

# STOCHASTIC CHOICE AND PREFERENCES FOR RANDOMIZATION\*

Marina Agranov<sup>†</sup>    Pietro Ortoleva<sup>‡</sup>

September 9, 2015

## Abstract

We conduct an experiment in which subjects face the same questions repeated multiple times, with repetitions of two types: 1) following the literature, the repetitions are distant from each other; 2) in a novel treatment, the repetitions are *in a row*, and subjects are told that the questions will be repeated. We find that a large majority of subjects exhibit stochastic choice in *both* cases. We discuss the implications for models of stochastic choice.

JEL: *C91, D11, D81.*

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

---

\*We thank the editor and the anonymous referees for their many useful suggestions that significantly improved the paper. We also thank Mark Dean, David Dillenberger, Federico Echenique, Kfir Eliaz, Guillaume Frechette, Yoram Halevy, Mark Machina, Daniel Martin, Efe Ok, Gil Riella, Kota Saito, Chris Shannon, Leeat Yariv, and participants at many seminars and conferences for helpful comments and suggestions. Ortoleva gratefully acknowledges the financial support of NSF grant SES-1156091 and of Caltech's Gordon and Betty Moore Foundation through Grant GBMF1158.

<sup>†</sup>Division of the Humanities and Social Sciences, California Institute of Technology, 1200 E. California Blvd, Pasadena, CA 91125 USA. Email: [magranov@hss.caltech.edu](mailto:magranov@hss.caltech.edu).

<sup>‡</sup>Department of Economics, Columbia University, 420 West 118th St., New York NY 10027 USA. Email: [pietro.ortoleva@columbia.edu](mailto:pietro.ortoleva@columbia.edu).

# 1. Introduction

A consistent finding regarding individual decision making is the phenomenon of stochastic, or random, choice: when asked to choose from the same set of options many times, subjects often make different choices.<sup>1</sup> Stochastic choice is documented in many environments, including those in which subjects have no value for experimentation (e.g., when there is no feedback) and those in which there are no bundle or portfolio effects (e.g., when only one choice is paid).

This robust finding has led to the development of a large body of theoretical models that capture this behavior. These models can be ascribed to three broad classes: 1) models of *Random Utility*, or Preferences, in which subjects' answers change because their preferences change stochastically; 2) models of *Bounded Rationality*, in which subjects have stable preferences but exhibit stochastic choice as they may fail to choose the best option for them; 3) models of *Deliberate Randomization*, in which subjects deliberately choose to report different answers because it is optimal for them to do so (e.g., to minimize regret or to hedge between options).

The goal of this paper is to shed light on the origin of stochastic choice and, in particular, to inform the three broad classes of theories described above.

We use the following design. Subjects are asked to make several choices between objective lotteries and are paid for one random decision round out of all rounds played. In the first part of the experiment, we replicate a standard design: subjects are asked a set of questions repeated several times, with the repetitions distant from each other (separated by other questions); subjects are not told in advance of these repetitions. The novelty of our experiment is to introduce a second part in which subjects face the same question three times *in a row* and are explicitly told that each question will be repeated three times.

As a further test of the desire to randomize, for some questions, subjects are allowed to choose either one of two lotteries or a (computer-simulated) coin flip to determine which lottery they are assigned; selecting the coin has a small fee.

We also elicit subjects' attitudes towards risk, compound lotteries, and their prudence to violate Expected Utility as captured by the Allais paradox. Finally, as we wish to study the motivation underlying stochastic choice, in a non-incentivized questionnaire distributed at the end of the experiment, we ask subjects directly if, and why, they choose different answers when questions are repeated in a row. Our

---

<sup>1</sup>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 large fraction of subjects gave different answers to the same question. Many studies have replicated this result, focusing on choices between risky gambles: 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).

main findings are the following.

First, in line with previous results, the vast majority of participants (90%) choose different lotteries in the three repetitions of the same question when these repetitions are distant from each other. They tend to do so for at least half of the questions within the class of questions for which stochastic choice is prevalent.

Second, a large majority of subjects (71%) select different lotteries also when questions are repeated three times *in a row* and they are explicitly told about the repetition. Those who exhibit stochastic choice do so multiple times. Stochastic choice behavior is strongly correlated in the two cases (distant and in-a-row repetitions).

Third, in both cases, stochastic choice is present almost exclusively in questions in which none of the available options is “clearly better” than the other (what we call “hard” questions); it is extremely frequent for these questions and virtually absent for others. This distinction is the strongest predictor of stochastic choice in our data. Hard questions are not necessarily the questions in which the expected values, or utilities, are the closest. In fact, differences in expected utility between the options have limited predictive power in determining the stochasticity of choice in our data and cannot account for the variation in stochastic choice.

Fourth, 29% of the subjects choose the option to flip a (costly) coin at least once (most of these subjects did so multiple times), again only for “hard” questions.

We note two additional patterns. Stochastic choice is significantly correlated with violations of Expected Utility *à la* Allais but generally not with risk aversion or attitude towards compound lotteries. Also, the analysis of response time shows that subjects behave very differently in distant versus in-a-row repetitions.

We conclude our analysis by looking at the answers to the final questionnaire in which subjects provided reasons for making different choices in the in-a-row repetitions of the same question. We find that almost all subjects who gave different answers with in-a-row repetitions reported doing so in the questionnaire and that the vast majority (79%) reported doing so deliberately. Typical motivations given by subjects were about hedging and diversification.

To frame our analysis, we extend existing models of stochastic choice so that they make predictions about distant and in-a-row choices. For models of Random Utility, we consider the Random Expected Utility model of Gul and Pesendorfer (2006), and we posit that the stochastic component of the utility does not change for in-a-row repetitions. For models of Bounded Rationality, we consider the Drift Diffusion Model of Ratcliff (1978) and Ratcliff and McKoon (2008), and we posit that the agent does not collect more information for those repetitions. For models of Deliberate Randomization, we consider the Cautious Stochastic Choice (CSC) model of Cerreia-Vioglio et al. (2015b).

We interpret our results as suggesting that the main driving force behind stochastic choice in our data is the *deliberate desire of the subjects to choose different answers*, which is consistent with models of Deliberate Randomization and not with models in other classes. We find that a small fraction of subjects, 6%, never exhibit stochastic choice. We also find that 23% report stochastic choice for distant but not for in-a-row repetitions, consistent with models of Random Expected Utility and Bounded Rationality. However, neither of these two classes of models, under our assumptions, can account for the behavior of the majority of subjects, 61%, that report stochastic choice for both distant and in-a-row repetitions. Models of Deliberate Randomization instead correctly predict stochastic choice behavior in both cases. They also predict the correlation between stochastic choice and the Allais-like behavior observed in the data. This interpretation of stochastic choice is also supported by the open-ended answers to the questionnaire.

It is important to stress that we can only test among these classes of models because of the intertemporal structure that we add to the Random Utility and Bounded Rationality Models. Our data do not, for example, rule out a Random Utility model in which choice-specific utility shocks arise for consecutive questions that the agent knows are identical. Our implementation is consistent with how random utility models are often interpreted (see Luce 1958, Becker et al. 1963b and the discussion in Section 2), but under other interpretations, our tests would not be decisive.

To test the robustness of our findings, we ran three additional short experiments. In the first, we asked only one of the questions from our main experiment, repeated three times in a row, with ten times higher stakes. Half of the subjects chose different answers in the three repetitions, a fraction very similar to that observed in the main experiment for the same question. The second and the third short experiments asked again only one question repeated three times, but questions were no longer about lotteries: in the second short experiment, the question offered a choice between different amounts of money at different moments in time (time preferences); in the third, it was a choice between different allocations to different participants (social preferences). We find that 34% and 42%, respectively, report stochastic answers in each of these experiments.

This paper is related to the experimental literature on choice under uncertainty (Camerer, 1995) and in particular to the studies on stochastic choice and preferences for randomization. Hey and Carbone (1995) test experimentally whether preferences are deterministic while choice is stochastic; their results rule out this possibility. Becker et al. (1963a) and Sopher and Narramore (2000) ask subjects to choose between two lotteries or a convex combination of them and document a strong tendency to choose the latter. Rubinstein (2002) documents a deliberate desire to report “diversified” answers even when this leads them to strictly dominated choices.<sup>2</sup> We

---

<sup>2</sup>Other recent studies document “false diversification.” In the experiment of Chen and Corter

document similar patterns, but in our case, the tendency to diversify appears only for “hard” questions and does not generate violations of first-order stochastic dominance. The experiment of Kircher et al. (2013) is a version of the dictator game in which the dictators can choose between 7.5 Euros for themselves and 0 to the recipient, 5 Euros to both, or a lottery between the options above. They find that approximately one third of the subjects chose to randomize. This finding is in line with the result of our robustness test on social preferences and with the broader notion of preferences for randomization. Finally, in a recent and independent study, Dwenger et al. (2013) explore, both in the laboratory and in real data, whether subjects wish to delegate their choice to an external device to avoid making decisions. They find that 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. They also show similar patterns using the data from a clearinghouse for university admissions in Germany, where the application process is such that students must submit multiple rankings of the universities they would like to attend. These rankings are submitted at the same moment in time, but only one of them matters. They find that a significant fraction of students report inconsistent rankings, even when there are no strategic reasons to do so.

The remainder of the paper is organized as follows. Section 2 discusses the implications of the theoretical models in each class. Section 3 presents the experimental design, and the results are analyzed in Section 4. Section 5 discusses the three robustness experiments. Section 6 concludes the paper. The Online Appendix contains additional analysis and experimental instructions.

## 2. Models of Stochastic Choice

We now turn to describe the three main classes of models of Stochastic Choice: Random Utility/Preferences, Bounded Rationality, and Deliberate Randomization. While each class includes many, often dozens, of models, they tend to have many features in common. For each class, we select one prominent model, describe it in detail, and discuss its implications for our behavior of interest.

Consider an interval  $[w, b]$  of monetary prizes and the set  $\Delta$  of probability distributions (lotteries) over them, with generic elements  $p, q$ . We are interested in the stochastic choice function  $\rho$  of the agent: a map that associates a probability measure  $\rho(A)$  over  $A$  to each finite subset  $A$  of  $\Delta$ . It represents the frequency with which each element of  $A$  is chosen assuming that the choice from  $A$  is asked repeatedly. While

---

(2006), subjects chose an irrational mixture of options, including dominated ones, in multiple-trial decisions over pure bundles. In Eliaz and Fréchet (2008), a significant fraction of subjects paid to switch from a lottery that pays in only one state to one that pays in more states, even though the overall distribution remains constant.

typically no discussion is made regarding the frequency of these repetitions, the key feature of our experiment is that we study stochastic choice in two distinct cases:

**Distant Repetitions** The agent is asked to choose between two lotteries multiple times, with each repetition distant from each other; the agent is not told that the questions are repeated;

**Repetitions in a Row** The agent is asked to choose between two lotteries multiple times, with repetitions in a row; the agent is explicitly told that each question will be repeated multiple times in a row and the number of repetitions.

*Random Utility* A well-known class of models of stochastic choice is that of Random Utility, or Random Preferences, according to which when subjects make a decision, they maximize a well-defined utility function (or preference), but this changes stochastically over time.<sup>3</sup> The relevant model for our analysis of choice over lotteries is *Random Expected Utility* (REU), studied in Gul and Pesendorfer (2006):<sup>4</sup> the agent has a probability distribution  $\mu$  over strictly increasing utility functions over money, with support  $U$ . The probability of choosing  $p$  from a set  $A$  is equal to the probability that the agent has a utility that, among the elements of  $A$ , is uniquely maximized at  $p$ .<sup>5</sup> That is, for all  $p \in A$ ,

$$\rho(A)(p) = \mu(\{u \in U : u(p) > u(q) \text{ for all } q \in A\}).$$

The common interpretation of why the utility function is stochastic suggests that the subject's utility changes due to changes in exogenous, unobservable subjective and objective conditions, such as information, mood, social situation, framing, *etc.* (see, among many, Harsanyi 1973; Gul and Pesendorfer 2006; McFadden 2006).<sup>6</sup>

---

<sup>3</sup>See Thurstone (1927), Luce (1959), Becker et al. (1963b), 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), Apesteguia and Ballester (2015), Fudenberg and Strzalecki (2015).

<sup>4</sup>In principle, one can consider a Random Utility model with non-Expected Utility preferences. This type of more general model would include features of both Random Utility and Deliberate Randomization models (discussed below) because subjects may have both a changing utility and an explicit desire to randomize.

<sup>5</sup>This is what Gul and Pesendorfer (2006) call a regular Random Utility Function with countably additive measure. The only difference is that in Gul and Pesendorfer (2006), lotteries are defined over arbitrary prizes, while we focus on monetary lotteries; we thus posit that utilities are strictly increasing.

<sup>6</sup>Since the early contributions in this literature, there has been a discussion on how to test models with changing utilities experimentally and estimate the choice probabilities as independent. Various papers suggested that for this to be meaningful, repetitions should be far apart and subjects should be unaware of them (e.g., Luce 1958, p.217, Becker et al. 1963b, p.45).

Motivated by the interpretation above, we denote by REU\* the REU model in which the utility function is fixed for repetitions in a row but can vary across distant repetitions. It is REU\* that we will test in our experiment. According to this model, we can observe stochastic choice with distant repetitions but not with repetitions in a row. Note also that under REU\*, subjects will never choose a first-order stochastically dominated option.

*Bounded Rationality* A second class of models assume that subjects have a well-defined and stable ranking of the available options but may not choose the alternative that maximizes it because of some form of bounded rationality.<sup>7</sup>

We illustrate this class of models using the Drift Diffusion Model (DDM) of Ratcliff (1978) and Ratcliff and McKoon (2008). Suppose that the agent must choose between two options,  $p$  and  $q$ , with values  $u(p)$  and  $u(q)$ . At every instant, she receives a noisy signal in favor of one of them. Assuming that positive values indicate that  $p$  is better (negative indicating the opposite), at each instant the agent adds all the evidence accumulated and does one of the following: chooses  $p$  if the accumulated evidence is above a threshold  $b > 0$ ; chooses  $q$  if it is below  $-b$ ; continues to acquire information otherwise. The accumulated evidence  $X(t)$  starts at 0 and evolves according to

$$X(t) = X(t - 1) + \alpha(u(p) - u(q)) + \epsilon(t),$$

where  $\alpha \in (0, 1)$ ,  $\alpha(u(p) - u(q))$  is the drift rate of the process, and  $\epsilon(t)$  is an i.i.d. normally distributed noise (mean zero, unitary variance).

The predictions of the DDM for the case of distant repetitions are clear. Agents will choose more frequently the option with a higher utility but may make mistakes and exhibit stochastic choice. This can happen for any question, including those involving FOSD. Stochastic choice will be more frequent the smaller the expected utility difference between the two options, as the drift is smaller in this case. For these questions, the response time should also be longer (as the process is more likely to take longer to reach the threshold).

Denote by DDM\* the DDM model with the modification that no additional information is collected when the same question is asked multiple times in a row and the agent is aware that it is the same question.<sup>8</sup> It is DDM\* that we will test in our experiment. According to this model, stochastic choice can be observed with distant repetitions but not with repetitions in a row. Note that under DDM\*, response times should be shorter in the second and third consecutive repetitions of the same choice.

---

<sup>7</sup>See, among many, in Busemeyer and Townsend (1993), Harless and Camerer (1994), Hey and Orme (1994), Camerer and Ho (1994), Wu and Gonzalez (1996), Ratcliff and McKoon (2008), Wilcox (2011), Woodford (2014), Gul et al. (2014), Fudenberg and Strzalecki (2015), Natenzon (2015); see also Johnson and Ratcliff (2013) for reviews.

<sup>8</sup>Indeed, if it was optimal for the agent to collect more information, she should have done so before answering the first of the repetitions.

*Deliberate Randomization* A third class of models of stochastic choice postulates that the stochasticity is a *deliberate choice* of the agent.<sup>9</sup> We illustrate this class using the Cautious Stochastic Choice (CSC) model of Cerreia-Vioglio et al. (2015b); identical predictions hold for the other models.

We add the following notation. For any set  $A$ , note that  $\rho(A)$  is a probability distribution over lotteries, thus a compound lottery, and denote by  $\overline{\rho(A)}$  the lottery it induces over final outcomes, that is,  $\overline{\rho(A)} := \sum_{q \in A} \rho(q)q$ . For any subset  $A$  of  $\Delta$ , denote by  $co(A)$  its convex hull.

In the CSC model, the agent has a compact *set* of utility functions  $\mathcal{W}$  over monetary amounts, all of which are continuous, strictly increasing, and concave. The stochastic choice is then represented by

$$\overline{\rho(A)} \in \arg \max_{p \in co(A)} V(p),$$

where

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

The model has two components. First, when subjects are asked to choose from a set  $A$ , they do not just consider the options in  $A$  but rather consider all possible randomizations over them and choose the optimal one according to  $V$ . They may deliberately choose to randomize in case the utility of the mixture is higher than that of the two options: if  $V(\frac{1}{2}p + \frac{1}{2}q) > V(p), V(q)$ , then from  $\{p, q\}$ , they would choose sometimes  $p$  and sometimes  $q$ . Their preference for randomization comes from the second component of the model, the shape of  $V$ , that follows the Cautious Expected Utility model of Cerreia-Vioglio et al. (2015a). The agent has a *set* of utility functions  $\mathcal{W}$ , and she values a lottery  $p$  as follows: for every utility  $v$  in  $\mathcal{W}$ , she computes the certainty equivalent of the lottery,  $v^{-1}(\mathbb{E}_p(v))$ ; she evaluates  $p$  as the *smallest* of these certainty equivalents. This may lead to a desire to hedge between options: if  $p$  is good for one utility but bad for another, and  $q$  is the opposite, by mixing between  $p$  and  $q$ , the agent obtains a lottery that is not too bad for both utilities. This agent may therefore have a desire to randomize between  $p$  and  $q$ , leading to stochastic choice.

The predictions of the CSC model are clear: the agent may exhibit stochastic choice both when repetitions are distant and when they are in a row.

---

<sup>9</sup>First suggested in Machina (1985), models in this class proposed different reasons for the desire to randomize: to minimize regret (Dwenger et al., 2013) or because subjects have non-Expected Utility preferences and wish to hedge between options (Marley, 1997; Swait and Marley, 2013; Henderson et al., 2014; Cerreia-Vioglio et al., 2015b; Fudenberg et al., 2015).

### 3. Design of the Experiment

The experiment is composed of four parts. The complete instructions and screenshots are presented in Online Appendices B and C. 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.

The main parts of the experiment are Parts I and III, in which subjects were asked many questions repeated multiple times. In Part I, repetitions of the same question were far apart, and subjects were not told that questions would be repeated. In Part III, the three repetitions were asked in a row, and subjects were explicitly told that each question would be repeated three times. Between Parts I and III, we measured subjects' risk and compound lottery attitudes using an investment task; this part was meant to break the repetitiveness of the questions in Parts I and III and to reduce fatigue. After Part III, we measured violations of Expected Utility. We concluded the experiment with a non-incentivized questionnaire. In most rounds, subjects were asked to select one of the lotteries displayed on the screen. Each lottery paid a certain number 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 had at most four different outcomes occurring with probabilities 0.25, 0.5, 0.75, or 1. Each lottery was presented as a table that listed the number of tokens paid for each face of the die.

Throughout the experiment, we repeatedly asked the same 10 basic questions, each involving the choice between two lotteries. Table 1 contains the 10 questions. They can be divided into three groups: FOSD, EASY, and HARD. The first included two lotteries, one of which was first-order stochastically dominated by the other. EASY questions involved lotteries with no first-order stochastic dominance but in which one lottery was “clearly better” than the other. By contrast, each of the four HARD questions had no “obvious winner” between two lotteries.<sup>10</sup>

In Part I, subjects played 40 decision rounds with each of the ten questions above repeated four times in different orders. These were designed to guarantee that each question would not appear too close to its repetition. In three of the four repetitions, subjects observed exactly the same question displayed in exactly the same way.<sup>11</sup> In the remaining repetition, in addition to the two lotteries, subjects were also offered

---

<sup>10</sup>The design of EASY and HARD questions was inspired by a pilot experiment (see Online Appendix E) in which we recorded choices and response times to different questions. EASY questions were designed so that most participants choose the same option with a short response time; HARD ones instead had a choice distribution close to uniform and a long response time.

<sup>11</sup>To test for order effects, each subject was randomly assigned to one of the two possible orders with a different sequence of questions. In Online Appendix D, we present the exact orders of questions used in the experiment and show that all results hold if we focus on any of these two orders.

TABLE 1 List of Questions Asked

Question	Lottery 1				<i>vs.</i>	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

*Notes:* Each question is one row: subjects were asked to choose between lotteries 1 and 2. Each lottery is described by the four amounts of tokens that it can pay out with equal probability. The tokens are converted into US dollars at the rate 20 tokens = \$1. The last column contains the absolute value of the difference in Expected Values. A screenshot is presented in Online Appendix C.

the option to have the computer flip a (simulated, fair) coin to determine which of the two lotteries would be assigned to them. This third option, however, was not free: if chosen, 1 token would be subtracted from the amount won.

Part II included two investment tasks. In the first, developed by Gneezy and Potters (1997) and Charness and Gneezy (2010), subjects were endowed with 100 tokens and chose how many to invest in a risky project that has 50% chance of success and returns 2.5 the investment if successful, and nothing otherwise. The amount of tokens invested can be used to estimate the subject's risk aversion. The second task was identical to the first except that the success of the investment was determined by a compound lottery that reduces to a 50% chance of success. Subjects with neither aversion nor attraction to compound lotteries should invest the same number of tokens in both questions; subjects who dislike (like) compound lotteries might decide to invest less (more) in the second investment task.

In Part III, subjects were again asked seven out of the ten questions used in Part I, repeated multiple times: FOSD1, EASY1, EASY2, HARD1, HARD2, HARD3 and HARD4. As opposed to Part I, the questions were asked three times in a row, for a total of 21 decision rounds. Subjects were explicitly told that each question would be repeated three times in a row.

Finally, Part IV of the experiment asked standard variations of the Common Ratio and Common Consequences effects of Allais (1953). The experiment concluded with a questionnaire in which subjects were asked questions regarding their choices. In particular, subjects were asked whether they chose different answers in the repetitions of the same question in Part III and, if so, why they did so.

All sessions were conducted at the California Social Science Experimental Laboratory (CASSEL) at University of California, Los Angeles in January 2013.<sup>12</sup> Subjects were recruited from a database of volunteer undergraduate students. Four identical sessions were run, for a total of 80 subjects. No subject participated in more than one session. Each session lasted approximately 45 minutes, and average earnings were \$19.

The payment of subjects was comprised of several parts. First, they received a \$10 show-up fee for completing the experiment. Second, subjects were paid for *one* decision round, randomly selected with equal probability from all rounds in Parts I and III (combined). Because the amounts paid by each lottery were described in tokens, these were converted into US dollars using the rate 20 tokens = \$1. The procedure in which only one decision round is paid is standard and is used to avoid incentives to choose different answers to the same questions to create a portfolio. Third, subjects were paid for all decisions made in Parts II and IV, which are less prone to portfolio concerns. These tokens were converted into US dollars using the rate 100 Tokens = \$1. The different conversion rate in Parts II and IV was chosen to reflect the difference in the scales of earnings and to create strong incentives for subjects to think hard about choices in Parts I and III, which are the primary focus of the experiment.

Subjects received no feedback about their earnings throughout the experiment. However, at the end of each decision round, the lottery chosen by a subject was played out by the computer, and its outcome was recorded by the software but not revealed to the subject. Subjects were informed about this. At the end of the experiment, subjects learned how much money they earned in total and were paid in cash privately.

## 4. Results

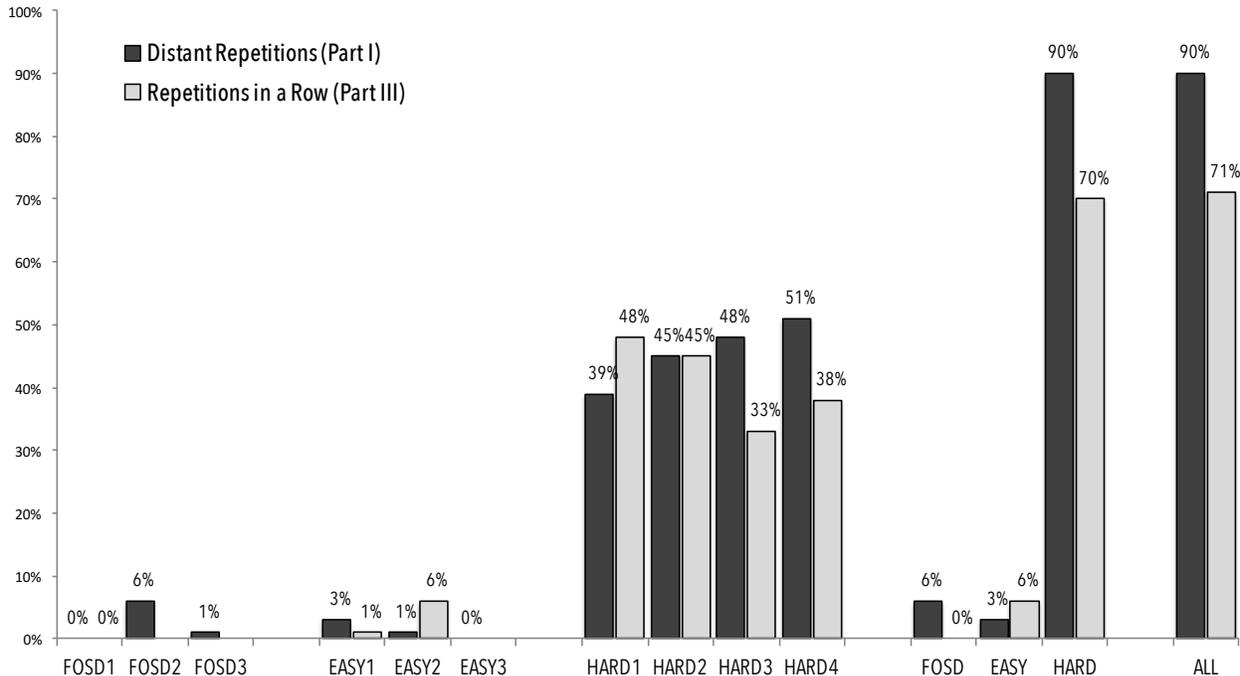
### 4.1 Behavior in Parts I and III

Figure 1 presents the fraction of subjects who give inconsistent answers in Parts I and III in the repetitions of each question, at least once for questions in each class (FOSD, EASY, and HARD) and at least once overall (ALL).

---

<sup>12</sup>The software was programmed as server/client applications in Java, using the open source ex-

FIGURE 1 Fraction of Subjects who Reported Stochastic Choice in the Main Experiment



*Notes:* In Part I, ten questions were repeated three times, with the repetitions distant from each other. In Part III, subjects were asked seven out of the ten questions, repeated three times in a row. In three questions, labeled FOSD, one lottery first-order stochastically dominates the other. In other three, labeled EASY, one lottery is ‘clearly’ better than the other. In the last four questions, labeled HARD, each of the lotteries is appealing in its own way. The results are presented for each question, for each category (FOSD, EASY, HARD), and overall (ALL): when multiple questions are considered, the bar indicates the fraction of subjects who reported stochastic choice in at least one of the questions included.

In Part I, where repetitions are distant, the vast majority of participants (90%) choose different lotteries in the three repetitions of the same question. This is in line with existing evidence. This stochastic behavior is not only widespread across subjects but also frequent for each subject: the majority (60%) of those that exhibit stochastic choice do so for at least half of the HARD questions (two or more out of four). A strong additional pattern emerges: stochastic behavior is present essentially only in HARD questions and is almost absent in EASY and FOSD ones. Fisher exact tests confirm that the proportion of subjects who give inconsistent answers in HARD questions is significantly higher than that in EASY questions ( $p < 0.01$ ) and that in FOSD questions ( $p < 0.01$ ), while there is no statistical difference between those reporting inconsistent answers in EASY and in FOSD questions ( $p = 0.443$ ).

perimental software Multistage ( <http://multistage.ssel.caltech.edu/> ).

The main finding of our experiment is the behavior in Part III, in which questions are repeated three times in a row. A large majority of subjects (71%) reported inconsistent answers in this case as well. Just as in Part I, this behavior takes place essentially only for HARD questions (Fisher exact test:  $p < 0.01$  for HARD vs. EASY;  $p < 0.01$  for HARD vs. FOSD; and  $p = 0.059$  for EASY vs. FOSD). While indeed the proportion of subjects who reported inconsistent answers in Part III is smaller than that in Part I (Fisher exact test,  $p = 0.005$ ), it still involves the large majority of the population. Similar to Part I, in Part III not only do most subjects report inconsistent answers, but they do so more than once: the majority (70%) of subjects who report inconsistent answers for HARD questions does so for at least two out of four of them.

Stochastic behavior in Part I and Part III is highly correlated at the individual level. Table 2 presents a regression analysis in which the indicator for stochastic choice in Part I is regressed on the indicator for stochastic choice for the *same* question in Part III while clustering standard errors by subject. The first regression considers all questions; the second one focuses on HARD ones. It shows a positive and significant correlation between the tendency to exhibit stochastic choice in Parts I and III at the individual level.

TABLE 2 Regression Analysis of Stochastic Choice

	SC Distant Repetitions (Part I)			SC Rep. in a row (Part III)
	(1)	(2)	(3)	(4)
SC in Rep. in a row	0.35** (0.04)	0.14** (0.06)		
Flip Coin			0.23** (0.06)	0.27** (0.07)
Constant	0.18** (0.02)	0.40** (0.04)	0.18** (0.01)	0.22** (0.02)
# of obs.	560	320	800	560
# of sub.	80	80	80	80
R-sq	0.1139	0.0223	0.0169	0.0262
Sample of Questions	All	HARD	All	All

*Notes:* Random-effects GLS regressions with standard errors clustered at the subject level. Each observation corresponds to an individual behavior for one of the ten (seven) questions subjects faced repeatedly in Part I with Distant Repetitions (Part III with Repetitions in a Row). The dependent variable in panels (1) - (3) is an indicator of Stochastic Choice with Distant Repetitions (Part I). The dependent variable in panel (4) is an indicator of Stochastic Choice with Repetitions in a Row (Part III). The Flip Coin variable is a dummy variable that takes value one if the subject chose the costly coin in that question and zero otherwise. \*\* indicates significance at the 5% level.

Table 3 classifies subjects according to their behavior in Parts I and III in light of the model predictions discussed in Section 2. Several patterns emerge. First, the behavior of the vast majority of subjects (90%) is consistent with Expected Utility (EU) or with one of the three models described in Section 2: REU\*, DDM\* or CSC

models. Second, very few subjects (6%) never report inconsistent answers and are categorized as being consistent with standard Expected Utility. Third, there is a significant fraction of subjects (23%) who exhibit stochastic choice in EASY and/or HARD questions in Part I but never in Part III, which is consistent with both REU\* and DDM\* models. Finally, the majority of subjects (61%) exhibit stochastic choice behavior in both Parts I and III, which is consistent only with models of Deliberate Randomization.<sup>13</sup>

TABLE 3 Classification of Subjects according to the Theoretical Models

Model	Stochastic Choice Distant Repetitions	Stochastic Choice Repetitions in a row	Fraction of subjects	Number of subjects
EU	No	No	6%	5
REU*	Yes except FOSD	No	} 23%	18
DDM*	Yes	No		
CSC	Yes except FOSD	Yes except FOSD	61%	49
Other (1)	Yes in FOSD	Yes in EASY and/or HARD	6%	5
Other (2)	No	Yes in HARD	4%	3
<i>Total</i>			100%	80

## 4.2 Flip of the Costly Coin

A much smaller but still sizable fraction of the subjects (29%) choose the option to flip a costly coin. Similar to the behavior observed in Parts I and III, the coin was chosen virtually only for HARD questions: 24% of subjects selected it in HARD questions, while only 8% did so in EASY and 4% in FOSD questions (Fisher exact test,  $p = 0.008$  for HARD vs. EASY;  $p = 0.0004$  for HARD vs. FOSD; and  $p = 0.495$  for EASY vs. FOSD). Those who choose the coin tend to do so frequently: 42% of those who ever chose it did so in at least two out of four HARD questions.

<sup>13</sup>These models also predict the observed correlation of stochastic choice in the two parts: intuitively, it is the same desire to hedge that applies in both cases (even though it may not necessarily lead to the same behavior because the agent does not know the number of repetitions in Part I).

Table 2 presents a regression analysis in which the indicator for stochastic choice in Part I or Part III is regressed on the indicator for choosing a coin in the *same* question while clustering standard errors by subject. The results indicate that subjects who choose the coin are significantly more likely to report inconsistent answers in both parts.

The choice of the costly coin has implications for the theoretical models. According to REU (and not only REU\*), subjects should never choose the costly coin because they follow Expected Utility once they face a question. Models of bounded rationality and of deliberate randomization are instead compatible with both choosing and not choosing the coin.<sup>14</sup> This allows us to further refine the classification in Table 3: of the 18 subjects classified as either REU\* or DDM\*, 3 chose the coin at least once and thus cannot belong to the former group.

### 4.3 Stochastic Choice and Expected Values

Table 4 summarizes the results of a regression analysis in which the indicator for stochastic choice in Part I or in Part III is regressed on dummy variables for the difficulty of questions and on question-type specific absolute differences in individual expected utilities. For the latter, we consider risk neutral – thus indicating the difference in expected values – as well as CRRA and CARA functional forms. For these, the individual-specific parameters were estimated using the answers to the investment task in Part II (question 1). These estimations are meaningful for subjects who reported a strictly positive risky investment, which leaves us with a sample of 78 subjects. Of these, 30 subjects are either risk-neutral or risk-loving because they invested their entire budget. Regressions 1, 3, 5, 7, 9 and 11 include these 30 subjects, treating them as risk-neutral, while Regressions 2, 4, 6, 8, 10 and 12 exclude them as well as the subjects who violated Expected Utility as measured by Allais-type behavior in Part IV of the experiment.

The analysis shows that the similarity in expected values or expected utilities does *not* account for the full variation in the stochastic behavior between HARD and other types of questions in either Part I or Part III: the coefficient of the dummy variable HARD is both large and highly significant in *all* regressions, even when we control for the differences in expected values or in subject-specific expected utilities. This holds when we consider either all subjects or those who do not exhibit Allais-type behavior and who are risk-averse. In some specifications (Regressions 3, 5 and 9), the differences in expected utilities significantly affect stochastic behavior within HARD questions, but this effect is not robust with respect to the utility specification or to

---

<sup>14</sup>With models of deliberate randomization, on the one hand, the agent may find an advantage in using a randomization device, especially if she does not know if questions will be repeated; on the other hand, the agent may be able to randomize herself, and thus not need to pay for the coin.

TABLE 4 Determinants of Stochastic Choice  
Stochastic Choice with Distant Repetitions (Part I)

Utility specification	Risk Neutral		CRRA		CARA	
	(1)	(2)	(3)	(4)	(5)	(6)
Dummy for EASY	-0.010 (0.06)	-0.045 (0.087)	0.0002 (0.042)	-0.04 (0.07)	-0.018 (0.05)	-0.03 (0.07)
Dummy for HARD	0.49** (0.04)	0.39** (0.07)	0.48** (0.03)	0.43** (0.05)	0.49** (0.03)	0.49** (0.06)
Diff in EU FOSD	-0.0003 (0.0005)	-0.0005 (0.0007)	-0.0003 (0.0006)	-0.001 (0.002)	-0.043 (0.13)	-0.59 (1.76)
Diff in EU EASY	-0.0002 (0.001)	1.13 e <sup>-17</sup> (0.002)	-0.0006 (0.001)	1.69 e <sup>-18</sup> (0.004)	0.048 (0.23)	2 e <sup>-16</sup> (0.49)
Diff in EU HARD	-0.012 (0.006)	0.008 (0.0009)	-0.023** (0.007)	0.0056 (0.03)	-3.37** (1.03)	-7.73 (4.91)
Const	0.027 (0.027)	0.045 (0.04)	0.024 (0.025)	0.04 (0.038)	0.02 (0.024)	0.03 (0.04)
# of obs	780	350	780	350	780	350
# of subjects	78	35	78	35	78	35
Overall R-sq	0.3089	0.32	0.3164	0.3183	0.3156	0.3228
Set of subjects	All	Risk A&Not Allais	All	Risk A&Not Allais	All	Risk A&Not Allais
Sample of questions	All	All	All	All	All	All

Stochastic Choice with Repetitions in a row (Part III)

Utility specification	Risk Neutral		CRRA		CARA	
	(7)	(8)	(9)	(10)	(11)	(12)
Dummy HARD	0.40** (0.07)	0.38** (0.11)	0.47** (0.05)	0.43** (0.09)	0.45** (0.06)	0.38** (0.09)
Diff in EU EASY	0.001 (0.002)	0.0008 (0.002)	0.0017 (0.002)	0.001 (0.006)	0.29 (0.37)	0.18 (0.75)
Diff in EU HARD	0.004 (0.007)	0.02 (0.01)	-0.026** (0.008)	0.027 (0.04)	-2.14 (1.28)	10.01 (5.46)
Const	-0.003 (0.07)	0.02 (0.10)	-0.0007 (0.05)	0.025 (0.09)	-0.003 (0.06)	0.03 (0.09)
# of obs	468	210	468	210	468	210
# of subjects	78	35	78	35	78	35
Overall R-sq	0.1593	0.1967	0.1822	0.1905	0.168	0.1955
Set of subjects	All	Risk A&Not Allais	All	Risk A&Not Allais	All	Risk A&Not Allais
Sample of questions	EASY & HARD	EASY & HARD	EASY & HARD	EASY & HARD	EASY & HARD	EASY & HARD

*Notes:* Random-effects GLS regressions with standard errors clustered at the subject level. Each observation corresponds to the behavior of one subject in one of the ten (seven) questions subjects faced repeatedly in Part I with Distant Repetitions (in Part III with Repetitions in a Row). For Risk Neutral utilities, we used the identity function. For CRRA, we estimated the parameter  $\rho > 0$  of the utility function  $u(x) = \frac{x^{1-\rho}}{1-\rho}$ . For CARA, we estimated the parameter  $\beta > 0$  of the utility function  $u(x) = 1 - e^{-\beta x}$ . These estimates are based on the subject-specific answer to the risky investment task in Part II (question 1). (Two subjects are excluded because they reported an investment of zero and thus the parameters cannot be estimated.) Dummy HARD denotes a dummy variable that takes the value 1 if a question is in the HARD category, 0 otherwise. The Diff in EU FOSD (EASY, HARD) variable is the interaction between the absolute difference in expected utilities of lotteries and the dummy for FOSD (EASY, HARD). Risk A & Not Allais indicates the subset of subjects who are risk-averse according to the risky investment task in Part II (question 1) and who do not violate the principles of Expected Utility in their answers in Part IV. \*\* indicates significance at the 5% level.

the subset of considered subjects. In particular, it disappears if we focus on risk-averse subjects who do not exhibit Allais-type behavior (the group for which we have reasonable measures of the utility). In Online Appendix A, we present additional regressions that show that these results are robust to different specifications.

Our finding is not compatible with the many models of stochastic choice that imply a relation between stochastic choice and expected utility difference. These include the DDM (and not only DDM\*) and many other models of bounded rationality, e.g., models of trembles like Harless and Camerer (1994).

Our results above may appear incompatible with the finding documented in many experiments according to which stochastic choice is more prevalent when the difference in expected values or expected utilities of lotteries is smaller (e.g., Mosteller and Nogee 1951; Kable and Glimcher 2007). To reconcile our findings with this literature, note that these experiments have predominantly focused on the differences in expected values and/or utilities. Because these tend to be smaller for “hard” questions, the two effects are conflated. By contrast, in our experiment, questions were designed in a way that allows us to separate these two forces.

Before proceeding, we briefly mention two other possible interpretations of stochastic choice. First, subjects could be simply indifferent between the available options. This implies that stochastic choice should be present only when these have very similar expected utilities. Moreover, because indifference is such a knife-edge case, stochastic choice must be limited. As we have seen, neither implication is true in our data. Alternatively, stochasticity of choice could be due to ‘preference discovery:’ subjects could be forming their preferences in the course of the experiment. This implies that the stochasticity of choice should become less frequent as we proceed in the experiment. This is also not supported in our data, as 71% of subjects report stochastic choice in Part III of the experiment.

## 4.4 Questionnaires

One way to gain some insights into the motivation behind subjects’ behavior is to ask them directly. We have done so at the end of the experiment in a non-incentivized questionnaire in which we asked “In Part III of the experiment each question was asked to you three times. If you chose different options, could you please tell us why did you do it? (Please elaborate).”<sup>15</sup>

As the question had an open answer, we used the following procedure to analyze the responses. Two undergraduate students from Columbia University were hired

---

<sup>15</sup>The questionnaire also asked other questions (the question above was the second one; see Online Appendix B for a complete list). We focus on this question as it allows us to discuss the motivations for stochastic choice.

as research assistants to code them. Each was given written guidelines instructing him or her to classify subjects into one of six categories, which are listed in the first column of Table 5.<sup>16</sup> These research assistants were not privy to the research questions posed in this paper. Overall, we observe a very high consensus among the two research assistants on how to classify subjects: they disagree only for 4 out of 79 questionnaires (one subject did not fill out the questionnaire at all), and the correlation in their classifications was 0.994 ( $p < 0.01$ ). In the analysis below, we focus on 75 subjects for which there were no disagreements (none of our qualitative results would change if we included subjects with disagreements following any of the two coders; see Online Appendix H). Furthermore, for subjects classified as deliberately reporting stochastic answers, we gave the assistants a list of possible reasons for this choice – arguments invoking hedging, diversification, riskiness of different options, optimizing ones’ earnings, delegation of decision and the difficulty of choice – and asked them to select the reasons, possibly more than one, closest to the subject’s answer. For this task, we observe more variability in the coders’ responses; the correlation between coders was 0.62 ( $p < 0.01$ ). However, the frequencies of motives indicated by two coders were similar, as we will see below.

Table 5 reports the classification of subjects into the six categories, focusing on all subjects as well as on those who exhibited stochastic choice in Part III.<sup>17</sup> The first check is for the consistency between the answers and actual behavior: approximately two-third of subjects reported giving inconsistent answers in Part III, and among them, all but one actually did so (48 out of 49 subjects); conversely, among those who actually reported inconsistent answers in Part III, only two subjects reported that they never did, and two did not answer.

The vast majority of subjects who reported giving inconsistent answers in Part III indicate that they did so *deliberately*: only one subject reported giving inconsistent answers because he/she was indifferent between two options, 12% reported changing their mind, and the remaining 79% said that they did it on purpose. To illustrate, here are two randomly chosen answers assigned to the latter group, category 6: “I wanted to increase my chances of getting more tokens with varied options” and “some options were not clearly better, would give each a try.”

We now turn to the reasons given for reporting inconsistent answers, conditional on choosing to do so deliberately. In this exercise, subjects could belong to more than one group, as they could give more than one reason. The vast majority of subjects appeal to hedging, diversification and differences in the riskiness/safeness of the presented options: 71% or 69% depending on the coder. The second most popular

---

<sup>16</sup>Because all of our experiments were conducted in California, this eliminates the possibility that one of the research assistants participated as a subject in any of our experimental sessions. The complete instructions for the coders are presented in Online Appendix F.

<sup>17</sup>Online Appendix G includes a list of randomly chosen quotes from questionnaires for each of the categories.

TABLE 5 Reasons for Stochastic Choice Reported in the Final Questionnaire

Category	Explanation	All subjects	Subjects with SC in Part III
1	No answer	9 (12%)	2 (4%)
2	No SC in Part III	17 (23%)	2 (4%)
3	SC b/c they changed their mind about which option is better	6 (8%)	6 (12%)
4	SC unintentionally, by mistake	0 (0%)	0 (0%)
5	SC b/c they were indifferent between the two options	1 (1%)	1 (2%)
6	SC b/c they deliberately chose to do so	42 (56%)	41 (79%)
Total number of subjects		75	52

*Notes:* Categorization of subjects according to their answer to the question: ‘In Part III of the experiment [with repetitions in a row] each question was asked to you three times. If you chose different options, could you please tell us why did you do it? (Please elaborate).’ This question was asked at the very end of the experiment and was not incentivized. It was an open question. The answers were coded by two independent research assistants who were asked to assign each subject to one of the six mutually exclusive categories listed in the table. (One subject refused to fill out the questionnaire.) In this table, we report the classification of the 75 subjects that both coders classified in the same way. The coders disagreed about how to classify 4 subjects (see Online Appendix H).

explanation mentions some optimization of the earnings: 40% or 48%. Finally, some subjects mention a desire to delegate the decision: 10% or 12%. All other explanations were recorded fewer than 7% of the times.

## 4.5 Response Time

In this section, we present the response time analysis. Among other things, this provides a direct test of the extra assumption we imposed on the DDM to make predictions regarding the case of repetitions in a row.

We begin by noting that we find no statistical difference in the distribution of response times between participants who reported stochastic answers for HARD questions and those who did not in Parts I or III (with the exception of HARD1 in Part I, in which subjects who gave inconsistent answers took significantly longer).<sup>18</sup> These

<sup>18</sup>For each question, we run a separate Random-effects GLS regression in which we regress the observed response times on the dummy variable that indicates whether or not a subject gave inconsistent answers for the same question, clustering standard errors at the individual level. The estimated coefficients on the dummy variables in each regression are not significantly different from zero at the 5% level: for Part I,  $p = 0.087$  in HARD1,  $p = 0.082$  in HARD2,  $p = 0.264$  in HARD3

findings refute the hypothesis that subjects who reported inconsistent answers in either part of the experiment did so because they did not spend time thinking about their choice.

Table 6 reports summary statistics of reaction-time data for each type of question in seconds. In Online Appendix A (Table A3), we report the data for each question separately. Several patterns emerge from this and from the regression analysis (see Online Appendix A for details). First, subjects take longer to make a decision for HARD than for FOSD or EASY questions the first time they encounter a question in both Parts I and III ( $p < 0.05$  in HARD vs. EASY and HARD vs. FOSD in each part). Second, the first repetition takes longer than the second ( $p < 0.05$  in all pairwise comparisons), while response times in later repetitions are often not statistically different from each other at the 5% level.

TABLE 6 Mean (Median) Response Time for each Type of Question, in seconds

Repetition	FOSD		EASY		HARD	
	Part I	Part III	Part I	Part III	Part I	Part III
First	6.5 (5)	2.9 (3)	4.8 (4)	3.6 (3)	14.3 (10)	10 (7)
Second	4.3 (3)	1.7 (1)	3.9 (3)	1.9 (2)	9.6 (7)	2.6 (2)
Third	3.4 (3)	1.7 (1)	3.4 (3)	1.6 (2)	7.2 (6)	2.4 (2)

*Notes:* Category FOSD reports response times for questions FOSD1, FOSD2 and FOSD3. Category EASY reports response times for questions EASY1, EASY2 and EASY3. Category HARD reports response times for questions HARD1, HARD2, HARD3 and HARD4.

An important finding for our analysis is a comparison between response times in the repetitions of HARD questions in Parts I and III. In Part I, subjects take a relatively long time to answer each of the three repetitions. Instead, in Part III, they take a relatively long time only the first time they encounter a question – a median of 7 seconds – while the next two repetitions are answered *almost immediately* – with a median of 2 seconds and with no difference between the repetitions (HARD questions in Part III:  $p < 0.01$  when comparing first vs. second repetitions and  $p = 0.49$  when comparing second vs. third repetitions). Subjects take significantly longer to answer

and  $p = 0.876$  in HARD4; and for Part III,  $p = 0.274$  in HARD1,  $p = 0.834$  in HARD2,  $p = 0.826$  in HARD3 and  $p = 0.179$  in HARD4. Throughout this section, we will use a regression analysis of this type to compare response times between different types of questions, different repetitions of the questions and different subsets of subjects. When we say that there is a significant difference in response times between two groups (or two types of questions or two repetitions of questions), we mean that the estimated coefficient on the dummy variable that indicates one of the groups (types or repetitions) is significantly different from zero at the 5% level. In all such comparisons, we report the p-value associated with the dummy variable in the regression. Tables A4 and A5 in Online Appendix A present a detailed summary of these regressions and summary statistics for the response times of subjects that reported consistent and inconsistent answers in HARD questions in Parts I and III.

the second or the third repetition of the HARD questions in Part I than in Part III ( $p < 0.01$  in both tests). These results hold also when we focus on the subset of subjects who gave inconsistent answers in Part III (see Online Appendix A).

To interpret these numbers, we take as a benchmark the response times in the second and third repetitions of the FOSD question in Part III: these are naturally the easiest questions in the experiment, in which we expect subjects to recognize the dominant option in the first repetition and to simply implement the same choice in later repetitions without much thought (indeed, all subjects choose the dominant option all of the times). Response time in these repetitions can be seen as the time necessary to move the mouse across the screen to make a choice.<sup>19</sup> We find that response times in the second and third repetitions of HARD questions in Part III are statistically *indistinguishable* from those in the second and third repetitions of FOSD questions in Part III.<sup>20</sup>

The analysis of response times is consistent with the view that in-a-row repetitions of the same question are not treated as entirely new questions and are different from distant repetitions. This is in line with the assumptions that we have made to extend existing models to our setup in Section 2. These findings also allow us to directly test the assumption made to extend the DDM model to the case of repetitions in a row (DDM\*). Within that model, the very short response times observed in the second and third consecutive repetitions are consistent with the assumption that subjects do not gather more information when they observe the same question asked repeatedly in a row.

## 4.6 Relation with Risk Attitudes and Violations of EU

We conclude by studying the relation between stochastic choice, the attitudes towards risk and compound lotteries, measured in Part II, and Allais-type questions, measured in Part IV.

Looking at Part II, we find that the majority of subjects (63%) are risk-averse and a large majority either strictly dislike compound lotteries (43% of subjects) or are neutral (54% of subjects). Risk averse subjects are those who invest fewer than

---

<sup>19</sup>Our software recorded response times approximated to seconds (and not milliseconds). Moreover, 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. A response time of 2 seconds is thus consistent with the time necessary to move the mouse. This is in contrast with other experiments, especially in psychology and neuroscience (e.g., De Martino et al. 2006; Levy et al. 2013), where the choice can be made much more quickly by pressing a physical button on which subjects are resting their fingers.

<sup>20</sup>Regression analysis confirms that there is no statistical difference (at the 5% level) between the median decision time subjects took in the second and third repetitions of FOSD versus HARD questions in Part III ( $p > 0.05$ ).

TABLE 7 Correlation between Stochastic Choice and Preferences under Risk

	SC Distant Rep. (Part I)		SC Rep. in row (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**	( $p = 0.0208$ )	0.14	( $p > 0.1$ )	0.15	( $p > 0.1$ )
Compound Investment	-0.08	( $p > 0.1$ )	-0.05	( $p > 0.1$ )	0.01	( $p > 0.1$ )
Compound (Indicator)	0.12	( $p > 0.1$ )	0.10	( $p > 0.1$ )	-0.04	( $p > 0.1$ )
Allais-like behavior	0.19*	( $p = 0.0985$ )	0.22**	( $p = 0.0451$ )	0.16	( $p > 0.1$ )

*Notes:* Risky Investment is a variable that records how many points (out of 100 available) a subject chose to invest in a risky project that pays back 2.5 times the invested amount with probability 50%, and zero otherwise (question 1, Part II). The indicator for Risk-aversion is a dichotomous variable that takes value one if the subject invested fewer than the maximum number of points, and zero otherwise. The compound Investment variable records how many points (out of the 100 available) a subject chose to invest in a risky project that has the same odds as the one described in the Risky Investment except that the success of the investment was determined by a compound lottery that reduces to a 50% chance of success (question 2, Part II). The indicator for Compound-aversion is a dichotomous variable that takes value one if the subject invested less in the compound lottery case than in the simple lottery case, and zero otherwise. Allais-like behavior is an indicator that the subject’s answers in Part IV of the experiment violated Expected Utility principles in either the Common Consequences question or in the Common Ratio question. \* and \*\* denote significance at the 10% and 5% levels, respectively.

the maximum number of tokens in the first investment task in Part II. We observe an average investment of 69.6 tokens, in line with typical results (Dreber and Hoffman 2007, Charness and Gneezy 2010, Langer and Weber 2008). The attitude towards compound lotteries is measured by comparing the investments in the two questions in Part II: a compound averse (neutral, loving) subject invests strictly less (equal, more) in the compound lottery question than in the risk question. A Wilcoxon matched-pairs signed-ranks test detects a statistical difference between the investment in the two questions at the 1% level ( $z = 4.963$  and  $p < 0.01$ ).

Looking at Part IV, 25% of subjects violate expected utility principles according to either the *Common Consequences* effect and/or the *Common Ratio* effect, the so-called Allais paradoxes, which are the most widely used tests in the literature. This proportion is in line with other experiments that use similar incentive structures (e.g., Camerer 1989b, Conlisk 1989, Burke et al. 1996, Fan 2002, Huck and Müller 2012).

Table 7 reports the Spearman correlations between the tendency to report stochastic choice and attitudes towards risk, compound lotteries and Allais-type behavior. While the former two are, with one exception, not related to the stochastic behavior in Parts I and III, we find a positive and significant correlation between these and the tendency to exhibit Allais-like violations of expected utility:  $\text{corr} = 0.19$  ( $p = 0.0985$ ) and  $\text{corr} = 0.22$  ( $p = 0.0451$ ) for Parts I and III.

We conclude by noting that the documented relation between stochastic choice and Allais-type behavior is one of the predictions of the CSC model (see Cerreia-Vioglio et al. 2015a). Intuitively, in this model, the stochasticity of choice emerges because the agent has a desire to “hedge” between options, which is due to the multiplicity of utilities (the set  $\mathcal{W}$ ). This same multiplicity generates an Allais-like behavior with an attraction towards degenerate lotteries: in this model, when the agent has multiple utilities, she is “pessimistic” in her evaluation of non-degenerate lotteries, while degenerate ones are not affected (as their certainty equivalent is the same for any utility). This generates an Allais-like behavior. Thus, subjects with a larger set of utilities should be (weakly) more prone to both stochastic choice and Allais-type behavior – generating the relation documented in our data.<sup>21</sup>

## 5. Robustness: Three Short Experiments

In this section, we report the results of three short experiments designed to investigate the robustness of our findings. Short Experiment 1 asks questions similar to those of our main experiment (a choice between lotteries) but with much higher stakes and many fewer questions. Short Experiments 2 and 3 follow a similar methodology but ask questions in different domains: time preferences and social preferences. The complete instructions for all short experiments are presented in Online Appendix B.

*Short Experiment 1: Lotteries, High Stakes* The experiment was conducted at the California Institute of Technology with 26 undergraduate students using pen and paper. After the instructions were read aloud to the subjects, they were asked to answer question HARD2 (see Table 1) three times in a row, each time on a separate sheet of paper. As in Part III of our main experiment, subjects were explicitly told about this repetition. At the end of the experiment, one participant rolled a die to determine which of the three repetitions would determine the payments. To further increase the stakes, we used a much higher conversion rate: 2 tokens = \$1 (instead of 20 tokens = \$1, as in the main experiment). This change made the stakes unusually high (winnings up to \$45 with probability 0.5), considering that the entire experiment,

---

<sup>21</sup>More precisely, in the CSC model, the larger the set of utilities  $\mathcal{W}$ , the (weakly) larger the tendency to exhibit both behaviors. Moreover, in this model, subjects who exhibit stochastic choice must exhibit Allais-like behavior for some range of prizes; conversely, if  $|\mathcal{W}| < \infty$  and the agent exhibits Allais-like behavior, then she should also exhibit stochastic choice, again in some range of prizes (see Cerreia-Vioglio et al. 2015b.) However, note that this does not imply that these subjects who exhibit stochastic choice must exhibit Allais-like behavior for *any* range of prizes or questions: first, because in the CSC model, subjects can abide by Expected Utility in some range of prizes and not in others; second, because specific Allais questions may not be properly calibrated to capture non-Expected Utility in a given range, a well-known issue with small incentives (see Cerreia-Vioglio et al. 2015a and Cerreia-Vioglio et al. 2015b).

including the instruction period, lasted 15 minutes in total.

Short Experiment 1 has several differences from our main experiment: (a) it involves extremely high stakes; (b) only one question was asked, three times, sharply reducing the cognitive load; (c) it is performed in isolation from Part I of the main treatment (in which subjects were exposed to the option of flipping a coin between available lotteries); (d) it involves a different population, with high average computational skills;<sup>22</sup> and (e) the randomization that determined which repetition matters for payments was performed by a participant, eliminating trust issues.

We observe that 50% of subjects (13 out of 26) chose different lotteries in the repetition of the same question. This fraction is not significantly different, at the 5% level, from the 45% of subjects who reported different answers in the same HARD2 question in Part III of the main experiment (Fisher exact test,  $p = 0.821$ ).

*Short Experiment 2: Time Preferences* Short Experiment 2 has a very similar design as Short Experiment 1, but instead of lotteries, the choice is between bundles of payments delivered at different dates. Subjects face the choice between

**Option 1** \$12 today and \$10 in two weeks;

**Option 2** \$3 today and \$22 in two weeks.

The two options were designed to keep the transaction cost constant, as both require subjects to come back to the lab and pick up the second portion of their payment two weeks later. Except for the difference in the objects of choice, Short Experiment 2 was identical to Short Experiment 1: subjects were asked to choose between the options above three times in a row after being explicitly told that the question would be repeated three times. This was the only task. At the end of the experiment, one of the repetitions was randomly chosen to determine the final payments. Short Experiment 2 was conducted at the Experimental Social Science Laboratory (ESSL) at University of California at Irvine with 47 undergraduate students.<sup>23</sup>

We find that 34% of subjects (16 out of 47 participants) reported different answers in the three repetitions of the question. While smaller (at the 5% level) than the

---

<sup>22</sup>Undergraduate students at Caltech have very high average computational skills (SAT scores) and receive substantial training in math and probabilities. The advantage of confirming our results with this population is that it shows that the desire to randomize is present also for subjects who understand probabilities and distributions. This should eliminate the concern that the desire to randomize was found because subjects were asked to make choices with objects that they were deeply unfamiliar with.

<sup>23</sup>Short Experiments 2 and 3 were conducted at the end of experimental sessions in which the subjects participated in an unrelated experiment. Because these short experiments lasted less than 15 minutes and were not the main task for which students were invited to the lab, we paid only a fraction of students for this additional task (6 out of 47 students were randomly selected to receive payments for Short Experiment 2).

fraction of subjects who reported different answers in Short Experiment 1 (Fisher exact test,  $p = 0.045$ ), this still represents a sizable fraction of the population, similar to that observed for some HARD questions in our main experiment.

*Short Experiment 3: Social Preferences* Short Experiment 3 follows an identical design, but subjects were asked to choose between vectors of monetary allocations to themselves and to two other subjects. They were asked to choose between

**Option 1** You get \$13    Person 1 gets \$13    Person 2 gets \$13;

**Option 2** You get \$17    Person 1 gets \$27    Person 2 gets \$1.

The roles of Person 1 and Person 2 were randomly allocated to two other participants, keeping their identities secret. The two options above capture the well-known equality-efficiency trade-off: Option 1 gives an equal allocation but a lower total surplus (\$39); Option 2 gives a higher surplus (\$45) and a higher payment to the decision-maker, but members have very unequal shares. We used the exact same amounts as in Bolton and Ockenfels (2006), except we used US dollars instead of DM.

The structure of Short Experiment 3 was identical to the structure of Short Experiment 2: subjects had to choose one of the two presented options three times in a row and were told in advance of the repetitions. Short Experiment 3 was conducted at ESSL at University of California at Irvine with 57 undergraduate students, and no student participated in both Short Experiment 2 and Short Experiment 3.<sup>24</sup>

We find that 42% of subjects (24 out of 57 participants) reported stochastic answers. This fraction is not significantly different (at a 5% level) from the fraction of students who reported stochastic answers in Short Experiment 2 (Fisher exact test,  $p = 0.578$ ) or those who reported stochastic answers in Short Experiment 1 (Fisher exact test,  $p = 0.096$ ).

## 6. Conclusion

In this paper, we study experimentally the origin of stochastic choice. We ask subjects to choose from the same set of options multiple times under two conditions: with repetitions of the same question distant from each other, and with repetitions in a row. We find that subjects exhibit stochastic choice in both cases, with stochasticity highly correlated. In three robustness tests, we confirm that a similar behavior holds also for high stakes, different subject pools, and questions in different domains. Overall,

---

<sup>24</sup>Similar to Short Experiment 2, in Short Experiment 3, twelve out of 57 students were randomly selected to receive payments for this experiment.

our results indicate that the desire to randomize plays an important role in driving stochastic choice.

## 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.
- APESTEGUIA, J. AND M. A. BALLESTER (2015): “Well-Defined Stochastic Choice Models: The Case of Risk and Time Preferences,” Manuscript, Universitat Pompeu Fabra.
- BALLINGER, T. P. AND N. T. WILCOX (1997): “Decisions, error and heterogeneity,” *The Economic Journal*, 107, 1090–1105.
- BARBERÁ, S. AND P. K. PATTANAİK (1986): “Falmagne and the rationalizability of stochastic choices in terms of random orderings,” *Econometrica*, 707–715.
- BECKER, G. M., M. H. DEGROOT, AND J. MARSCHAK (1963a): “An experimental study of some stochastic models for wagers,” *Behavioral Science*, 8, 199–202.
- (1963b): “Stochastic models of choice behavior,” *Behavioral science*, 8, 41–55.
- BOLTON, G. E. AND A. OCKENFELS (2006): “Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: comment,” *The American economic review*, 1906–1911.
- 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.
- BUSEMEYER, J. R. AND J. T. TOWNSEND (1993): “Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment.” *Psychological review*, 100, 432.
- CAMERER, C. F. (1989a): “Does the Basketball Market Believe in the ‘Hot Hand’?” *American Economic Review*, 79, pp. 1257–1261.
- (1989b): “An experimental test of several generalized utility theories,” *Journal of Risk and uncertainty*, 2, 61–104.
- (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 (2015a): “Cautious Expected Utility and the Certainty Effect,” *Econometrica*, 83, 693–728.
- CERREIA-VIOGLIO, S., D. DILLENBERGER, P. ORTOLEVA, AND G. RIELLA (2015b): “Deliberately Stochastic,” Manuscript 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,” *American Economic Review*, 79, 392–407.
- DE MARTINO, B., D. KUMARAN, B. SEYMOUR, AND R. J. DOLAN (2006): “Frames, biases, and rational decision-making in the human brain,” *Science*, 313, 684–687.
- DREBER, A. AND M. HOFFMAN (2007): “2D: 4D and Risk Aversion: Evidence that the Gender Gap in Preferences is Partly Biological,” Manuscript, Stockholm School of Economics.
- DWENGER, N., D. KÜBLER, AND G. WEIZSÄCKER (2013): “Flipping a Coin: Theory and Evidence,” Manuscript 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,” Manuscript 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. IJIMA, AND T. STRZALECKI (2015): “Stochastic Choice and Revealed Perturbed Utility,” *Econometrica* (forthcoming).
- FUDENBERG, D. AND T. STRZALECKI (2015): “Recursive Logit with Choice Aversion,” *Econometrica*, 83, 651–691.
- GNEEZY, U. AND J. POTTERS (1997): “An experiment on risk taking and evaluation periods,” *The Quarterly Journal of Economics*, 112, 631–645.
- GUL, F., P. NATENZON, AND W. PESENDORFER (2014): “Random choice as behavioral optimization,” *Econometrica*, 82, 1873–1912.
- GUL, F. AND W. PESENDORFER (2006): “Random Expected Utility,” *Econometrica*, 74, 121–146.
- 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.
- HENDERSON, V., D. HOBSON, AND A. S. TSE (2014): “Randomized strategies and prospect theory in a dynamic context,” Available at SSRN 2531457.

- HEY, J. (2001): “Does repetition improve consistency?” *Experimental economics*, 4, 5–54.
- HEY, J. AND C. ORME (1994): “Investigating generalizations of expected utility theory using experimental data,” *Econometrica*, 1291–1326.
- HEY, J. D. AND E. CARBONE (1995): “Stochastic choice with deterministic preferences: An experimental investigation,” *Economics Letters*, 47, 161–167.
- 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.
- JOHNSON, E. J. AND R. RATCLIFF (2013): “Computational and Process Models of Decision Making in Psychology and Behavioral Economics,” in *Neuroeconomics: Decision Making and the Brain*, ed. by P. W. Glimcher and E. Fehr, Academic Press, 35–48.
- KABLE, J. W. AND P. W. GLIMCHER (2007): “The neural correlates of subjective value during intertemporal choice,” *Nature neuroscience*, 10, 1625–1633.
- 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.
- LEVY, D. J., A. C. THAVIKULWAT, AND P. W. GLIMCHER (2013): “State dependent valuation: the effect of deprivation on risk preferences,” *PloS one*, 8, e53978.
- LOOMES, G. AND R. SUGDEN (1995): “Incorporating a stochastic element into decision theories,” *European Economic Review*, 39, 641–648.
- LUCE, R. D. (1958): “A probabilistic theory of utility,” *Econometrica*, 193–224.
- (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.
- MOSTELLER, F. AND P. NOGEE (1951): “An experimental measurement of utility,” *Journal of Political Economy*, 59, 371–404.
- NATENZON, P. (2015): “Random choice and learning,” Manuscript, Washington University.
- RATCLIFF, R. (1978): “A theory of memory retrieval.” *Psychological review*, 85, 59.

- RATCLIFF, R. AND G. MCKOON (2008): “The diffusion decision model: theory and data for two-choice decision tasks,” *Neural computation*, 20, 873–922.
- 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.
- WOODFORD, M. (2014): “Stochastic Choice: An Optimizing Neuroeconomic Model,” *American Economic Review: Papers & Proceedings*, 104, 495–500.
- WU, G. AND R. GONZALEZ (1996): “Curvature of the probability weighting function,” *Management science*, 42, 1676–1690.