, because in other parts of the experiment subjects are offered the possibiliy to flip a coin, this may induce them to want to randomize in other questions. Lastly, subjects may not trust the randomization by the computer, and may thus prefer to hedge their options between different answers.

To address them, in our robustness experiment we have asked only one question, repeated three times, where subjects were told that the same question would be repeated. That is, we have a task very similar to the part of the main treatment with explicit repetitions, but with only that task and for only one question (asked repeatedly). We have also used much higher stakes (10 times higher for each question, with only three instead of tens of questions asked). In addition, the experiment was conducted with pen and paper, and subjects were asked to perform the randomization themselves (rolling a die). Reassuringly, we find that also in this much simpler experiment the same main result holds: 50% of the subjects choose different answers to the same question – a number very similar to the 45% observed for the same question in the main treatment. Thus, our main finding appear to be robust to various changes including fewer questions and much higher stakes.[10]

We then use the findings above to distinguish between existing models of stochastic choice. We focus on the three classes of models discussed above: Random Expected Utility, Mistakes, and Deliberate Randomization. The predictions of these models

difference in EV/EU.

[9]As this was an open question, the precise wording was very heterogeneous, but the general meaning for a large majority of subjects was rather stable and in line with what reported above.

[10]Moreover, this experiment was conducted at a different institution than the main treatment: it was run at the California Institute of Technology, whose students are well known for their high computational skills.

depend on whether, in the part of the experiment when questions are repeated three times in a row and subjects are explicitly told about it, they treat each repetition as a whole new question, or they view the three questions together. Our analysis of the response time suggests that the latter is the case; this is also supported by the answers to the questionnaire, and it is quite natural, since this point was made very explicit in the instructions.[11] Under this interpretation, models of Random Expected Utility predict that subjects should not report stochastic answers in this part of the experiment – since in these models agents have complete and well-defined preference in any moment in time – in contrast to what we observe (our second main finding described above). Similarly, under models of Mistakes for these questions we should observe only 'trembling hand' kind of mistakes, which again is not compatible with our data since stochastic choice is present only for some types of questions. Models of Mistakes, moreover, also predict that stochastic choice should be correlated with the difference in Expected Values or Utilities of available options. As discussed above, this is also not the case in our data.[12] By contrast, models of Deliberate Randomization are compatible with stochastic choice behavior in both parts of our experiment. They also predict that such stochasticity should be correlated in the two parts of the experiment, question by question. This prediction also finds strong support in our data. The motivation given by subjects for reporting stochastic answers in the final questionnaire is also compatible with models of Deliberate Randomization.

We should emphasize that we do not interpret our findings as suggestive that mistakes or to changes in the utility do not happen – not only we do not have sufficient evidence to claim this, we also believe that both occur frequently.[13] Rather, we interpret our results as suggestive of the fact that some *other* force is also at play in generating stochastic choice: namely, the *desire* to choose different options. This desire for stochastic answers appears to be a strong driving force of stochastic behavior in our data. This suggests that models that account for such desire capture an important feature of behavior.

Our paper relates to the experimental literature that studies decision making under uncertainty (see Camerer 1995 for a comprehensive survey), and in particular, to the papers that investigate the source of stochastic choice behavior and preferences for randomization. Hey and Carbone (1995) design an experiment to test the possibility that preferences are deterministic but choice is actually stochastic; their results rule out the proposed framework. Sopher and Narramore (2000) design an experiment in

[11]That is, subjects were explicitly told that the same question would be repeated three times in a row, and saw this happening in all questions in this part. This is particularly strong in the robustness test, where subjects faced *only* one question repeated three times in a row.

[12]Notice that this rules out also other possible explanations, such as the fact that subjects are just indifferent, or are discovering their preferences.

[13]Coherently, we find that the proportion of subjects who report inconsistent answers when questions are asked in a row is smaller than the one observed when repetitions are separated (71% vs. 90%), in line with the presence of either mistakes or random utility.

which subjects can choose between two lotteries or any convex combination of them. They document a strong tendency to choose a mixture of lotteries rather than the extremes, in line with convexity of preferences.

Rubinstein (2002) documents a deliberate desire to report 'diversified' answers, even when this is explicitly disadvantageous: in hypothetical questions, many subjects choose to diversify even when this leads them to strictly First Order Stochastically dominated choices.[14] We document similar patterns, even though in our case the tendency to diversify takes place virtually only for 'hard' questions, and does not generate violations of FOSD, and we link it to stochastic choice.

Kircher et al. (2013) conduct an experiment where subjects play a version of the dictator game in which the dictators can choose one the following actions: option A gives 7.5 Euros to him/her and 0 to the recipient; option B gives 5 Euros to both; and option C is a randomization between options A and B with equal chances. They find that about one third of the subjects chose option C, which is incompatible with Expected Utility and shows a strict preference for randomization. These results suggest that the preference for randomization may play an important role in social settings. Our results show that they may play a role also in settings in which no social component is present.[15]

Finally, in a recent and independent study, Dwenger et al. (2013) explore the possibility that subjects may prefer to avoid making decisions and thus choose to delegate their choice to an external device both in the laboratory and in real data. They find that relevant fractions (between 15% and 53%) of subjects choose lotteries between available allocations, indicating an explicit preference for randomization, and discuss how their experimental data is consistent with a theory of responsibility aversion, a variant of the regret theory by Loomes and Sugden (1982). As opposed to our work, they don't discuss the relation with stochastic choice.[16]

The remainder of the paper is organized as follows. Section 2 describes the experiment design, for both the main treatment and the robustness test, and Section

[14]Several recent studies document a pattern of behavior related to false diversification. Chen and Corter (2006) document that subjects chose irrational mixture of options, including the dominated ones, in multiple-trial decisions over pure bundles of either option. Eliaz and Fréchette (2008) show that a significant fraction of subjects are willing to pay to switch from a lottery that pays a prize in only one state of nature to a lottery that pays a prize in more than one state, even though the overall distribution over prizes remained constant.

[15]Kircher et al. (2013) also conduct a treatment in which the decision involves only one player who has to choose if she wants to buy a mug for a given price or not, or to randomize. In this case they document a very small preference for randomization. This compatible with our results given that we observed such preferences only for 'hard' choices, but not for others.

[16]We should also mention that the popular blog of Freakonomics has now started offering a service that allows visitors to flip a (simulated) coin to take hard and important decisions. As of this writing, tens of thousands of visitors have used this service, again in line with a desire for randomization when facing hard decisions. See `www.freakonomicsexperiments.com`.

3 presents our results. Section 4 discusses the implications on existing models of stochastic choice data. Section 5 concludes. The appendix contains a more detailed analysis of the response time, screenshots of the experiment, and the instructions given to the subjects.

2. Experiment

2.1 Design: Main Treatment

2.1.1 Basic Questions In most rounds of the experiment subjects were asked to select one of two lotteries displayed on the screen. Each lottery pays a certain amount of tokens depending on the roll of a four-sided fair die with faces named A, B, C and D (simulated by a computer). Thus, all the lotteries have at most four different outcomes, occurring with probabilities 0.25, 0.5, 0.75, or 1. Each lottery was presented as a table that lists the number of tokens paid for each face of the die. Figure 1 contains one example of a lottery that pays 16 tokens if faces A or B are obtained, and 94 tokens if faces C or D are obtained.

A	B	C	D
16	16	94	94

FIGURE 1 Example of a Lottery

Throughout the experiment, we ask, repeatedly, the same 10 basic questions, each involving the choice between two lotteries. These 10 questions can be divided into three groups: FOSD, EASY, and HARD. The first group (FOSD), comprised of three questions, involved the choice between two lotteries one of which was first-order stochastically dominated by the other. The EASY questions had two lotteries with no first order stochastic dominance, but in which one was 'clearly better' than the other, in the sense that it was designed so that most subjects would have little doubt about which option would be best. By contrast, each of the four HARD questions had no 'obvious winner' among two lotteries. In practice, to design these questions we run a pilot experiment with many questions in which we recorded both choices and the response time. We then identified EASY and HARD questions looking at two dimensions. First, EASY questions had a unanimous consensus on which lottery is better, while HARD questions had choice distribution close to uniform. Second, response time was higher in HARD than in EASY questions. Table 1 lists all the questions we asked.

Question	*Lottery 1*				*vs.* *Lottery 2*				*Diff. in EV*
FOSD1	98	98	98	98	103	103	103	103	5
FOSD2	17	17	18	18	17	17	17	17	0.5
FOSD3	10	20	30	30	70	100	120	190	97.5
EASY1	23	23	30	30	5	5	5	31	15
EASY2	12	14	16	96	85	85	85	85	50.5
EASY3	100	100	100	100	20	20	20	101	59.75
HARD1	38	38	38	77	16	16	94	94	7.25
HARD2	10	10	90	90	32	45	45	56	5.5
HARD3	6	84	105	200	54	60	117	135	7.25
HARD4	13	30	51	81	19	32	38	86	0

Table 1 List of Questions Asked. Each question is one row: subjects were asked to choose between lotteries 1 and 2. Each lottery is described by the four money amounts that it can pay out with equal probability. (Appendix B presents a screenshot of a typical question). The last column contains the absolute value of the difference in Expected Values.

2.1.2 Structure of the experiment The experiment is comprised of four parts. We present below the details of each part of the experiment, and refer the reader to Appendix C for the complete instructions. Subjects received general instructions about the experiment and specific instructions about Part I when they entered the room. Separate instructions were distributed (or displayed on the screen) at the beginning of each of the following parts. In Parts I, III, and IV subjects faced several decision rounds in which they were asked to choose one of the lotteries presented on the screen; some of these questions appeared repeatedly. In Parts II, we separately elicited subject's attitudes towards risk and compound lotteries. Subjects received no feedback throughout the experiment.[17] Only at the end participants learned how much money they earned, and were paid in cash privately.

The main parts of the experiment are Part I and III, in which we observe subjects' stochastic choice behavior following different procedures. In Part I subjects played 40 decision rounds, constructed by asking four times each of the ten questions described above (Table 1). During each of the four repetition of the list of questions, the order of questions within the list was changed, and the orders were designed to guarantee that

[17]However, after each decision round, the lottery selected by a subject was played out and its outcome recorded by the software (but not revealed to the subject). Subjects were informed in the instructions that the realization of the selected lottery took place right after they made every choice. This was done to elicit one's preferences over lotteries at the moment she encounters the problem, rather than her belief about which lottery she would prefer at the end of the experiment when the choice is actually realized.

each question would not appear too close to its repetition.[18] This means that in Part I, each pair of lotteries was presented four times. During three of these four times, subjects observed exactly the same question, displayed in exactly the same way. The other time they saw each question, in addition to the two lotteries they were also offered the option to have the computer flip a (simulated, fair) coin to determine which of the two lotteries would be assigned to them. (Figure 5 in Appendix B contains a screenshot.) This third option, however, was not free: if this question was selected for payment, 1 token would be subtracted from the amount won. The cost of flipping a computer simulated coin was explained in details in the instructions.

The second main part of the experiment is Part III, in which subjects were asked again seven out of the ten basic questions: FOSD1, EASY1, EASY2, HARD1, HARD2, HARD3 and HARD4. The main difference with Part I is that here each of the questions was asked three times *in a row* (one after the other), for a total of 21 decision rounds. Subjects were explicitly told that each question would be repeated three times in a row, and this point was highly emphasized in the instructions.

Between Part I and III subjects faced Part II that included two different tasks designed both to break the repetitiveness of the questions, with the goal of reducing fatigue, and to measure risk aversion and attitude towards compound lotteries. In the first task, developed by Gneezy and Potters (1997) and Charness and Gneezy (2010), subjects were endowed with 100 tokens, any number of which could be invested in a risky asset that has a 0.5 chance of success, and a payoff of 2.5 times the investment if successful; whatever is not invested is kept. The amount of tokens not invested is proportional to the degree of risk-aversion of a subject who obeys Expected Utility. The second investment task was identical to the first, except that the success of the investment was determined by a compound lottery which reduces to a 0.5 chance of success. A subject with no aversion nor attraction to compound lotteries should invest the same number of tokens in both questions, while a subject who dislikes (likes) compound lotteries might decide to invest less (more) in the second rather than in the first investment task.

Finally, in Part IV of the experiment we employ standard technique to elicit subjects' non-Expected Utility preferences, by asking variations of the Common Ratio and Common Consequences effects of Allais (1953).[19]

[18]To test for the order effects in the presentation of questions, each subject was randomly assigned to one of the two possible orders, in which we vary both the order of questions, and the order of options within each question. Importantly, the latter was varied across individuals, but was not changed when questions were repeated for a given subject: each repetition of the same question was displayed exactly in the same way. See Appendix E for the exact order of questions. We find no order effects of any kind. (For example, all results hold if we focus only on specific orders.)

[19]Appendix C presents the exact specification of these questions.

2.1.3 Payment Scheme and Location Subjects received a total payment composed of several parts. First, subjects were paid for *one* randomly selected decision round from all rounds in Parts I and III combined. Each of these decision rounds was equally likely to be selected for payment. This procedure was used to avoid the incentive to choose different lotteries from the same question in order to create a portfolio (and distribute risks) – such choice behavior would be consistent with standard Expected Utility, and it is different from stochastic choice. By contrast, if only one question is paid, any Expected Utility agent would maximize her return by choosing in each round her preferred lottery.[20] Once a decision round is selected, the subject was assigned a certain number of tokens depending on her choice and on the realization of the lottery. These were then converted into US dollars using the rate 20 Tokens = $1, and the dollar amount was added to final payment.

Since Parts II and IV of the experiment are less prone to portfolio concerns, subjects were paid for all decisions made in these parts of the experiment. The conversion rate for the tokens earned in Parts II and IV, which were on average much higher, was 100 Tokens = $1.[21] In addition, participants received a $10 show-up fee for completing the experiment.

At the end of the experiment, subjects were asked to fill a non-incentivized questionnaire that contained several questions regarding what motivated them to make their choices in Parts I and III of the experiment.[22]

All sessions of the main treatment were conducted at the California Social Science Experimental Laboratory (CASSEL) at University of California, Los Angeles in January 2013.[23] Subjects were recruited from a database of volunteer undergraduate students. Four identical sessions were run, using a total of 80 subjects. No subject participated in more than one session. Each session lasted about 45 minutes, and average earnings were $19.

We conclude this section by discussing some features of our experimental design. First, Part I of the experiment is in line with many other experiments in the literature that record the phenomenon of stochastic choice. We use this part to reproduce stochastic choice behavior. The explicit and immediate repetition of the same ques-

[20]Subjects who violate Expected Utility may instead have a different attitude: for example, they may want to report different answers to the same question repeated multiple times. This is important for the interpretation of our results, and it will be discussed at lengths when we analyze the relation between of results and theories of Deliberate Randomization in Section 4.3.

[21]The smaller conversion rate used in Parts II and IV than in Parts I and III was chosen to reflect the difference in the scales of earnings in this part relative to the possible lotteries' outcomes, and to create strong incentives for subjects to think hard about their choices in Parts I and III, that are the main focus of the experiment.

[22]Appendix C includes the complete list of questions we used in the questionnaire.

[23]The software was programmed as server/client applications in Java, using the open source experimental software Multistage (http://multistage.ssel.caltech.edu/).

tion present in Part III is instead novel, at least to our knowledge. The advantage of this kind of repetition is two-folds. The fact that repetitions in Part III are very explicit eliminates the concern that the inference of the stochastic choices is, in part, a result of subjects not being aware (or have forgotten) that they have encountered same questions in the previous rounds of the experiment and/or making a mistake in recollecting their past answers. The fact that repetitions are immediate and, more importantly, that the presentation of lotteries is identical in each of these repetitions, reduces the plausibility that subjects's valuation of available lotteries changes between questions, and that they make mistakes in answering. We will address this point at length when we present the results of our experiment. Finally, since Parts I and III contain the same questions, our design provides an opportunity to study the individual behavior of subjects in different parts, rather than infer sample statistics from observed distribution of choices.

2.2 Design: Robustness Test

To test the robustness of our main result we have conducted another experiment. We aim to address the potential concern that subjects can choose different options in Part III of our main treatment for reasons related to specific features of our design and not to their desire to randomize.[24] First, this may happen because subjects in our experiment are asked very many questions, the stakes can be too low to incentivize subjects to think hard about their choice. If this were the case, the presence of inconsistent answers in Part III could simply indicate that subjects did not care enough about their choices to think hard about them. A second concern is that, before reaching Part III, subjects were exposed to the option to flip a coin in Part I. This may have led them to think about randomizing, something that they otherwise wouldn't do. Finally, one may think that subjects have a limited trust in the experimenter and worry that the randomizations done by the computer are not performed fairly. Under specific assumptions, this could potentially generate a desire to express different answers to the same question in Part III.

Our robustness experiment was conducted at the California Institute of Technology with 26 undergraduate students using pen and paper.[25] After the introductions were read aloud to the subjects (see Appendix D), they were asked to answer question HARD2 three times in a row;[26] like in Part III of the main treatment, they are explicitly told that the same question would be repeated three times. That is, this

[24]We thank the audiences at various seminars and conferences for raising them.

[25]We should note that, compared to other schools, the average student at Caltech is considered as having very high computational skills (e.g., much higher SAT scores).

[26]Recall that in question HARD2 subjects are asked to choose between lottery 1, which pays 10 tokens with probability 50% and 90 tokens with probability 50% and lottery 2, which pays 32 tokens with probability 25%, 45 tokens with probability 50% and 56 tokens with probability 25%.

robustness test mimics the design of one repetition of questions in Part III – but includes only that. Each repetition of the question was printed on a separate sheet of paper, and subjects had to circle their selected lottery in each sheet.[27] At the end of the experiment, each participant was asked to perform herself the randomization that determined the payment: first, she rolled a die to determine which question will be paid; then, she rolled a die that determined her payment in the selected lottery in the selected question. Finally, subjects were asked an un-incentivized questionnaire on the reasons behind their choices. For this experiment we used much higher stakes, including a much higher conversion rate: 2 tokens = \$1 (while in the main treatment it was 20 tokens = \$1). Note that this makes the stakes in this robustness test unusually high.[28] The experiment lasted 15 minutes in total (including the instruction period).

The robustness treatment thus contains the key novel test of the main treatment – the same question asked three times in a row, where subjects are told about the repetition, as in Part III. As opposed to the main treatment, however: it involves stakes that are enormously higher for each question, and quite high in general for experiments (there are only 3 questions instead of tens; and a 10-times-higher conversion rate); there is only one repetition of questions (reducing the cognitive load), that does not follow Part I of the main treatment (thus subjects are not suggested the option to flip a coin); it involves a different population, with extremely high average computational skills; finally, subjects performed the randomizations themselves (thus eliminating, or reducing, trust issues). We thus believe that documenting in the robustness test a behavior similar to that of the main experiment would address the concerns expressed above.

3. Results

We present the results of our experiments in the following order. We start by describing the behavior in Parts I and III of the main treatment, when subjects faced the same questions repeatedly, with repetitions that are far apart (Part I) or adjacent (Part III). We then discuss the results of the robustness test. Next, we go back to the main treatment and we analyze the connection between choice patterns in the two parts, as well as with the choice of the costly coin. We then discuss the answers to the un-incentivized questionnaire. Finally, we examine the response time and whether stochastic choice is related to risk attitudes or Allais-like behavior.

[27]This was done to mimic as close as possible the design of Part III of our original experiment, in which each repetition of a question in Part III appeared on a different screen.

[28]Applying the conversion rate, lottery 1 in this experiment pays \$5 with probability 50% and \$45 with probability 50%, while lottery 2 pays \$16 with probability 25%, \$22.5 with probability 50% and \$28 with probability 25%.

3.1 Behavior in Each Part: Main Treatment

Figure 2 presents the fraction of subjects who give inconsistent answers in each question in Parts I and III, as well as combined measures to indicate subjects who switched at least once for questions in each class (FOSD, EASY, and HARD). Table 2 reports these results as well as the average and the median number of questions for which each subject selected inconsistent answers, and the proportion of subjects choosing inconsistent answers at least *twice*, conditional on being inconsistent in that part of the experiment. It also includes summary statistics on the choice of the computerized coin in Part I.

In line with existing evidence, we find that the vast majority of subjects (90%) exhibit a stochastic choice behavior in Part I of the experiment. In our data such switching behavior is not only widespread across subject, but frequent for each subjects: not only most subjects 'switch', but they tend to do so multiple times. As evident from Table 2, the majority of subjects (60%) that exhibit stochastic choice reports inconsistent answers for at least half of the HARD questions (2 or more out of 4). That is, virtually every subject exhibits stochastic choice for HARD questions, and the majority do so more than once.

A strong additional pattern emerges: stochastic behavior is virtually only observed for HARD questions, while it is almost absent for EASY and FOSD ones. Indeed, only 5 out of 80 subjects (6.25%) ever choose the dominated option in FOSD questions, and only 2 out of 80 subjects (3%) change their answers in EASY questions. As we will discuss in detail in Section 3.4, the key determinant of stochasticity in our experiment appears to be the 'difficulty' of the questions and not the differences in expected values (or utilities) of presented options, as often suggested by the literature.

The key finding of our experiment is the behavior in Part III, where questions are repeated three times in a row and subjects are explicitly told about it. As evident from Figure 2 and Table 2, we find that a large majority of subjects (71%) reported inconsistent answers in this case as well. Just like in Part I, this takes place virtually only for HARD questions, while this almost never happens for FOSD and EASY ones. While indeed in this case the proportion is smaller than the one observed in Part I (71% vs. 90%), it still involves a large majority of the population. (In Section 3.3 we discuss the relation between switching behavior in different parts.) From Table 2 we also see that, like in Part I, also in Part III not only most subjects switch, but they tend to do it more than once: the majority (70%) of subjects who report inconsistent answers for HARD questions do so for at least 2 out of 4 of them. That is, also when they are aware of the fact that it is the same question, most subjects report inconsistent answers, and most of them do so multiple times. This is the main finding of our experiment.

Finally, we find that a much smaller, but still sizable fraction (29%) of the subjects

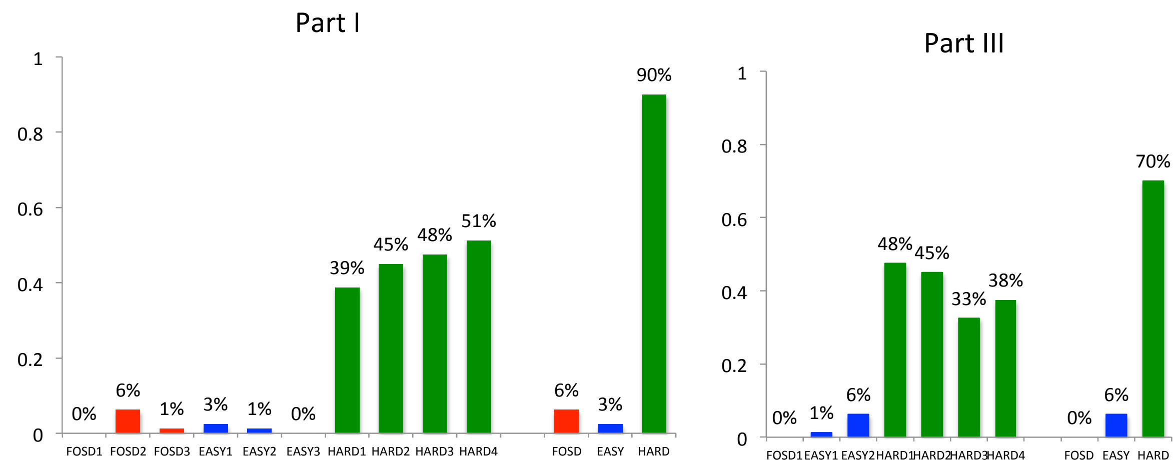

FIGURE 2 Fraction of Subjects who gave inconsistent answers, by question

choose to the option to flip a costly coin. Again this is done virtually only for HARD questions, and those who choose the coin tend to do so frequently (about 40% of those who chose it do it at least twice, as shown in Table 2).

We conclude this section by noting that we do not find any prevalent 'pattern of switching', either in Part I or III,[29] and that for each pair of HARD questions, there is no correlation in the switching patterns of subjects who reported stochastic answers in both questions.

Result 1.a. *The vast majority of subjects (90%) exhibit stochastic choice behavior in Part I of the experiment. This behavior is observed almost exclusively for HARD questions and very rare for FOSD and EASY questions. Most of the subjects that report stochastic answers do so for at least 2 of the 4 HARD questions.*

Result 1.b. *A large majority of subjects (71%) select different lotteries in the repetition of the same question in Part III of the experiment, in which they are aware of this repetition. Again, this behavior is observed almost exclusively for HARD questions, and again, most of the subjects that report stochastic answers, do so for at least 2 of the 4 HARD questions.*

Result 1.c. *A non-trivial fraction of subjects (29%) choose the computerized costly coin in Part I of the experiment, mostly in HARD questions.*

[29]By 'pattern of switching' we mean the sequence of subject's choices in the three repetitions of the same question. For instance, one pattern is choose lottery 1 in the first repetition; then lottery 2; and then lottery 1 again (pattern 121). We find no prevalent pattern.

Part I	% of subjects that switched at least once (# of subjects)	*Conditional on Switching* mean # of switches	median # of switches	% with 2 or more switches
FOSD	6 % (5)	1.2	1	–
EASY	3% (2)	1.5	1.5	–
HARD	90% (72)	2.03	2	60%
Total	90% (72)	2.15	2	61%

Part III	% of subjects that switched at least once (# of subjects)	*Conditional on Switching* mean # of switches	median # of switches	% with 2 or more switches
FOSD	0 % (0)	0	0	–
EASY	6% (5)	1.2	1	–
HARD	70% (56)	2.32	2	70%
Total	71% (57)	2.36	2	68%

Coin Flip	% of subjects that chose coin at least once (# of subjects)	*Conditional on Choosing the Coin* mean # of coin choices	median # of coin choices	% with 2 or more coin choice
FOSD	4 % (3)	1.33	1	–
EASY	8% (6)	1	1	–
HARD	24% (19)	1.63	1	42%
Total	29% (23)	1.78	1	43%

Table 2 Frequency of Switches and Coin Choices. The right columns report the frequency of switching behavior and coin choice conditional on a subject having switched (or chosen a coin) at least once in each category.

3.2 Behavior in the Robustness Test

We now turn to discuss the results of the robustness test. Recall that the key differences of this treatment with the main one are that subjects were only asked to choose from one question repeated three times (question HARD2), with much higher stakes.

We find the same pattern of behavior in the robustness test: 50% of the participating students (13 out of 26) chose different lotteries in the repetition of the same question. This is very similar to the 45% observed in Part III of the main treatment for the same question. This result suggests that our main findings are robust and hold also when stakes are very much higher, and subjects are asked only one repetition of questions without confounding effects from the rest of the experiment.

	Switch in Part I	
	All quest.	*HARD quest.*
Switch in III	0.98***	0.38***
	($p < 0.01$)	($p < 0.01$)
Constant	-0.91***	-0.27***
	($p < 0.01$)	($p < 0.01$)
# of obs.	560	320
# of clusters	80	80
Log Likel.	-294.93	-217.01

(a)

	Switch in I	Switch in III
Flip a Coin	0.69***	0.76***
	($p < 0.01$)	($p < 0.01$)
Constant	-0.91***	-0.76***
	($p < 0.01$)	($p < 0.01$)
# of obs.	800	560
# of clusters	80	80
Log Like.	-387.69	-304.09

(b)

TABLE 3 (a) Probability of switching in Part I of the experiment, clustering by individuals. (b) Probability of switching in Part I or III regressed on the dummy indicating that coin was selected, clustering by individuals.

3.3 Relation between Part I, Part III, and Coin Flip: main treatment

We now explore the relation between the switching behavior in Part I and III, and with the tendency to choose the costly coin. This will allow us to shed light on the behavior of the agent in the experiment as a whole.

Table 3a presents a PROBIT regression of switching in Part I on the (dichotomous) variable that indicates whether a subject changed her answer in the *same* question in Part III, clustering by individuals. We find a strong and significant correlation between switching behavior observed in Parts I and III. The regressions indicate that switching behavior in both parts of the experiment are strongly and positively correlated at the 1% level.[30] [31]

Table 3b presents the results on the relation between the choice to flip a coin and the behavior in the rest of the experiment. It shows two regressions, in which the flipping-a-coin option is used to explain the switching behavior observed in Parts I and III (again observations are clustered by individuals). The results clearly indicate that subjects that choose the option to flip a costly coin are also significantly more likely to exhibit a switching behavior in either part of the experiment.[32]

[30]Notice that we are not regressing the tendency to switch for at least some question in Part I, and at least some question in Part III – this would be rather complicated, also from an estimation prospective, since the vast majority of subjects exhibit both behaviors. Rather, we run this analysis by question (clustering by subject), where the average switch is almost always below 50%.

[31]The same conclusion holds when estimating the Spearman correlation between the two dichotomous variables that indicate whether a subject switched at least once in Part I and and Part III. This correlation coefficient is positive (corr = 0.25) and significant at 5% level ($p = 0.0262$).

[32]The Spearman correlation coefficients between the tendency of a subject to flip a coin and switch in Part I or in Part III of the experiment are also both positive and significant: we find a correlation of 0.21 with $p = 0.0594$ for Part I, and of 0.22 with $p = 0.0494$ for Part III.

We summarize these findings in the following result.

Result 2. *The Switching behavior in Parts I, the switching behavior in Part III, and the choice to flip a coin are strongly and significantly correlated.*

3.4 Stochastic Choice and Expected Values

A very robust finding in the experimental literature on Stochastic Choice is that switching behavior tends to be more frequent the closer the Expected Values (or Expected Utilities) of the two involved lotteries. On the other hand, we have already seen that in our experiment stochastic choice behavior is observed only in HARD questions and is almost entirely absent in EASY and FOSD ones. These two patterns together suggests two possible fundamental determinants for stochastic choice. On the one hand, following the typical suggestion the literature, the key one could be the difference in Expected Value (or EU) – and in our experiment HARD questions are those that options with a smaller difference in expected values (or utilities). On the other hand, opposite to the typical suggestion in the literature, the key determinant could be the actual *difficulty* of the question – which is related but distinct from the similarity of the EV or EU. We will now use our data to try to distinguish the two, by estimating whether differences in Expected Values or Utilities affected switching behavior. The distinction is made possible because in our experiment there are questions with similar Expected Values but different 'difficulty,' and questions with similar 'difficulty' but different Expected Values.

Table 4 summarizes the results of a regression analysis where the dummy variable of whether a subject switched in Part I or in Part III is regressed on the dichotomous variable that indicates the difficulty of the question, the absolute difference in Expected Values of the two lotteries, the absolute difference in variances of the lotteries, and the absolute difference in the Expected Utilities of the lotteries.[33] All presented regressions are Random Effects PROBIT regressions, where the standard errors are clustered by the individual.[34]

Regressions 1 and 3 confirm the pattern we have already noted: subjects are much more likely to exhibit switching behavior in HARD than in EASY or FOSD questions in both Part I and in Part III of the experiment.

[33] To calculate the latter we use the risk question from Part II of the experiment to estimate subject-specific CRRA parameter, which allows us to evaluate the expected utility of lotteries for each subject. This estimation is possible for all subjects who reported strictly positive number of points that they want to invest in the risky project, which is true for 78 out of 80 subjects. Thus, our full sample in these regressions is 78 subjects. (Of these, 30 subjects that invested the whole budget; they are treated as the risk-neutral ones with linear utility.)

[34] Regressions for switching in Part III focus on EASY and HARD questions and omit the only FOSD question we have in this part of the experiment (FOSD1) since all subjects reported consistent answers in all three repetitions.

	Dependendent Variable: Switch in Part I						
	Reg. 1	Reg. 2	Reg. 3	Reg. 4	Reg. 5	Reg. 6	Reg. 7
Dummy EASY	-0.36 (0.37)	0.16 (0.53)	-0.29 (0.39)				
Dummy HARD	2.20** (0.24)	2.03** (0.27)	2.14** (0.25)	1.99** (0.53)	2.51** (0.37)		
Diff in EV		- 0.02 (0.02)		-0.02 (0.02)		-0.07 (0.04)	
Diff in Var		0.00003 (0.00007)		0.00002 (0.00007)		0.00001 (0.00008)	
Diff in EU			0.009 (0.008)		-0.003 (0.01)		0.01 (0.01)
Const	-2.33** (0.24)	-2.10** (0.25)	-2.22** (0.25)	-2.06** (0.54)	- 2.62** (0.40)	0.04 (0.15)	-0.18 (0.11)
# of obs	780	780	780	546	546	312	312
# of subjects	78	78	78	78	78	78	78
Sample Quest.	ALL	ALL	ALL	EASY, HARD	EASY, HARD	HARD	HARD

	Dependendent Variable: Switch in Part III						
	Reg. 1	Reg. 2	Reg. 3	Reg. 4	Reg. 5	Reg. 6	Reg. 7
Dummy HARD	2.16** (0.28)			3.61** (0.79)	2.36** (0.35)		
Diff in EV				0.04** (0.02)		0.10** (0.04)	
Diff in Var				-0.0002** (0.00007)		-0.0003** (0.0001)	
Diff in EU					0.01 (0.01)		-0.02 (0.02)
Const	-2.47** (0.29)			-3.88** (0.82)	-2.73** (0.41)	-0.41** (0.21)	-0.25 (0.17)
# of obs	468			468	468	312	312
# of subjects	78			78	78	78	78
Sample Quest.	EASY, HARD			EASY, HARD	EASY, HARD	HARD	HARD

TABLE 4 Errors are clustered at the individual level; ** indicates significance at 5% level; All the differences (in expected values, expected utilities and variances) are absolute differences; Expected utilities of the lotteries are computed based on the estimation of CRRA utility for each subject given her reported risky investment in Part II of the experiment; Regressions 1 and 2 are meaningless for the switching in Part III since there was only one FOSD question in Part III.

The novel part are Regressions 2, 4 and 6: they clearly show that this switching behavior is *not* explained by the similarity in Expected Values of the lotteries, or Variances. In Part I, neither of the differences is statistically significant, while in Part III the estimated coefficients on the differences in expected values have the *opposite* sign – higher difference in expected value leads to more switching behavior. This results hold if we focus on all questions, on both EASY and HARD, or only on HARD. The differences in the variances of lotteries instead does affect switching behavior in Part III, but the effect is rather small.

Regressions 3, 5 and 7 show a similar pattern for differences in the Expected Utilities – it is not correlated with switching behavior in both Part I and in Part III of the experiment. Again, this holds irrespectively of the subset of questions included in the regressions.

To summarize, our regression analysis reveals that the stochastic choice behavior observed in both Part I and Part III of our experiment is *not* related to the differences in Expected Utilities or Expected Values of the lotteries. Instead, it appears to be related to some other feature of the questions, such as their distinction into the EASY or HARD category – which may possibly capture their actual difficulty.

Result 5. *Stochastic Choice behavior in either part of the experiment in not related to the differences in Expected Value or Utility of the two available options. Instead, it it is related to whether the question is classified as HARD, EASY, or FOSD.*

We conclude by emphasizing that we do not consider this finding incompatible with the established result that stochastic choice is linked to the similarity of options in terms of Expected utilities or Values. This apparent inconsistency can have a simple explanation: existing experiments have almost only focused on the differences in Expected Value/Utility; but because these tend to be smaller for 'hard' questions, it is difficult to disentangle which is the key factor using their data. In particular, even if stochastic choice is related only the difficulty of the question, as suggested by our result, because in their design this is in turn related to the difference in Expected Values, then the latter is naturally found to be linked with stochastic choice.

3.5 Answer to the Questionnaires

An informal indication of the reason why subjects choose different options in Part III comes from their answers to the final questionnaire, in which subjects were asked if they reported different answers in Part III and, if so, why did they do so. We emphasize that the answers to this questionnaire were not incentivized and took place at the very end of the experiment, so their content should be treated with the appropriate caution.[35]

Of the (majority) of subjects who reported choosing different answers in the repetitions of questions in Part III, a large majority writes a motivation which could be summarized as: "I didn't know which option was best, and therefore I've chosen not to commit to one, picking sometimes one option, and sometimes the other." (Since these were open questions, the precise wording varied greatly between subjects, but the meaning was rather stable for a large majority of subjects.) This indicates two important points. First, subjects appear to be *aware* of their choice of different answers – which can be seen as an actual choice itself. Second, stochastic choice appears to be the reaction to some form of incompleteness of the preferences, or incomparability of the available options. As we shall discuss in Section 4.3, this is in line with some existing models of Deliberate Randomization.

A similar questionnaire (adapted for the specifics of that experiment) was asked at the end of the robustness test as well. Very similar answers were given in that case as well: again subjects appeared to be unsure of which of the options to choose, and to purposely select sometimes one lottery and sometimes the other.

3.6 Response Time

We now turn to analyze the time that subjects took to respond to each question in the various parts of the experiment, which could be taken as a proxy for how long they took to think about each decision problem they face.

We begin by noting that we find no statistical difference in the distribution of reaction times between participants that reported stochastic answers and those that were consistent in all three repetitions of the same question.[36] This finding refutes the hypothesis that subjects that reported stochastic choice in Part III did so because they did not spend time thinking about their choice; in fact, the time they spent thinking was not distinguishable from that of the subjects who were consistent.

Table 5 contains some summary statistics on reaction-time data for each question. Three patterns emerge. First, as one would expect, decision times increase as we move from FOSD, to EASY, to HARD. Second, response time is weakly decreasing as subjects encounter the same question over and over again – potentially indicating the effect of experience or memory of solving the question in the past. Both patterns are standard findings.

[35] For example, one might worry that, once we ask subjects the question above, they might realize that they gave inconsistent choices, and write in the questionnaire an answer meant to ex-post rationalize their behavior. While we find this implausible, of course we have no way of knowing.

[36] To test this we performed a Wilcoxon RankSum test between reaction times of subjects that switched and those that did not, separately in each repetition of the questions that appeared in Part III of the experiment. None of these are significant even at the 10%.

Mean (Median) Response Times								
	FOSD1		FOSD2		FOSD3			
Repetition	Part I	Part III	Part I	Part III	Part I	Part III		
First	9 (8)	3 (3)	6 (5)		8 (7)			
Second	6 (4)	2 (1)	4 (3)		5 (4)			
Third	3 (3)	2 (1)	3 (3)		5 (4)			
Fourth	4 (3)		3 (3)		4 (3)			
	EASY1		EASY2		EASY3			
Repetition	Part I	Part III	Part I	Part III	Part I	Part III		
First	7 (6)	4 (3)	7 (6)	4 (3)	5 (5)			
Second	4 (4)	2 (2)	4 (4)	2 (1)	4 (3)			
Third	4 (3)	2 (2)	4 (3)	2 (1)	4 (3)			
Fourth	5 (3)		3 (3)		3 (3)			
	HARD1		HARD2		HARD3		HARD4	
Repetition	Part I	Part III	Part I	Part III	Part I	Part III	Part I	Part III
First	13 (9)	10 (6)	12 (9)	9 (5)	14 (10)	12 (8)	18 (12)	9 (7)
Second	8 (6)	3 (2)	9 (6)	2 (2)	12 (7)	3 (2)	10 (8)	3 (2)
Third	8 (5)	2 (2)	8 (6)	2 (2)	8 (7)	2 (2)	8 (7)	3 (2)
Fourth	7 (6)		6 (5)		8 (6)		10 (8)	

Table 5 Mean (Median) Response Times, in seconds, per question in Part I and III. Response Times are rounded to seconds by the software.

The third pattern that emerges is that the response times for the three repetitions of HARD questions in Part I are very different from those of the same question in Part III. In Part I, subjects take a relatively long time to think about each of the three repetitions. By contrast, in Part III they think long about the question only the very first time they encounter it – a median of 6 to 8 seconds – while the next two repetitions are answered almost immediately – a median of 2 seconds.[37] To better interpret it we can take as a benchmark the response times in the second and third repetitions of FOSD questions also in Part III. These are naturally the easiest questions in our experiments, where we can expect subjects to recognize the dominated option in the first repetition and just implement the same choice in later ones (and indeed we observe that all subjects choose the dominant option all the times). Thus, the response time in this case can be seen as a proxy of the time

[37]This difference is confirmed also by statistical analysis. Wilcoxon Signrank tests reject the hypothesis that the response times in the first repetition in Parts I and III are the identical; the same holds for the second and third repetitions ($p < 0.01$ for all three tests). Also rejected is the hypothesis that the first repetition in Part III is identical to the second and third repetitions ($p < 0.01$ for both). Finally, the same tests accept the Hypothesis that the second and third repetitions in Part III have the same distribution of response times. As noted above, virtually identical results hold if we focus on subjects who switched answer for that particular question.

necessary to move the mouse across the screen to make a choice.[38] The key observation is that the response times in the 2nd and 3rd repetitions of HARD questions in Part III are very similar, and statistically indistinguishable, from this benchmark. They are thus akin to that of questions in which the agent has an 'obvious' choice to simply implement. Overall in our experiment, for any type of question, subject appear to have very similar reaction times in the second and third repetitions in Part III. By contrast, in the first repetition of HARD questions in Part III subjects take considerably longer, and their answer is statistically indistinguishable from the second and third repetition of the question in Part I.

Thus, to summarize, for HARD questions in Part III subjects take a relatively long time to answer the first repetition – similar to that of Part I – and then a much shorter time to answer the following two repetitions – a time akin to that for the repetitions of FOSD questions. These results confirm the intuition that subjects do not treat the repetitions of the same question in Part III as whole new question: rather, aware of the fact that it is the same question asked multiple time, their behavior is consistent with simply implementing a plan made the first time they observe the question in this part. As we have seen, the interesting thing is that , while for FOSD or EASY questions subjects simply choose the same lottery in each repetition, in HARD questions they often choose different ones.

We conclude by noting that there is no relationship between decision time in Part III of the experiment and the tendency to report a stochastic answer in either part of the experiment.

3.7 Relation with Risk Attitudes and Violations of EU

Finally, we connect the tendency to report stochastic answers with the attitudes towards risk and compound lotteries, as measured in Part II, and with their answers to the Allais questions, as measured in Part IV.

Regarding risk attitudes, consistent with previous studies we find that the majority of subjects (63%) are risk-averse.[39]

[38]Notice that in our experiment response times are approximated to second (and not milliseconds). Moreover, in our experiment the initial position of the cursor in each screen was the bottom left corner, which means that subjects had to move the mouse to select the desired alternative and then move it back to click the "Submit" button. Thus a response time of 2 seconds is consistent with the time necessary to move the mouse. (This is in contrast with other experiments, especially in psychology, where the initial position of the cursor is on top of the options and thus a choice can be made in less than a second.)

[39]These are the subjects that choose to invest less than 100 tokens in the risky asset (which pays 2.5 times the investment with probability 50%). The remaining subjects invested all available budget in the risky asset, which indicates that they are either risk-neutral or risk-loving. The average investment in the risky asset was 69.6, which matches typical results in the literature. See,

	Switch in Part I		Switch in Part III		Flip a Coin	
Risky Investment	-0.12	($p > 0.1$)	-0.09	($p > 0.1$)	-0.04	($p > 0.1$)
Risk-averse (Indicator)	**0.26****	($p = 0.02$)	0.14	($p > 0.1$)	0.15	($p > 0.1$)
Compound Investment	-0.04	($p > 0.1$)	-0.03	($p > 0.1$)	-0.06	($p > 0.1$)
Compound (Indicator)	0.12	($p > 0.1$)	0.10	($p > 0.1$)	-0.04	($p > 0.1$)
Allais-like behavior	**0.19***	($p = 0.09$)	**0.23****	($p = 0.04$)	0.16	($p > 0.1$)

Table 6 Spearman correlations between stochastic choice and preferences under risk

Comparing the investment in the risky asset between risk and compound question (first and the second question in Part II) allows us to estimate subjects' attitude towards compound lotteries. Only 3 out of 80 subjects (4%) invested more in the compound lottery question, which suggests that these subjects value compound lotteries more than the corresponding simple ones. As expected, the vast majority of subjects (the remaining 96%) do not derive additional utility from compound lotteries. Among them, 43% strictly dislike compound lotteries (they invest strictly less in compound lottery than in the risk question), and 54% appear to be neutral (they invest the same amount of tokens in both questions).[40]

In Part IV we ask subjects typical questions linked to the paradox of Allais (1953). We find that 11 subjects exhibit the *Common Consequences* effect, 7 subjects exhibit the *Common Ratio* effect and 1 subject exhibits both. In other words, 20 subjects in total (25%) violate expected utility principles according to the most widely used tests in the literature. This rather small proportion is in line with what reported by other experiments with similar small incentives incentives (e.g., Camerer 1989b, Conlisk 1989, Burke et al. 1996, Fan 2002, Huck and Müller 2012).

Table 6 reports the Spearman correlations between the tendency to give inconsistent answers in Parts I and III and the measures above, where the attitudes towards risk and compound lotteries are analyzed both using the amount of tokens invested in each case, and the indicator variables on whether the subject is risk or compound averse.

We find that in general the attitudes of subjects toward risk and compound lotteries are not related to switching behavior observed in Parts I and III of our experiment, as well as to the tendency to choose a computer-simulated coin. The only exception is the risk-averse indicator, which is positively and significantly correlated with the switching behavior in Part I of the experiment.[41] At the same time, there is positive

for example, Dreber and Hoffman (2007), Charness and Gneezy (2010), Langer and Weber (2008).

[40]Wilcoxon matched-pairs signed-ranks test detects a statistical difference between investment in the risk and compound questions at 1% level ($z = 4.963$ and p-value < 0.01).

[41]We should emphasize, however, that contrary to the standard methods used in the literature (e.g. Halevy 2007), we measure attitude towards compound lotteries by describing the probability

and significant correlation between the tendency to exhibit Allais-like violations of Expected Utility and reporting inconsistent answers in Parts I and III of the experiment.[42]

4. Implications for Models of Stochastic Choice

In this last substantive section we discuss the implications of our results for existing models of stochastic choice. We focus in details on models of Random Expected Utility, Mistakes, and Deliberate Randomization.

Before proceeding, we briefly mention two other possible interpretations: indifference and preference discovery. According to the former, what we consider stochastic choice is simply due to the fact that subjects are indifferent, or almost indifferent, between the available options and 'break' such indifference randomly. Indeed this is compatible with stochastic choice for any question except FOSD and the choice of the coin (which is costly). However, such indifference, and thus stochastic choice, should be more common for those questions in which options have a more similar Expected Value or Utility – but this is in contrast with what we find (see Section 3.4). Alternatively, stochasticity of choice could be due to 'preference discovery:' subjects could be *forming* their preferences on the course of the experiment, and the stochasticity is due to this process. This interpretation predicts that, since in our experiment most questions are asked many times, the stochasticity of choice should disappear, or be significantly smaller, in Part III of the experiment. As we discussed above, this is not what we observe.

4.1 Random Expected Utility

One of the most common interpretations of stochastic choice is following the (class of) models of *Random Utility.* According to them, every time subjects make a decision they maximize a well-defined utility function, but this utility changes over time, stochastically, possibly even within one experimental session. That is, randomness in choice is linked to randomness in utility, which is modeled by replacing the utility function with a probability distribution over possible utilities. Models of this kind appear in Thurstone (1927), Harsanyi (1973), Falmagne (1978), Cohen (1980), Barberá and Pattanaik (1986), McFadden and Richter (1991), Loomes and Sugden (1995),

of a successful investment as the compound lottery, rather than have a lottery with two (or more) steps of resolution. It is possible that this negative result is due to this design choice, and we leave for the future research a more thorough investigation of the connection between stochastic choice and attitudes toward compound lotteries.

[42] The raw correlation is corr = 0.19 ($p = 0.09$) for Part I and corr = 0.23 ($p = 0.04$) for Part III.

Clark (1996), McFadden (2006), Gul and Pesendorfer (2006), Ahn and Sarver (2013), and Fudenberg and Strzalecki (2014). A relevant special case for us is that of choice between risky prospects, called Random Expected Utility, in which the agent follows Expected Utility using a stochastically-determined Bernoulli utility function.

The implications of the Random Expected Utility model for Part I of our experiment are clear: subjects can change their answers for EASY or HARD questions, but not for FOSD ones (the dominant option should always be chosen for any utility). We should also observe more frequent 'switching' for HARD rather than for EASY questions, as for the latter one option is better for a much larger set of utilities. Finally, subjects should never opt to flip a costly coin since according to this interpretation once the subject faces a question, she knows which utility to use and behaves like an Expected Utility maximizer.

If we compare these predictions for Part I to our data, we find that, as predicted, stochastic choice behavior is more frequent for hard question; contrary to the model, we do observe sizable fraction of subjects choosing to flip the costly coin.

With regards to Part III of the experiment, where subjects are explicitly told that they will observe the same question three times in a row, the predictions depend on the interpretation of the model. If we are willing to assume that, even though subjects are aware that it's the same question, they treat each repetition as a whole new decision, and thus we allow their utility to change; and if we further assume that the utility of the agents changes with such high frequency between two adjacent questions; then the model is compatible with switching behavior in Part III.

However, we find these assumptions implausible: in particular, we believe it wouldn't make sense for a subject to treat each instance of the same question as a new decision problem: she knows it's the *same* question for three times in a row – she is told in the instructions and she has seen it for all other questions Part III – and we find it unrealistic that she would just disregard this and treat each repetition as new question. As we have seen in Section 3.6, this view finds support in the analysis of response times, where repetitions of questions in Part III are treated very differently from those in Part I: subjects seem not to spend time to 'rethink' the problem. In particular, the response time for repetitions of HARD questions in Part III is indistinguishable from that of repetitions on FOSD questions in Part III, where subject do not have to think about what to choose but simply 'implement a plan.'[43]

Under the more realistic view that subjects do not treat each instance of the same question in Part III as a new decision problem, the Random Expected Utility

[43]One may believe that the difference in behavior in the receptions in Part I and III of HARD questions is due to the fact that subjects do not need to read the question again – as they know they have just read it. If they are choosing again, however, they would still need time to think about what they want. But then, response time should be different from FOSD questions, where this is not needed. As we have seen, this is not what we find.

model would then predict no 'switching' behavior in Part III, as the utility of the agent should not change. As we have seen, this is contradicted by our data: a large majority of our subjects exhibit stochastic choice in Part III, and do so consistently (it is often exhibited for the majority of HARD questions). Thus the Random Expected Utility model, under the interpretation above, cannot explain the behavior of a large majority of subjects in our experiment.[44]

We should also emphasize that this interpretation is also inconsistent with subjects' answers to the questionnaire, where they state that the choice of different options in Part III was a deliberate decision due to a difficulty to choose.[45]

4.2 Mistakes

A second common interpretation of stochastic choice is that of *Mistakes*, which includes a large number of models that assume that subjects have a well-defined utility function, but by mistake may fail to choose the option that maximizes this utility. These mistakes can be of many forms, but the most common assumptions are that every option has a positive probability of being chosen; and that the frequency of the mistakes, and hence of stochastic choice, should depend on the difference in utility between the available options, and mistakes should be more frequent for questions where the options have similar utilities.[46] This is one possible interpretation of the well-known model of Luce (1959), and models of this kind appear, amongst many, in Harless and Camerer (1994), Hey and Orme (1994), Camerer and Ho (1994), Wu and Gonzalez (1996), and Fudenberg and Strzalecki (2014). Indeed under this interpretation, subjects may exhibit violations of First Order Stochastic Dominance (FOSD): if one option is dominated by another, but their utilities are nevertheless very similar, the dominated option should be chosen a significant fraction of times.

Testing this interpretation in our data gives a very poor fit: as discussed in Section 3.4, the difference in Expected Value or Expected Utility between options does not explain the stochastic choice behavior in our data – both for all questions, and focusing only on EASY and HARD, or only on HARD ones.

In terms of the the predictions specific for Part III, again they depend on our interpretation of the model. If, like for Random Utility, we assume that subjects do *not* treat each repetition as an wholly different question and compute everything

[44] We note that the behavior in our experiment can be compatible with a Random Utility model in which subjects have non-Expected Utility preferences, since specific forms of non-Expected Utility can induce a desire to report stochastic answers. In this case, such models would be part of the class of models of Deliberate Randomization discussed in Section 4.3. (We refer to that section for more discussion on how non-EU preferences can lead to a desire to randomize.)

[45] Though this should be treated with caution as it can be an ex-post rationalization.

[46] Exceptions are Wilcox (2011) and Blavatskyy (2014).

again (an assumption that, we have argued above, find support in agent's response time), then the only mistakes we should observe are of the 'trembling-hand' kind. But this appears to be incompatible with our data, for two reasons. First, such mistakes should be independent of the type of question asked, while in our data they are concentrated only in HARD questions. Second, we find it hard to believe that a majority of subjects exhibit such 'trembling-hand' mistakes for the majority of HARD questions (two or more out of four), but only for them. Thus, under the assumption above the Mistakes interpretation does not appear to be compatible with the behavior of the majority of the subjects in our experiment, both in Part I and III.

As for Random Utility, we conclude by emphasizing that the Mistakes model is also inconsistent with subjects' answers to the questionnaire where they state that the choice of different options in Part III was a deliberate decision due to a difficulty to choose – even though these answers should be treated with caution.

4.3 Deliberate Randomization

A third class of models of stochastic choice postulate that the stochasticity is a *deliberate choice* of the agent. This could take place because subjects do not have standard complete Expected Utility preferences, and would rather choose stochastic answers, randomizing between the two alternatives, than receive either of them.

Some of these models, such as Dwenger et al. (2013), link this behavior with the desire to minimize *regret*: if the agent feels more regret for ex-post suboptimal choices when they choose one option rather than it being chosen by chance, then they will prefer to randomize their answers.

Other models link this behavior to preferences that violate Expected Utility, in particular those studied in relation with the Allais paradox. This avenue was originally suggested by Machina (1985) who noted that convexity of preferences, which could generate Allais-like behavior, would also generate a desire to 'mix' between some of the available options. One example is the model in Cerreia-Vioglio et al. (2014b). It has two key components. First, when subjects are asked to choose between two lotteries, they consider *all* possible probability distributions over them and choose the optimal one, where they evaluate such randomizations by reducing them to standard lotteries over final outcomes.[47] The second component is that subjects have non-Expected Utility, and convex, preferences over lotteries. In particular, they follow the model of *Cautious Expected Utility*, developed in Cerreia-Vioglio et al. (2014a), in which subjects have not one, but many utility functions over money with

[47]That is, when asked to choose between p and q, they choose the probability of choosing p, denoted by λ, that maximizes their valuation of the monetary lottery $\lambda p + (1-\lambda)q$, which is the lottery that returns a monetary amount x with probability $\lambda p(x) + (1-\lambda)q(x)$. Thus they are deliberately randomizing, and choosing the 'optimal' randomization.

different curvatures, as if they didn't know how risk averse they should be, and evaluate lotteries by displaying a *cautious behavior*: to evaluate a lottery p, they compute its certainty equivalent with respect to each possible function in the set, and pick the smallest one.[48] These preferences lead the agent to *choose* to report a stochastic answer: as they don't know which utility function to use, they may prefer to 'hedge' between the available options, and thus may strictly prefer $\frac{1}{2}p+\frac{1}{2}q$ to either p or q. As discussed in Cerreia-Vioglio et al. (2014a), this could in turn be related to incomplete preferences. Other models of this kind appear in Marley (1997), Fudenberg et al. (2014), Fudenberg and Strzalecki (2014), Swait and Marley (2013).

The predictions of these models in our experiment are clear. In Part I we should observe no stochastic answer for FOSD questions: these models have forms of monotonicity for which the agent should never chooses a dominated option. We should also expect (much) more frequent changes for HARD rather than for EASY questions, as for HARD questions non-EU preference have a stronger pull to choose to report stochastic answers. As opposed to the models in the previous sections, for these models the same predictions apply to Part III of our experiment: it is especially in this case that we should observe different answers to the same questions. Finally, these models make no predictions on the choice to flip a costly coin (unless they involve FOSD). On the one hand, they allow for it, as it provides another form of randomization. On the other hand, they are also compatible with the agent not choosing it because: 1) the coin is costly (the preference for randomizing may not be strong enough); 2) instead of choosing the costly coin in the software, the agent could instead choose to flip 'a coin in her head,' or a physical coin in her hands, for free;[49] 3) the coin presents a very specific randomization (50/50), that might be suboptimal for the agent (her optimal hedging could be, say, 95/5).

It is easy to see how these predictions find a strong support in our data: subjects exhibit a stochastic choice behavior in Part I and III, and a sizable fraction of subjects chose the option to flip a coin. Importantly, moreover, the answers to the questionnaire at the end of the experiment support interpretations along these lines: subjects report of choosing different answers as they didn't know which option was best and *preferred* to choose different options.

Models of Deliberate Randomization also predict that the stochasticity of answers in Part I and III of the experiment should be correlated, per subject and per questions. Again, this is precisely what we observe in our data.

Finally, the models that link stochastic choice to Non-Expected Utility preferences

[48]That is, if $\mathcal{U}$ denotes the set of utility functions that the agent use, they evaluate each lottery following

$$V(p) := \min_{v\in\mathcal{U}} \; v^{-1}(\mathbb{E}_p(v)).$$

See Cerreia-Vioglio et al. (2014a) for more.

[49]This was commonly given in the final questionnaire as a reason for not choosing the coin.

also predict a connection with Allais-like behavior.[50] As we have seen in Section 3.7, we document this relationship in our experiment.

5. Conclusion

In this paper, we propose an experimental design that allows us to distinguish between three broad classes of theories suggested in the literature to account for stochastic choice: Random Expected Utility; Mistakes; and Deliberate Randomization. Our results, both in the main experiment and in the robustness test, support the class of models of Deliberate Randomization.

We conclude by noting that our results investigate only the case of stochastic choice of objective lotteries. Whether similar results would hold in different domains, like the choice of food items, or ambiguous bets, is a natural question. We believe our methodology could be easily adapted to study these cases as well.

References

Ahn, D. S. and T. Sarver (2013): "Preference for flexibility and random choice," *Econometrica*, 81, 341–361.

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

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

Barberá, S. and P. K. Pattanaik (1986): "Falmagne and the rationalizability of stochastic choices in terms of random orderings," *Econometrica*, 707–715.

Blavatskyy, P. R. (2014): "Stronger utility," *Theory and Decision*, 76, 265–286.

Burke, M. S., J. R. Carter, R. D. Gominiak, and D. F. Ohl (1996): "An experimental note on the allais paradox and monetary incentives," *Empirical Economics*, 21, 617–632.

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

——— (1989b): "An experimental test of several generalized utility theories," *Journal of Risk and uncertainty*, 2, 61–104.

[50]To be more precise, these models connect the choice to report stochastic answers to Non-Expected Utility preferences, which could also lead to Allais behavior. At the same time, for most of these models (including the one of Cerreia-Vioglio et al. 2014b), the agent may exhibit the Allais paradox for some range of prizes but not for others. Thus, these models are compatible with subjects exhibiting stochastic choice but no Allais behavior in different ranges of prizes, as is the case with the majority of our subjects (see Cerreia-Vioglio et al. 2014b and Cerreia-Vioglio et al. 2014a).

——— (1995): "Individual decision making," in *Handbook of Experimental Economics*, ed. by A. R. J. Kagel, Princeton University Press, Princeton, vol. 2.

Camerer, C. F. and T.-H. Ho (1994): "Violations of the betweenness axiom and nonlinearity in probability," *Journal of risk and uncertainty*, 8, 167–196.

Cerreia-Vioglio, S., D. Dillenberger, and P. Ortoleva (2014a): "Cautious Expected Utility and the Certainty Effect," *Econometrica (forthcoming).*

Cerreia-Vioglio, S., D. Dillenberger, P. Ortoleva, and G. Riella (2014b): "Deliberately Stochastic: Random Choice and Preferences for Hedging," Mimeo Columbia University.

Charness, G. and U. Gneezy (2010): "Portfolio choice and risk attitudes: An experiment," *Economic Inquiry*, 48, 133–146.

Chen, Y.-J. and J. E. Corter (2006): "When mixed options are preferred in multiple-trial decisions," *Journal of Behavioral Decision Making*, 19, 17–42.

Clark, S. A. (1996): "The random utility model with an infinite choice space," *Economic Theory*, 7, 179–189.

Cohen, M. A. (1980): "Random utility systems—the infinite case," *Journal of Mathematical Psychology*, 22, 1–23.

Conlisk, J. (1989): "Three variants on the Allais example," *The American Economic Review*, 79, 392–407.

Dreber, A. and M. Hoffman (2007): "2D: 4D and Risk Aversion: Evidence that the Gender Gap in Preferences is Partly Biological," Mimeo, Stockholm School of Economics.

Dwenger, N., D. Kübler, and G. Weizsäcker (2013): "Flipping a Coin: Theory and Evidence," Mimeo Humboldt-Universität zu Berlin.

Eliaz, K. and G. Fréchette (2008): "Don't Put All Your Eggs in One Basket!: An Experimental Study of False Diversification," Mimeo New York University.

Falmagne, J. C. (1978): "A Representation Theorem for Random Finite Scale Systems," *Journal of Mathematical Psychology*, 18, 52–72.

Fan, C.-P. (2002): "Allais paradox in the small," *Journal of Economic Behavior & Organization*, 49, 411–421.

Fudenberg, D., R. Iijima, and T. Strzalecki (2014): "Stochastic Choice and Revealed Perturbed Utility," Mimeo Harvard University.

Fudenberg, D. and T. Strzalecki (2014): "Recursive Stochastic Choice," Mimeo, Harvard University.

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

Gul, F. and W. Pesendorfer (2006): "Random Expected Utility," *Econometrica*, 74, 121–146.

Halevy, Y. (2007): "Ellsberg Revisited: An Experimental Study," *Econometrica*, 75, 503–536.

HARLESS, D. W. AND C. F. CAMERER (1994): "The predictive utility of generalized expected utility theories," *Econometrica*, 1251–1289.

HARSANYI, J. C. (1973): "Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points," *International Journal of Game Theory*, 2, 1–23.

HEY, J. D. (2001): "Does repetition improve consistency?" *Experimental economics*, 4, 5–54.

HEY, J. D. AND E. CARBONE (1995): "Stochastic choice with deterministic preferences: An experimental investigation," *Economics Letters*, 47, 161–167.

HEY, J. D. AND C. ORME (1994): "Investigating generalizations of expected utility theory using experimental data," *Econometrica*, 1291–1326.

HUCK, S. AND W. MÜLLER (2012): "Allais for all: Revisiting the paradox in a large representative sample," *Journal of Risk and Uncertainty*, 44, 261–293.

KIRCHER, P., S. LUDWIG, AND A. SANDRONI (2013): "On the Difference between Social and Private Goods," *The BE Journal of Theoretical Economics*, 13.

LANGER, T. AND M. WEBER (2008): "Does commitment or feedback influence myopic loss aversion?: An experimental analysis," *Journal of Economic Behavior & Organization*, 67, 810–819.

LOOMES, G. AND R. SUGDEN (1982): "Regret theory: An alternative theory of rational choice under uncertainty," *The Economic Journal*, 92, 805–824.

——— (1995): "Incorporating a stochastic element into decision theories," *European Economic Review*, 39, 641–648.

LUCE, R. D. (1959): *Individual choice behavior: A theoretical analysis*, Wiley (New York).

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

MARLEY, A. (1997): "Probabilistic choice as a consequence of nonlinear (sub) optimization," *Journal of Mathematical Psychology*, 41, 382–391.

MCFADDEN, D. AND K. RICHTER (1991): "Stochastic Rationality and Revealed Stochastic Preference," in *Preferences, Uncertainty, and Rationality*, ed. by J. Chipman, D. McFadden, and K. Richter, Westview Press, 161–186.

MCFADDEN, D. L. (2006): "Revealed stochastic preference: a synthesis," in *Rationality and Equilibrium*, ed. by C. Aliprantis, R. Matzkin, D. McFadden, J. Moore, and N. Yannelis, Springer Berlin Heidelberg, vol. 26 of *Studies in Economic Theory*, 1–20.

REGENWETTER, M., J. DANA, AND C. P. DAVIS-STOBER (2011): "Transitivity of preferences," *Psychol Rev*, 118, 42–56.

REGENWETTER, M. AND C. P. DAVIS-STOBER (2012): "Behavioral variability of choices versus structural inconsistency of preferences," *Psychol Rev*, 119, 408–16.

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

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

Starmer, C. and R. Sugden (1989): "Probability and juxtaposition effects: An experimental investigation of the common ratio effect," *Journal of Risk and Uncertainty*, 2, 159–178.

Swait, J. and A. Marley (2013): "Probabilistic choice (models) as a result of balancing multiple goals," *Journal of Mathematical Psychology.*

Thurstone, L. L. (1927): "A law of comparative judgment." *Psychological review*, 34, 273.

Tversky, A. (1969): "Intransitivity of preferences." *Psychological review*, 76, 31.

Wilcox, N. T. (2011): "'Stochastically more risk averse:' A contextual theory of stochastic discrete choice under risk," *Journal of Econometrics*, 162, 89–104.

Wu, G. and R. Gonzalez (1996): "Curvature of the probability weighting function," *Management science*, 42, 1676–1690.

Appendix: For Online Publication Only

Appendix A: Analysis of the Response Time

In this section we analyze the response time of subjects to better understand stochastic choice behavior observed in Parts I and III of the experiment. Our software recorded how much time subjects took to make their choices in each round of the experiment (in seconds). Given natural heterogeneity between subjects, we will compare the response times exhibited by the same subject for different repetitions of the same question.

Recall that throughout the experiment subjects saw the exact same question repeated 6 times: three repetitions in Part I, with each repetition separated by other questions, and three repetitions in Part III, in which all three repetitions were back-to-back. Considering all six repetitions of the same question throughout the experiment, we find that the response time is weakly decreasing as subjects encounter the same question over and over again. This might partially indicate the effect of experience or memory of solving the question in the past. The longest response times are observed for the very first time the question is presented irrespectively of the difficulty of the question. As an example, Table 7 illustrates this trend by presenting the summary statistics of the response times in each repetition of the HARD1 question. (Notice that this trend of weakly decreasing response times does not change if one considers subjects that switched in Part I or in Part III of the experiment.) The data for other questions looks very similar.

Part I	*All subjects*			*Switched in Part I*		
Repetition	mean	median	st dev	mean	median	st dev
1st	13.05	9	12.30	14.29	10	14.07
2nd	8	6	5.20	9.41	8	6.64
3rd	7.51	5	5.22	9.16	6	6.03

	All subjects			*Switched in Part III*		
Part III	mean	median	st dev	mean	median	st dev
1st	10.41	6	11.45	8.16	6	7.6
2nd	2.85	2	2.33	3.02	2	2.72
3rd	2.3	2	1.66	2.66	2	1.94

Table 7 Response Time for HARD1, in seconds (the raw data approximated in seconds)

Second, the patterns of response times for three repetitions of the same question in Part I are very different from the ones corresponding to the three repetitions of the same question in Part III of the experiment. In particular, subjects take a relatively long time to think about each of the three repetitions in Part I. By contrast, in Part III they think long about the question only the very first time they encounter it. The next two repetitions are answered almost immediately: for example, in Table 7 we observe a median time of 2 seconds, which, considering that our response times are approximated to second (and not milliseconds), is not far from the minimum amount of time necessary to move the cursor across the screen to make a selection.

Taking again HARD1 question as an example, Table 8 shows the results of Wilcoxon Signrank test of equality of medians using matched observations (each row of this table is one test). The first and second column of Table 8 list the series that are compared, matched by subjects, while the third column reports the result of the test and the fourth column specifies p-value. We also depict some of the differences in the histograms 3.a below.

Finally, another way to interpret whether the second and the third repetition in Part III are fast or not is to compare how long it takes subjects to answer FOSD1 question in the second and the third repetition in Part III – which is naturally is the easiest question we have in our whole experiment, in which we can expect subjects to identify the optimal option in the 1st repetition and quickly move to choose it in the 2nd and 3rd repetitions. The histograms 3.b below show that the distributions of the difference in response times between second and third repetitions of HARD1 and FOSD1 questions are very similar to each other. Put differently, subjects take very short time to select the lottery in the second and the third repetition of any question in Part III of the experiment, especially once we take into account the time it takes to move the mouse and click on the lottery.

The data on response times is, therefore, suggestive of the fact that subjects treat each repetition as a new question in Part I of the experiment, but not on Part III. Speculating about possible explanation for the short response times in the second and third repetition of the same question, we hypothesize that subjects think about all three repetitions the first time they encounter the question in Part III, and don't re-think it in the subsequent repetitions. Of course, our evidence is only suggestive of this.

We conclude by noting that there is no relationship between decision time in Part III of the experiment and the tendency to report a stochastic answer in either part of the experiment.

	Series 1	Series 2	Hypothesis	p-value
Test 1	2nd Part I	2nd Part III	Rejected	$p < 0.01$
Test 2	3rd Part I	3rd Part III	Rejected	$p < 0.01$
Test 3	2nd Part I	1st Part III	Accepted	$p = 0.38$
Test 4	3rd Part I	1st Part III	Accepted	$p = 0.05$
Test 5	1st Part III	2nd Part III	Rejected	$p < 0.01$
Test 6	1st Part III	3rd Part III	Rejected	$p < 0.01$
Test 7	2nd Part III	3rd Part III	Accepted	$p = 0.05$

Table 8 Wilcoxon test of equality of medians with matched observations

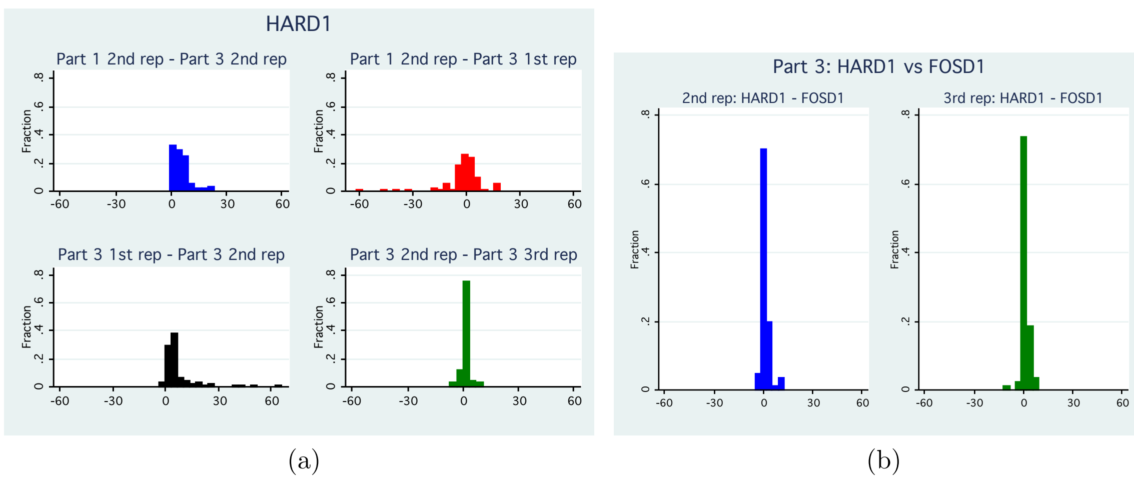

FIGURE 3 (a) Differences in median response times. (b) Difference in median response times, HARD vs FOSD.

Appendix B: Screenshots

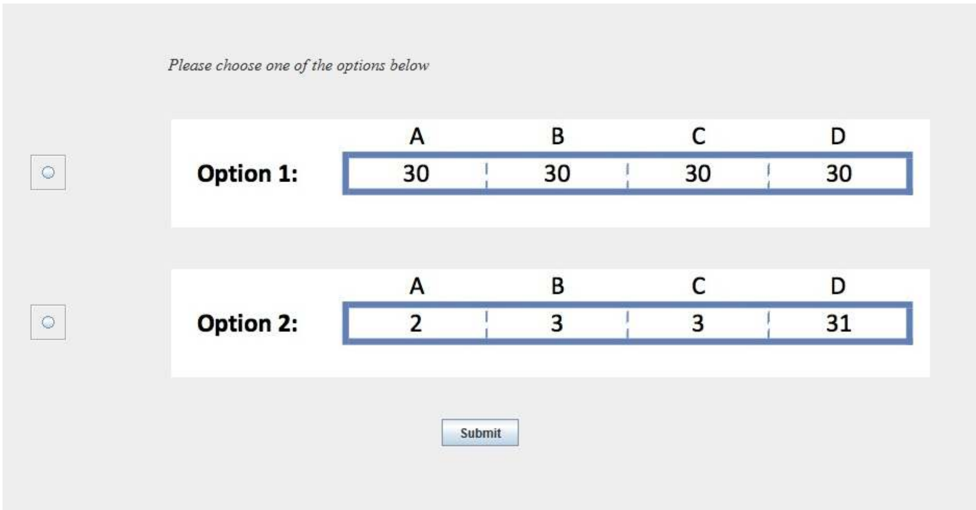

FIGURE 4 Screenshot of a typical question

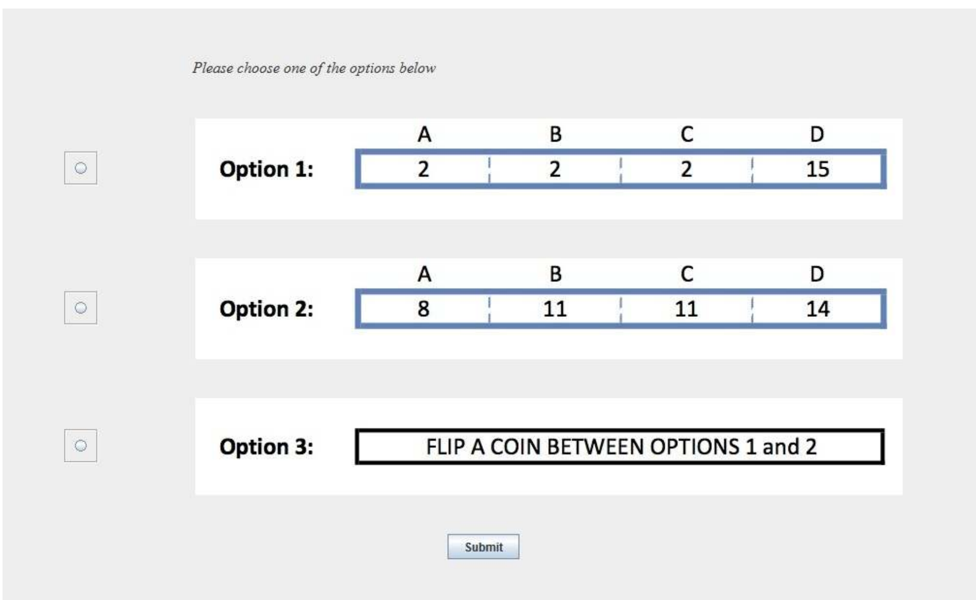

FIGURE 5 Screenshot of a typical question with the 'Flip a coin' option

Appendix C: Experimental Instructions, main treatment

This is an experiment in the economics of decision-making. The instructions are simple, and if you follow them carefully and make good decisions you may earn a CONSIDERABLE AMOUNT OF MONEY that will be PAID TO YOU IN CASH at the end of the experiment.

The currency in this experiment is called tokens. All payoffs are denominated in this currency.

There are four parts in this experiment. In Part I, III, and IV you will be asked to choose between lotteries (we will explain what they are shortly). In Part II, you will be asked to choose how much to invest in a risky investment. We will hand out specific instructions for each of the parts before the beginning of the part. You will see 40 questions in Part I, 2 questions in Part II, 21 questions in Part III, and 8 questions in Part IV.

Your payoff in the experiment will be determined as follows:

- The computer will choose one of the questions in Parts I or III with equal probability, and play out the lottery that you have chosen in that question. The amount of tokens that you receive for this lottery will then be converted to US dollars using the rate 20 Tokens = \$1.
- In addition, we will sum up all the tokens you earned in all the questions in Part II and IV, and convert them to US dollars using the rate 100 Tokens = \$1.
- Your total earning will consist of the amounts above plus a \$10 participation fee if you complete the experiment.

Part I

There are 40 questions in this part. In each question you will see either two or three options on the screen, depending on the question. Your task is to choose one of the available options. In every question, the first two options are two lotteries that pay a certain amount of tokens depending on chance. In particular, the computer will simulate the roll a four-faced die with faces named A, B, C, and D. Notice that each face is equally likely to occur. Depending on which face occurs, the lottery will pay a certain amount of tokens. Here is the example of such a lottery:

	A	B	C	D
Option 2:	95	120	170	185

That is: if we obtain face A then you will get 95 tokens; if we obtain face B then you will get 120 tokens; if we obtain face C then you will get 170 tokens; and if we obtain face D then you will get 185 tokens. Because each face is equally likely, and because there are only four faces, then this lottery pays: 95 tokens with probability 1/4, 120 tokens with probability 1/4, 170 tokens with probability 1/4, 185 tokens with probability 1/4.

Option 3: FLIP A COIN BETWEEN OPTIONS 1 and 2

In some of the questions, the following third option will also be available:

If you choose this option, the computer will simulate the flip of a coin to determine which of the two lotteries above you will be assigned. If HEADS come up, you will receive Option 1; if TAILS come up, you will receive Option 2. If you choose this third option, 1 token will be subtracted from number of tokens that are paid by the selected lottery. For example, if the lottery above is selected by the coin flip, and the die-roll selected D, then you get 184 tokens: the 185 tokens paid by the lottery minus 1 token.

In each screen you will see either two or three options, as described, and you will be asked to choose one of them. To choose an option click on the button next to the option and then click the SUBMIT button on the bottom of the screen. Once you clicked the SUBMIT button, your decision is recorded and you cannot change it. The computer will play out the option that you have selected right after you made your choice in every screen.

Notice:

- There is no right or wrong answer for any of these questions. We are interested in studying your preferences.
- Some of the questions might be repeated over the course of the experiment.
- After you click the submit button, there might be a short delay before the next question appears, due to the software. Please be patient.

Part II (*Displayed on the screen*)

Risk Question

You are endowed with 100 tokens and asked to choose the portion of this amount (between 0 and 100 tokens, inclusive) that you wish to invest in a risky option. Those tokens not invested are yours to keep.

If the risk investment is successful, you receive 2.5 times the amount you chose to invest; if the investment is unsuccessful, you lose the amount invested. To determine if the investment is successful or not, we will roll a four-faced die, with faces marked A, B, C, D. If we obtain face A or D the investment is successful. If we obtain faces B or C the investment is unsuccessful.

We now ask you to indicate the number of points that you wish to invest:

I wish to invest XXX tokens

Compound Lottery Question

You are endowed with 100 tokens and asked to choose the portion of this amount (between 0 and 100 tokens, inclusive) that you wish to invest in a risky option. Those tokens not invested are yours to keep.

If the risk investment is successful, you receive 2.5 times the amount you chose to invest; if the investment is unsuccessful, you lose the amount invested. To determine if the investment is successful or not, we will first of all roll a four-faced die, with faces marked A, B, C, D. If we obtain face A the investment is successful. If we obtain face D the investment is unsuccessful. If we obtain faces B or C, then roll the die again. If we obtain face A or B, the investment is successful. Otherwise, if we obtain face C or D, the investment in unsuccessful

We now ask you to indicate the number of points that you wish to invest:

I wish to invest XXX tokens

Part III (*Handed out after subjects finish Part II*)

In this part of the experiment you are asked again to choose between two lotteries, just like you did in Part I. The difference between this Part and Part I is that in this part you will see the same question three times, one after the other. That is: you will be asked one question; once you click the submit button, you will be asked the same question again; and once you click the submit button, the same question will appear for the third time. Once you click the submit button, then a new question will appear, which will also be asked three times. There are a total of 7 questions, each asked three times, for a total of 21 decisions to be made.

Please remember that there is no right or wrong answer.

Part IV (*Displayed on the Screen*)

Screen 1

Please choose one of the options below:

- Option 1: 100 tokens with certainty
- Option 2: 10% chance of 500 tokens, 89% chance of 100 tokens and 1% chance of 0 tokens

Screen 2

Please choose one of the options below:

- Option 1: 11% chance of 100 tokens and 89% chance of 0 tokens
- Option 2: 10% chance of 500 tokens and 90% chance of 0 tokens

Screen 3

Please choose one of the options below:

- Option 1: 100 tokens with certainty

- Option 2: 98% chance of 500 tokens and 2% chance of 0 tokens

Screen 4

Please choose one of the options below:

- Option 1: 1% chance of 100 tokens and 99% chance of 0 tokens
- Option 2: 0.98% chance of 500 tokens and 99.02% chance of 0 tokens

Final Questionnaire (*Handed out at the very end of the experiment*)

1. Did you have the impression that there were some questions that were 'easier than others, i.e., for which there was one option that was clearly better? (Yes or no is enough)
2. In Part III of the experiment each question was asked to you three times. If you choose different options, could you please tell us why did you do it? (Please elaborate)
3. Did you ever choose the 'flip a coin option? If so, why did you do it? If not, why didnt you choose it?
4. Would you have chosen to flip a coin if there were no 'cost to chose it, i.e., if it simply choose one of lotteries without imposing a cost?
5. If you did not chose the 'flip a coin option in the experiment but you would have chosen it if it didnt have a cost, why is it the case?
6. In Part I of the experiment, some questions were asked multiple times. Did you notice? Do you think you gave the same answer? If you gave different answers, why did you choose different answers? (Feel free to say same as above)

Appendix D: Instructions for Robustness Test

This is an experiment in the economics of decision-making. You may earn a CONSIDERABLE AMOUNT OF MONEY that will be PAID TO YOU IN CASH at the end of the experiment.

The currency in this experiment is called tokens. All payoffs are denominated in this currency.

In this experiment you will be asked to choose between two options, which are two lotteries that pay a certain amount of tokens depending on chance. In particular, at the end of the experiment, one of the participants will roll a four-faced die with faces named A, B, C, and D. Notice that each face is equally likely to occur. Depending on which face occurs, the lottery will pay a certain amount of tokens. Here is the example of such a lottery:

	A	B	C	D
Option 2:	95	120	170	185

That is: if we obtain face A then you will get 95 tokens; if we obtain face B then you will get 120 tokens; if we obtain face C then you will get 170 tokens; and if we obtain face D then you will get 185 tokens.

Because each face is equally likely, and because there are only four faces, then this lottery pays:

- 95 tokens with probability 1/4
- 120 tokens with probability 1/4
- 170 tokens with probability 1/4
- 185 tokens with probability 1/4

In each question you will see two lotteries, as described above, and you will be asked to choose one of them. To choose a lottery please circle it. Once you made your choice, you can turn the page and move on to the next question.

Importantly, in this experiment you will be asked to choose between two lotteries three times in a row. The three questions will be identical, one after the other. That is: you will be asked to choose between two lotteries; you will then proceed to the next sheet, in which you will be asked again to choose between the same two lotteries; and then, the same question will appear for the third time.

After you answer these three questions, the experiment will end.

Your payoff in the experiment will be determined as follows:

- At the end of the experiment, each participant will roll a six-sided die to determine which of the three questions will be paid. If we obtain 1 or 2, then the first question will be selected for payment. If we obtain 3 or 4, then the second question will be selected for payment. If we obtain 5 or 6 then the third question will be selected for payment. In other words, each of the three questions that you answer are equally likely to be selected for payment.

- Once a question is selected for payment, we will play out the lottery that you have chosen in that question. That is, each participant will roll a four-faced die to determine the amount of tokens that you win. The amount of tokens that you receive for this lottery will then be converted to US dollars using the rate 2 Tokens = $1.

Notice: There is no right or wrong answer for any of these questions. We are interested in studying your preferences.

Are there any questions?

Questionnaire

Did you choose different options in Questions 1 3?

If you did, could you please tell us why did you do it? Please elaborate.

Appendix E: Order of Questions

Part I of the experiment

In the table below we present two different orders of questions that were used in the experiment. Each subject was randomly assigned to one of the orders. EASY2 + coin" refers to the screen in which subjects were asked to select one of the three alternatives: the first two alternatives are the lotteries from EASY2 question and the third alternative is computerized coin which costs 1 token to select.

Table 9 Order of screens in Part I of the experiment

Screen	Order 1	Order 2	Screen	Order 1	Order 2
1	FOSD1	FOSD3	21	HARD1	EASY3 + coin
2	EASY2 + coin	EASY1	22	HARD3 + coin	FOSD3
3	EASY3	HARD3	23	FOSD1	HARD2 + coin
4	HARD4	FOSD1 + coin	24	HARD2	EASY2
5	EASY1 + coin	FOSD2	25	EASY3	HARD3
6	HARD1	HARD4	26	FOSD2	EASY1
7	FOSD3	EASY3	27	HARD4	HARD4 + coin
8	HARD3	HARD2	28	FOSD3 + coin	HARD1
9	FOSD2 + coin	EASY2 + coin	29	EASY2	FOSD2
10	HARD2	HARD1	30	EASY1	FOSD1
11	EASY2	EASY1	31	HARD3	EASY1 + coin
12	EASY1	FOSD2 + coin	32	EASY3 + coin	HARD2
13	HARD3	FOSD1	33	HARD4	EASY2
14	FOSD1 + coin	HARD3	34	FOSD2	FOSD3
15	HARD1	EASY2	35	EASY2	HARD3 + coin
16	EASY3	FOSD3 + coin	36	HARD1 + coin	EASY3
17	FOSD2	EASY3	37	EASY1	HARD4
18	HARD4 + coin	HARD2	38	FOSD3	FOSD1
19	FOSD3	HARD4	39	HARD2 + coin	FOSD2
20	HARD2	HARD1	40	FOSD1	HARD1 + coin

Part III of the experiment

Table 10 Order of screens in Part III of the experiment

Screen	Order 1	Order 2		Screen	Order 1	Order 2
1	HARD3	HARD1		13	EASY2	FOSD1
2	HARD3	HARD1		14	EASY2	FOSD1
3	HARD3	HARD1		15	EASY2	FOSD1
4	FOSD1	EASY1		16	HARD2	HARD4
5	FOSD1	EASY1		17	HARD2	HARD4
6	FOSD1	EASY1		18	HARD2	HARD4
7	HARD4	HARD2		19	EASY1	HARD3
8	HARD4	HARD2		20	EASY1	HARD3
9	HARD4	HARD2		21	EASY1	HARD3
10	HARD1	EASY2				
11	HARD1	EASY2				
12	HARD1	EASY2				