How Real is Hypothetical? A High-Stakes Test of the Allais Paradox[*]

Uri Gneezy[†] Yoram Halevy[‡] Brian Hall [§] Theo Offerman [¶]
Jeroen van de Ven [¶]

September 19, 2024

Abstract

Researchers in behavioral and experimental economics often argue that only incentive-compatible mechanisms can elicit effort and truthful responses from participants. Others argue that participants make less-biased decisions when the stakes are sufficiently high. Are these claims correct? We investigate the change in behavior as incentives are scaled up in the Allais paradox, and document an *increase*, not decrease, in deviations from expected utility with higher stakes. We also find that if one needs to approximate participants' behavior in real high-stakes Allais (which are often too expensive to conduct), it is better to use hypothetically high stakes than real low stakes, as is typically the practice today.

Keywords: high stakes, real and hypothetical incentives, Allais paradox, Expected Utility
JEL Codes: C91, D81

[*]We are grateful to the staff at the Busara Center for Behavioral Economics and in particular Raphael Malenya for their support in implementing the experiments. The study received IRB approval from the Busara Centre/Strathmore University (# SE-ISERC1459/22). Funding from the Amsterdam Center of Behavioral Change (project numbers 201711200311 and 202205301205) and the Rady School of Management (chair account Gneezy) is gratefully acknowledged. Our paper benefited from the comments of seminar audiences at the University of Vienna, 2023 SAET meeting (Paris), the 2023 ESA world meeting (Lyon), 2024 MBEES meeting (Maastricht), BRIC X (Venice, 2024), Israeli Economic Association annual conference (2024), and the 8th Solomon Lew Conference on Behavioral Economics (Tel Aviv University, 2024). Billur Görgülü provided helpful comments. We are also grateful to Random.org for providing us with a non-profit license to use their API for their website.

[†]Rady School of Management, University of California San Diego

[‡]Department of Economics, University of Toronto; and the Bogen Family Department of Economics and the Federmann Center for the Study of Rationality, Hebrew University of Jerusalem

[§]Harvard Business School

[¶]University of Amsterdam

1 Introduction

How people behave in high-stakes versus low-stakes environments, particularly in terms of risk-taking, has long fascinated economists (e.g., Markowitz, 1952; Samuelson, 1963; Pratt, 1964; Arrow, 1971; Rabin, 2000). As behavior in experiments is often inconsistent with expected utility predictions, some economists argue that this inconsistency in experiments results from low stakes, and for sufficiently high stakes, people will make decisions that align more closely with the expected utility model. However, the experimental literature studying this argument is surprisingly small.

So what happens as the stakes rise? We systematically investigate this question by studying the behavior of participants confronted with a famous choice problem designed in 1953 by the French economist Maurice Allais—a problem that routinely generates deviations from expected utility (Allais, 1953). The experiment, which called into question the descriptive validity of expected utility, has motivated important contributions to behavioral economics, among them Kahneman and Tversky's (1979) prospect theory and an extensive literature in decision theory.

Allais asked decision-makers to consider the following two lottery sets, and in each case to choose between Options A and B:

Lottery set 1 (L1):

Option A:	\$X with certainty
Option B:	\$0 with probability 0.01 \$X with probability 0.89 \$5X with probability 0.1

Lottery set 2 (L2):

Option A:	\$0 with probability 0.89 \$X with probability 0.11
Option B:	\$0 with probability 0.90 \$5X with probability 0.1

Expected utility predicts that people would choose either A or B in both cases. Deviating from this choice pattern violates the independence axiom. In the first lottery set, A and B share a .89 probability of winning \$X, so according to the independence axiom, a decision-maker would ignore this consequence when comparing the two options and focus instead on what remains: the .11 probability of winning \$X (in A) compared with the .1 probability of winning \$5X and the .01 probability of winning \$0 (in B). This reasoning leads to choices of either A or B in both choice sets.

Research has shown, however, that participants in experiments confronted with the Allais problem often do not behave in a manner that is consistent with this prediction. Instead, they choose A in the first lottery and B in the second—a violation of the independence axiom. Allais attributed this behavior to the high stakes that he built into his experiment: X in his original design equaled 100 million old French francs, or approximately US $3.25 million today. He deliberately used that amount because he felt it was important for earnings to "have a large value relative to the player's wealth" (p. 526).[1] This, he argued, would encourage people to fixate on certainty in their assessment of possible winnings and therefore act in a way that was inconsistent with expected-utility theory.

Savage (1972) made a similar case, writing that many people "do not find the chance of winning a *very* large fortune in place of receiving a large fortune outright adequate compensation for even a small risk of being left in the status quo" (Savage, 1972, p. 102, emphasis in original).

Were Allais and Savage right? In this paper, we systematically investigate the effect of incentive size (X), and in particular the effect of very large incentives, on choices in the Allais paradox. Our design has two payment levels: "small incentives," which amount to standard laboratory payments for a couple of hours, and "large incentives," which are 100 times larger and amount to more than a month's income, with a chance to earn the equivalent of 4-5 months of income for the mean participant. To the best of our knowledge, we are the first to test the paradox with such high incentives. In a recent meta-analysis reanalyzing data collected in 81 experiments from 29 studies, Blavatskyy et al. (2022) found that incentivized tests of the Allais Paradox have used relatively small amounts of money, and they conclude that the paradox is "a fragile empirical finding" where the likelihood of observing Allais-type behavior reflects the details of the experimental design.

The lack of tests with high incentives is striking, because when incentives are low (as with standard lab incentives), people are unlikely to succumb to the certainty effect. In addition, with relatively small incentives, there is a higher chance of mistakes, which might result in violations of expected utility, but not because of the Allais reasoning. Going back to at least Ballinger and Wilcox (1997), researchers have pointed out that noise can generate expected-utility violations in the common-ratio version of the Allais paradox when paired-choice tasks are used. McGranaghan et al. (2024) find empirical support for this argument in a choice version but not in a valuation version of their experiment. The incentives they used were small: The maximum payoff was $54, and one out of 42 decisions was paid for every 1 out of 5 participants, with an average incentive compatible payment per participant being $1.51. Note that in our experiment all incentives are a few orders of magnitude higher, and this type of argument does not apply.

The design of our experiment also led us to an important methodological question

[1] Our translation. The original text reads: "Tel est en particulier le cas des choix entre des gains certains et des gain aléatoires, lorsque les gains ont une grande valeur par rapport à la fortune du joueur. Dans de tels cas, on peut mettre en évidence l'importance psychologique considerable que peut avoir, considéré en lui-même, l'avantage de la certitude." (Allais, 1953, p. 526,)

that extends beyond the Allais paradox: How useful are hypothetical responses? That is, if one needs to approximate participant behavior in real high-stakes experiments (which often would be prohibitively expensive to conduct), is it better to use hypothetical high stakes or real low stakes as is standard in experimental economics? The answer to this question is central to the ongoing debate in behavioral and experimental economics about the relative effectiveness of real versus hypothetical incentives.

Many researchers in the field argue that real incentives are necessary to motivate genuine effort and truthful responses from participants, especially to counteract biases arising from inattention or non-compliance with expected behavior (e.g., Plott, 1986; Smith, 1982, 1991; Svorenčík and Maas, 2016). But this approach has been challenged. Kahneman and Tversky (1979, p. 265) criticized it, for example, writing, "Laboratory experiments have been designed to obtain precise measures of utility and probability from actual choices, but these experimental studies typically involve contrived gambles for small stakes, and a large number of repetitions of very similar problems. These features of laboratory gambling complicate the interpretation of the results and restrict their generality. By default, the method of hypothetical choices emerges as the simplest procedure by which a large number of theoretical questions can be investigated. The use of the method relies on the assumption that people often know how they would behave in actual situations of choice, and on the further assumption that the subjects have no special reason to disguise their true preferences." Rubinstein (2013) concurs, and adds that response time data can provide a useful indication if decisions are deliberate or instinctive (and often a mistake).

Thaler (1986) addresses another issue with incentives in experiments. Summing up the argument against traditional experiments that real high incentives are necessary to motivate genuine effort and truthful responses, he writes, "If the stakes are large enough, people will get it right. This comment is usually offered as a rebuttal to a demonstration of embarrassing inconsistency found when groups of undergraduate students participate in experiments . . . but is also, of course, an empirical question. Do people tend to make better decisions when the stakes are high?"

From a modeling perspective, Allais-type behavior has motivated the lion's share of the generalizations of expected utility theory (as Rank Dependent Utility (Quiggin, 1982), Betweenness models (Chew, 1983; Dekel, 1986; Gul, 1991) and Cautious Expected Utility (Cerreia-Vioglio et al., 2015)). Though all models could accommodate the stakes effect, solid empirical support for the main behavioral pattern that has been motivating these models remains lacking over 70 years after Allais proposed his original thought experiment.

To test how incentives affect behavior in the Allais paradox using high incentives, we conducted experiments in Nairobi, Kenya, which is a low-income country. In a recent study, Enke et al. (2023) used the same laboratory in Nairobi to study how the level of incentives affects behavior in cognitive biases, finding that cognitive effort, as measured by response time, increases by 40% with very high stakes. Performance, on the other hand, improves very mildly or not at all as incentives increase, with the largest improvements due to a reduced reliance on intuition. That study examines how stakes influence performance on tasks with an objectively correct answer. In contrast, the current study explores *preferences* in situations where no objective correct answer exists.

This approach of running experiments in low-income countries has proved to be effective in other domains testing for stake effects, notably in the ultimatum game (Slonim and Roth, 1998; Cameron, 1999; Munier and Zaharia, 2002; Andersen et al., 2011). Another literature using this approach studies the effect of stake size on risk taking. Binswanger (1980) found increased risk aversion with stake size in a sample of low-income farmers in India, as did Kachelmeier and Shehata (1992) in China.

2 Stake-dependence in models of non-expected utility

Before describing the experimental design, we review some well-known models of non-expected utility. The main purpose of the analysis is to examine whether these models could accommodate changes in violations of the Independence Axiom as stakes vary.

Let X be a finite set of outcomes. Elements of X are denoted by x, y, z. A simple lottery is a finitely supported distribution over X, denoted by $p = \left(x_i, p^i\right)_{i=1}^{n} \in \Delta(X)$ where $x_i \in X$ and $\sum_{i=1}^{n} p^i = 1$. Lotteries are denoted by p, q, r. The DM's preferences $\succsim$ are defined on $\Delta(X)$. Assume throughout that preferences are complete, transitive and continuous. Let $\alpha \in [0,1]$ and define the *α-mixture* of p and r as the lottery that assigns probability $\alpha p(x) + (1-\alpha) r(x)$ to every outcome $x \in X$ and denote it by $\alpha p \oplus (1-\alpha) r$. The *Independence Axiom* requires that for all $p, q, r \in \Delta(X)$: $p \succsim q$ if and only if $\alpha p \oplus (1-\alpha) r \succsim \alpha q \oplus (1-\alpha) r$. Von Neumann and Morgenstern (1944) showed that $\succsim$ satisfy the Independence Axiom if and only if $\succsim$ has an *Expected Utility* representation: $U(p) = \sum_{x \in X} p(x) u(x)$ where $u : X \to \Re$ is increasing and unique up to a positive linear transformation.

In the Allais paradox, the decision maker chooses between $p_1(x) = (x, 1)$ and $p_2(x) = (5x, 0.1; x, .89; 0, .01)$, and between $q_1(x) = (x, .11; 0, .89)$ and $q_2(x) = (5x, .1; 0, .9)$. The Independence Axiom implies that $p_1(x) \succsim p_2(x)$ if and only if $q_1(x) \succsim q_2(x)$ since $p_1(x) = .11(x, 1) \oplus .89(x, 1)$ and $p_2(x) = .11\left(5x, \frac{10}{11}; 0, \frac{1}{11}\right) \oplus .89(x, 1)$ while $q_1(x) = .11(x, 1) \oplus .89(0, 1)$ and $q_2(x) = .11\left(5x, \frac{10}{11}; 0, \frac{1}{11}\right) \oplus .89(0, 1)$. Allais' conjectured behavior, however, was $p_1(x) \succ p_2(x)$ and $q_2(x) \succ q_1(x)$.

Figure 1 demonstrates graphically the issue of stake-dependence in the Allais paradox. The focus of the literature has been on the behavior in a Marschak-Machina triangle, for given prizes. However, when the stakes vary the lotteries evaluated change, and the question arises whether models of non-expected utility that were developed to accommodate violations of the Independence Axiom impose uniformity when stakes change? In the figure, the probabilities of winning the lowest prize (\$0) and the highest prize (\$5x) are depicted on the probabilities plane (for every lottery, the probability of winning the intermediate prize - x, is given by 1 minus the probabilities of the extreme prizes), and when x changes the Marschak-Machina triangle shifts. In Figure 1, the value of x is captured by the vertical axis. The figure depicts two triangles for two different levels of x, where the darker shaded triangle corresponds to a higher value of x.[2] As a result, the lotteries p_i and q_i for $i = 1, 2$ lie on different triangles for different values of x, though their projections on the probabilities plane coincide. As a result, stake dependence would imply that the ranking between the lotteries could depend on the stake-triangle to which they belong. We show below that all mainstream models of non-expected utility can accommodate dependency of Allais-type behavior on stake size. That is, the preference $p_1(x) \succ p_2(x)$ and $q_2(x) \succ q_1(x)$ could (and usually will) depend on x.

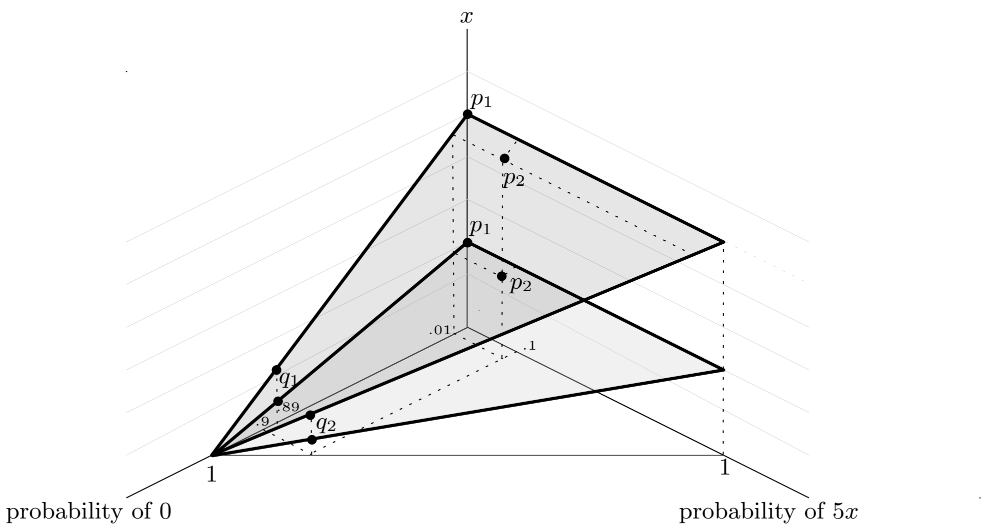

Figure 1: Stake-dependence in the Marschak-Machina Triangle

Rank Dependent Utility (Quiggin, 1982; adopted by Tversky and Kahneman, 1992 for Cumulative Prospect Theory) is considered by many the most common generalization of expected utility. Denote the weighting function by $\pi : [0, 1] \to [0, 1]$ and the value function by $v : X \to \Re$. A decision maker whose preferences are represented by RDU exhibits the Allais behavior if

$$v(x) > \pi(.1)\, v(5x) + [\pi(.99) - \pi(.1)]\, v(x) + [1 - \pi(.99)]\, v(0)$$

while

$$\pi(.1)\, v(5x) + [1 - \pi(.1)]\, v(0) > \pi(.11)\, v(x) + [1 - \pi(.11)]\, v(0)$$

[2]The triangles intersect at the lottery that pays \$0 with certainty.

Normalizing $v(0) = 0$ implies

$$\frac{1-(\pi(.99)-\pi(.1))}{\pi(.1)} = 1+\frac{1-\pi(.99)}{\pi(.1)} > \frac{v(5x)}{v(x)} > \frac{\pi(.11)}{\pi(.1)}$$

Obviously, satisfying these two inequalities depends on *both* the weighting function *and* the value function. It is natural to expect that $\frac{v(5x)}{v(x)}$ will be a decreasing function of x. This observation dates back to Markowitz (1952) and is captured by $v(\cdot)$ that exhibits increasing relative risk aversion (see Bouchouicha and Vieider, 2017), so the decision maker will exhibit the behavior typical for the Allais paradox for relatively high values of x, but not for low values of x.

Weighted Utility (Chew and MacCrimmon, 1979; Chew, 1983) is theoretically the mildest extension of expected utility that allows for Allais-type behavior. Here, the weighting is a function of the outcome $g : X \to \Re$, and the weight that is assigned to the utility $u(x_i)$ is: $\frac{p_i g(x_i)}{\sum_{x\in X} p(x)g(x)}$. For a decision maker to exhibit the Allais behavior

$$u(x) > \frac{.1g(5x)u(5x)+.89g(x)u(x)+.01g(0)u(0)}{.1g(5x)+.89g(x)+.01g(0)}$$

while
$$\frac{.1g(5x)u(5x)+.9g(0)u(0)}{.1g(5x)+.9g(0)} > \frac{.11g(x)u(x)+.89g(0)u(0)}{.11g(x)+.89g(0)}$$

Normalize $u(0) = 0$ and $g(0) = 1$ then

$$u(x)[.1g(5x)+.01] > .1g(5x)u(5x)$$

and
$$\frac{.1g(5x)u(5x)}{.1g(5x)+.9} > \frac{.11g(x)u(x)}{.11g(x)+.89}$$

implies

$$1+\frac{1}{10g(5x)} > \frac{u(5x)}{u(x)} >$$
$$\frac{.11g(x)[.1g(5x)+.9]}{.1g(5x)[.11g(x)+.89]} = \frac{11g(x)+89+99\frac{g(x)}{g(5x)}-89}{11g(x)+89} = 1+\frac{99\frac{g(x)}{g(5x)}-89}{11g(x)+89}$$

As $g(\cdot)$ is a function of x, Weighted Utility can accommodate cases where it is constant for low values of x (hence coincides with expected utility), but varies for high values of x (take, for example, $g(x) = 1.4$ and $g(5x) = 1.5$).

Disappointment Aversion (Gul, 1991) is a one-parameter generalization of expected utility. The support of every lottery is decomposed into elating outcomes – that are preferred to the lottery's certainty equivalent, and disappointing outcomes - to which the lottery's certainty equivalent is preferred. If the decision maker is disappointment averse, the functional representation assigns an extra (constant) weight to the objective probabilities of disappointing outcomes, and complementary lower (constant) weight to elating outcomes. If $p_1(x) \succ p_2(x)$ then outcomes $\{x, 5x\}$ in the support of $p_2(x)$ are elating, while $\{0\}$ is disappointing. The probability of elation is .99, while the probability of disappointment is .01.[3] Let $\beta > 0$ be the parameter which represents the decision maker's disappointment aversion,[4] then $p_1(x) \succ p_2(x)$ implies (normalizing $u(0)$ to 0) that

$$\frac{99/100}{(100+\beta)/100}\left(\frac{89}{99}u(x)+\frac{10}{99}u(5x)\right)=\frac{89u(x)+10u(5x)}{100+\beta}<u(x)$$

which holds if and only if[5]

$$\frac{10}{11+\beta}<\frac{u(x)}{u(5x)}$$

For $q_i(x)$ $i=1,2$ the 0 outcome is always disappointing while the outcomes x and $5x$ are elating. It follows that $q_1(x) \prec q_2(x)$ implies that $\frac{11}{100+89\beta}u(x) < \frac{10}{100+90\beta}u(5x)$. Combining the two inequalities:

$$\frac{10}{11+\beta}<\frac{u(x)}{u(5x)}<\frac{10}{11}\left(\frac{100+89\beta}{100+90\beta}\right)$$

It is easy to see that if $u(x)/u(5x)$ increases with x, the decision maker may exhibit Allais conjectured behavior only for high values of x, holding their disappointment aversion parameter β constant.

Obviously, Cautious Expected Utility preferences (Cerreia-Vioglio et al., 2015) that generalize Gul's Disappointment Aversion model for $\beta > 0$ (Cerreia-Vioglio et al., 2020) can accommodate stake-dependence as well.

3 Experimental design

We implemented two treatment variations: stake size (low or high) and incentives (real or hypothetical). We use a 2x2 between-subjects experimental design for the incentivized arm and a within-subjects design for the hypothetical arm of the experiment.

In the treatments with hypothetical amounts, participants made choices for both lottery sets (presented sequentially, in a random order). In the treatments with real amounts, participants made a single choice (for L1 or L2, determined randomly). We opted for a within-subjects design in the hypothetical treatments because it increases statistical power and because it facilitates a comparison with the literature, which mostly relies on such designs.

However, incentivizing multiple choices when evaluating departures from expected utility is problematic (Holt, 1986; Karni and Safra, 1987; Segal, 1988; Cox et al., 2015;

[3] Obviously, if $p_1(x) \prec p_2(x)$ then $\{5x\}$ in the support of $p_2(x)$ is elating, while $\{0, x\}$ are disappointing. The probability of elation is then .1, while the probability of disappointment is .9.

[4] $\beta = 0$ corresponds to expected utility and $-1 < \beta < 0$ to elation seeking

[5] The following inequality is reversed if $p_1(x) \prec p_2(x)$.

Freeman et al., 2019; Baillon et al., 2022a,b), especially when one of the options guarantees a certain payment (as Option A in L1). Intuitively, paying Lottery set 1 probabilistically will reduce the choice between options A and B to a choice between two risky payments, removing the attractive certainty attribute from Option A. Using a random incentive system (paying one choice randomly) could change the theoretical prediction of non–expected-utility models. Moreover, we wanted to avoid contamination between choices in L1 and L2 that may be a result of other motivations, such as that the (expected) outcome for the one choice would affect the choice for the other problem. We also worried about diluting the (expected) incentives. With real amounts, we therefore use a more conservative (and costly) between-subjects design, where each participant makes a single choice.

For the same reason, we opted for a paired-choice task rather than a valuation task. Multiple price lists are used to make valuation tasks incentive-compatible, but this method removes the certainty effect.

Table 1 lists the treatments. In the results section, for the treatments with real incentives, we often combine the data of the two lottery sets, and then we use the name Real Low to refer to the combination of Real Low L1 and Real Low L2, and we use the name Real High to refer to the combination of Real High L1 and L2.

Table 1: Treatment allocation

Treatment	Lottery set	# Participants
Hypo Low	L1 and L2	160
Hypo High	L1 and L2	159
Real Low L1	L1	162
Real Low L2	L2	158
Real High L1	L1	159
Real High L2	L2	157

Incentives. We conducted our study in Kenya, which has a GDP per capita of $2,082. This allowed us to offer very high incentives. All participants received a flat payment of 400 KES, or roughly $3 (at the time of the experiment, 1 KES=$0.008). The minimum wage in Kenya at the time of the experiment was 15,120 KES a month,[6] and the mean (median) self-reported monthly income of our participants was 13,782 KES (10,000 KES).

In the treatments with low monetary incentives, we used X=100 Kenyan Shillings (KES) in L1 and L2 (0 KES, 100 KES, or 500 KES). In the treatments with high incentives, all amounts were multiplied by a factor of 100 (0 KES, 10,000 KES, or 50,000 KES), which meant that participants had a chance to earn the equivalent of approximately $400, or roughly the equivalent to 4-5 months of income for the mean/median

[6]https://take-profit.org/en/statistics/minimum-wages/kenya/

participant. The average bonus paid out was 87 KES in Real Low and 7834 KES in Real High.

Recruitment of participants. Participants – all of whom were students attending the University of Nairobi and Strathmore University – were recruited by the Busara Center for Behavioral Economics. Those two universities were selected because their students are fluent in English. Each student could only participate in a single arm of the experiment and could only participate if they received an invitation from the Busara Center. Invitations were sent by text message to participants' phones. Each invitation contained a personal code that was valid for 24 hours. Participants needed this personal code and their unique Busara participant identifier to login. Participants were paid by electronic transfers on their mobile phones.

In total, 955 participants completed the study (45 percent female, mean age 24).[7] Table A1 in Appendix A provides a summary of some participant demographics. Table A2 shows the demographics by treatments, splitting by lottery set for the treatments with real incentives where we have a between-subjects design. As can be seen from the table, we cannot reject the hypothesis that the share of females is equal across all treatments. Age differences between treatments are significant but small in terms of point estimates. Reported income tends to be higher in treatments with large amounts. There is no indication that differences in demographics across treatments are driving any of our results. We will return to this when we discuss robustness.

Experimental procedures. The study took place online and was programmed in PHP/MySQL. After providing consent, participants were randomly assigned to one of the treatments. They were told that the study would take about 15-20 minutes, but that they would have up to 60 minutes to complete the study. The general instructions contained information about the type and number of questions they would be asked, and information about the possible amounts of money they could earn. Participants could only continue to the main part after correctly answering test questions about those general instructions. On the main decision screen, participants saw a description of the lotteries and made their choice. Probabilities were communicated in a frequentist manner –described as balls with prizes on them. The order in which the different options were presented on the screen was randomized.

After making their choice, participants answered a short survey about their age, gender, monthly income, and savings. Participants could skip the survey questions if they wished. In the treatments with real stakes, if participants had chosen a lottery, they were directed to a page explaining how the lottery would be implemented. Participants had to click on a widget from the website random.org to draw a random number. After

[7] Another 80 participants logged in but did not complete the study. 31 of those participants did not proceed beyond the consent form or entering their name. The other 49 participants dropped out during the study, most likely because they did not have a stable internet connection. The attrition rate does not differ across treatments. We removed 3 entries which we identified as duplicates based on names and phone numbers.

clicking on the button, the program retrieved a randomly generated number (using one of random.org's application programming interfaces) and the resulting number was displayed on the participant's screen. The last page showed participants their earnings. All instructions and screenshots are available in Appendix B.

Salience and credibility. We took various measures to ensure that the incentive levels were credible and salient. To ensure their salience, we included the possible payments in the general instructions and highlighted them in red. We also emphasized that participants would only make either a single choice (treatments with real incentives) or answer two questions (hypothetical). To ensure that they knew the possible amounts, participants had to answer test questions about the possible earnings and number of questions. Sixty percent of participants answered both test questions correctly on their first attempt, and another 33 percent on their second attempt. Participants could not continue to the main decision screen unless they answered those correctly.

We used the Busara lab to recruit participants because they have a no-deception policy and a good reputation among participants. Invitations to participate came directly from Busara. In the consent form, we added the following language on the study procedures: "The study you are participating in today is being conducted by economists, and our professional standards don't allow us to deceive research subjects. Thus, whatever you will read in the instructions is all true. Everything will actually happen as we describe." We highlighted the part underlined above. To guarantee that the lottery would be fair, we explained in the general instructions that the random draw would be performed on random.org website, over which we have no control. We included a link to that website in case they wanted to know more about how that works.

Pre-registration. The study was pre-registered on aspredicted.org.[8] We specified a sample size of 900 or more participants (including 90 participants in a pilot session to test the software), depending on the available budget that was still left after reaching the target of 900 participants. We still had budget left when we reached the target and so were able to recruit 55 additional participants.

Hypotheses. This design allows us to test the following two main (pre-registered) hypotheses.

(i) H1: Increasing real incentives results in more violations of the independence axiom in the direction predicted by Allais.

(ii) H2: Choices that are made under high real incentives are better approximated by high hypothetical than by low real incentives.

Our null hypothesis is that there is no difference.

[8] `https://aspredicted.org/JD2_H2H`

4 Results

In this section, following our pre-analysis plan, we will describe our main findings. In our plan we wrote that for directional hypotheses we would report p-values for one-sided tests, but for ease of exposition here we will report p-values for two-sided tests instead. Making this change changes none of our conclusions, which would have been the same had we reported p-values for one-sided tests.

4.1 Preliminaries

In the treatments with hypothetical stakes, participants made a choice for each lottery set. We did not detect any order effects depending on which lottery set was presented first.[9] Hence, we have pooled the data.

In Hypo High we re-scaled the amounts employed in previous studies with hypothetical questions to match the amounts in Real High. We start by establishing that, despite the re-scaling, choices in Hypo High replicate typical findings in the literature. Table 2 shows the distribution of choices in this treatment. We find that 46 percent of choices in Hypo High violate EU (choice combinations AB or BA). We can reject that the proportion of times that option A is chosen is equal across the two lottery sets (signed-rank test, $p < .001$). We also find more violations of the type AB (38.4 percent) than of type BA (7.5 percent), as is commonly found. For comparison, we include in the table the numbers from Huck and Müller (2012), who used the usual amounts. The data show very similar patterns, establishing that our subject pool resembles other populations in this respect.

Table 2: Percentages of choices in Hypo High

	Lottery set 2		Total
Lottery set 1	Option A	Option B	
Option A (safer choice)	9.4 [5.7]	38.4 [28.6]	47.8
Option B (riskier choice)	7.5 [7.1]	44.7 [58.6]	52.2
Total	17.0	83.0	100.0

Notes: Choices in the treatment with hypothetical high stakes. Data are pooled over the two rounds. Numbers in brackets are from Huck and Müller (2012), Table 4.

In the treatments with real stakes, participants made only a single choice and hence we cannot identify EU-violations at the individual level, but only at the population level. Following the literature, the prevalence of violations is measured by the risk-difference

[9] In treatment Hypo Low, the percentage of participants choosing the safer option in Lottery set 1 was 26 if it was presented first and 29 if it was presented second, and the difference is not significant ($p = .750$, proportions test). For Lottery set 2 the numbers are 13 and 8 ($p = .286$). In treatment Hypo High, for Lottery set 1 the numbers are 44 and 51 ($p = .388$) and for Lottery set 2 the numbers are 21 and 14 ($p = .245$).

(RD), defined as the difference in the percentage of "safe" choices across the two lottery sets. For treatment i, the risk-difference is defined as:

$$RD_i = p_{1,i} - p_{2,i},$$

where $p_(j,i)$ is the percentage of participants choosing option A in lottery set $j \in \{1,2\}$, respectively.

Table 3 reports, for each treatment, the RD and percentage of safe choices by lottery set. For each treatment, the risk-difference is positive, and we can reject the null-hypothesis that it is equal to zero. This reveals the presence of EU-violations in all treatments.

Table 3: Risk-differences by treatment

Treatment	% safe choices Lottery set 1	% safe choices Lottery set 2	Risk-difference (RD)	Test RD=0 (p-value)
Hypo Low	27.5	10.6	16.9	$< .001$
Hypo High	47.8	17.0	30.8	$< .001$
Real Low	14.8	5.1	9.8	.004
Real High	56.0	31.8	24.1	$< .001$

Notes: Safe choices are option A in each of the lottery sets. p-values in the last column are from tests of proportions. Data for the treatments with hypothetical stakes are pooled over the rounds.

These results are robust to including demographics as controls. Table A3 in Appendix A reports results from a linear probability model (LPM) in which the dependent variable "ChoiceA" is 1 if the participant chose the safer option (options A in Lottery sets 1 and 2). We estimate:

$$ChoiceA_i = b_0 + b_1 LotterySet1_i + \varepsilon_i, \tag{1}$$

where $LotterySet1$ is a dummy variable that equals 1 for lottery set 1 and 0 otherwise. The estimated coefficient b_1 captures how much more likely it is that a participant chooses the safer option in lottery set 1 compared to lottery set 2 (the omitted category), and thus identifies the RD in a treatment. The estimated RD is significant in each treatment, and very similar with and without controls.

4.2 The effect of real incentives

For each lottery set, we find that the safer option (option A) is chosen substantially more often when the stakes are high. The percentage of participants choosing option A increases from 14.8 to 56.0 in Lottery Set 1, and from 5.1 to 31.8 in Lottery Set 2, and the difference is significant in each case ($p < .001$, test of proportions).

To shed light on Hypothesis 1, we investigate how EU violations respond to an increase in the stakes. Faced with real incentives, the RD increases from 9.8 percent to 24.1 percent when the stakes increase from low to high. The difference in RDs is significant with a non-parametric Fleiss test ($Q = 5.11$, $p = 0.024$).

Table A4 in Appendix A shows similar results using a regression analysis. In column 1, we estimate a LPM as in (1), but combining all treatments:

$$\begin{aligned} choiceA_i = b_0 + b_1 LotterySet1_i + b_2 HypoLow_i + b_3 HypoHigh_i + b_4 RealLow_i \\ + b_{12} Lotteryset1_i \times HypoLow_i + b_{13} LotterySet1_i \times HypoHigh_i \\ + b_{14} Lotteryset1_i \times RealLow_i + \varepsilon_i. \quad (2) \end{aligned}$$

The omitted treatment is Real High. The coefficient of interaction term b_{14} captures the difference in RDs between treatment Real Low and Real High. The estimated coefficient b_{14} is negative (-14.4) and significantly different from 0, indicating that the RD in Real Low is lower than the RD in Real High, and robust to including demographics as controls (see column (2)).[10]

It is also instructive to consider whether the increase in risk differences with stakes is caused by a shift in preferences or can be attributed to less noise in behavior. Consider the random-utility approach, according to which decision makers maximize expected utility with an i.i.d. additive utility noise component (e.g., Becker et al., 1963; McFadden, 1974; Loomes, 2005; Butler and Loomes, 2007). Random errors push the choice levels in the direction of 50%. Given that with low stakes our participants select the safe option in L1 in 14.8% of cases and in L2 in only 5.1% of cases, their "true" preferences are even closer to 0% than the observed choice levels. Pure noise would give a risk difference of at most 14.8%. With high stakes it is less likely that noise reverses a preference of a participant. To the extent that noise still affected participants' decisions for high stakes, the 56.0% of safe choices in L1 is an underestimate of the true fraction preferring the certain outcome, while the 31.8% of safe choices in L2 is an overestimate of the true fraction with that preference. Consequently, under the random utility approach, the true risk difference for high stakes exceeds the one reported in the table. The takeaway is that random noise is likely to dampen the observed difference in risk differences. The choice percentages reported in the table represent an underestimation of the degree to which stakes enhance the frequencies of violations of the independence axiom.

Result 1. *Increasing real incentives results in more violations of the Independence Axiom.*

[10] All regression results reported in the main text are robust to including controls. The controls for which the treatments were not balanced - age, income and savings – neither affect the likelihood of a violation of Expected Utility in the treatments with hypothetical stakes, in which we can identify violations at the individual level. Pairwise Spearman correlations between violations and any of the background characteristics are all small (between -0.1 and 0.1) and insignificant (p-values all above 0.200).

4.3 Approximating high incentives

We next test our second hypothesis: that behavior under high real incentives is better approximated by hypothetical choices than by offering low real incentives. The first indication that this is the case, is that the RD in Hypo High is relatively close to the RD in Real High. The difference in RDs is 6.7 (= 30.8–24.1) and not significantly different from zero ($Q = .825$, $p = .364$, Fleiss test). Thus, whereas there is a substantial and significant difference in RDs between Real Low and Real High ($14.3 = 24.1 - 9.8$), the difference between Hypo High and Real High is only half that size and not significant.

Model 1 in Table A4 allows us to directly compare how far each of the RDs in Real Low and Hypo High is from the RD in Real High, and to test if the distance is the same. Consistent with our non-parametric estimates, the estimated RD in Hypo High is relatively close to that in Real High and the difference is not significant (see coefficient $b_{13} = 6.7$). The estimated RD in Real Low is further away from that in Real High and the difference is significant ($b_{14} = -14.4$). However, we cannot reject that the estimated coefficients b_{13} and b_{14} are equal in absolute terms ($p = 0.533$, Wald test). Thus, while we find that the RD in Hypo High is relatively close to that in Real High (difference of 6.7), the RD in Real Low is not significantly further away from the RD in Real High (difference of 14.4).

That Hypo High is a better approximation of choices in Real High is also clearly illustrated in Figure 2. Looking separately at each lottery set, the percentage of safer choices in Real High is better approximated by Hypo High than by Real Low. For each lottery set, there is a significant difference in the percentage of safer choices between Hypo High and Real Low ($p < .001$, test of proportions; this test was not pre-registered). Table A5 presents the same pattern in a regression analysis, where we regress the safe choice on the treatment split by lottery set. We reject that the estimated treatment coefficients of Hypo High and Real Low are equal, and results are again robust to including controls.[11]

Note that there is clear evidence that participants are responsive to the contents of the questions when the amounts are hypothetical. There is a significant difference between the proportion of safer choices across the two lottery sets in Hypo High ($p < .001$, proportions test), the proportion of safer choices is higher in Hypo High compared to Hypo Low for each lottery set ($p < .001$ for lottery set 1 and $p = .100$ for lottery set 2) and there is a significant difference in risk differences between Hypo High and Hypo Low ($p = .034$, Fleiss test).

We also consider participants' decision times to test whether Real High is better

[11] We have already shown that all of the above results are robust to controlling for reported background characteristics. In addition, we performed various other (non-preregistered) robustness tests. Table A6 in Appendix A replicates Table 3 for several subsamples. In Panel B we restrict the sample to the first round only. Choices in Hypo High are in this case even closer to those in Real High, with RDs that are virtually indistinguishable. Results are also robust to excluding participants who decided within 60 seconds (excluding about 15 percent of the sample), see Panel C. The RD in Real High is still significantly higher than in Real Low at the 5 percent level ($Q = 4.68$, $p = 0.031$, Fleiss test).

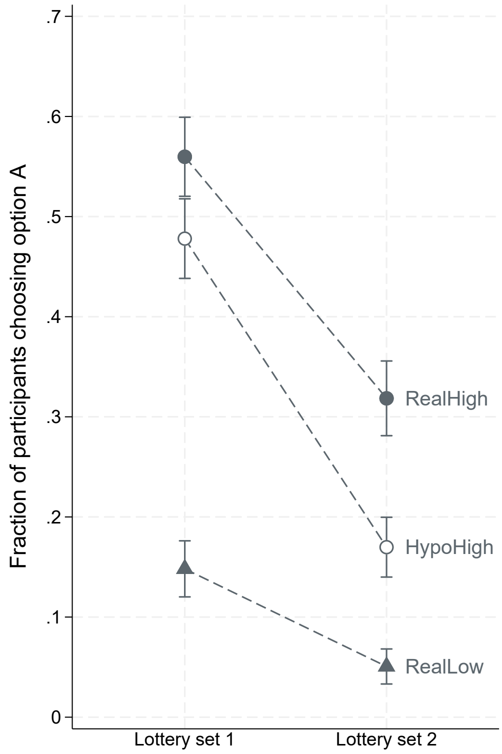

Figure 2: Fraction of safer choices (option A) by lottery set. Error bars indicate +/- 1 s.e. Data for Hypo High are pooled over the rounds.

approximated by Hypo High than Real Low. We recorded the time participants spent on the decision screen (this includes reading the question and making a choice). For a better comparison between the treatments with real incentives (where subjects make a single decision) and hypothetical incentives (where they make two decisions), we focus on the first decision. Table 4 presents the results.

Table 4: Response times (in seconds)

Treatment	Mean RT	Median RT
Hypo Low	153	115
Hypo High	176	136
Real Low	155	118
Real High	184	137

Notes: Response times (RT) in seconds. First choice only for Hypo low and Hypo high.

Participants spent more time on making a decision when the stakes (hypothetical or real) are higher. In Real Low, they spent on average 155 seconds (median 118) on the decision screen. They spent on average an additional 29 seconds in Real High. The median time spent increases by 19 seconds between Real Low and Real High, and the difference is significant ($p = .001$, Rank-sum test). Compared to Real Low, the mean decision time increases by 21 seconds in Hypo High, and the median decision time by 19 seconds. The difference is significant ($p = .010$, Rank-sum test). We again find that Hypo High is a better approximation of behavior in Real High, with very similar decision times. The difference between Hypo High and Real High is not significant ($p = .835$, Rank-sum test).

Result 2. *Both participants' choices and their decision times under real high incentives are better approximated by hypothetical high than by low real incentives.*

Even though Hypo High provides a better approximation of Real High than Real Low, there are systematic differences between Hypo High and Real High. The proportion of safe choices increases from 47.8 percent with hypothetical incentives to 56.0 percent with real incentives in lottery set 1, and from 17.0 percent to 31.8 percent in lottery set 2. The difference is not significant in lottery set 1 ($p = .145$, test of proportions) and significant in lottery set 2 ($p = .002$). This pattern does not carry over to low stakes. There, participants overestimate how risk averse they are with real incentives ($p = .005$ for lottery set 1 and $p = .065$ for lottery set 2).

5 Concluding discussion

Using the Allais choice problem, we set out to answer two questions in our study. First, we investigated whether, as economists often argue, sufficiently large incentives can eliminate deviations from expected utility, in particular the independence axiom. Contrary to this argument, and consistent with the intuitions of Allais and Savage, we found that when we increased incentives by a factor of 100, we generated an increase in deviations—a novel and important contribution to the ongoing debate regarding the impact of high stakes on behavioral biases.

One of the foundations of experimental economics is the belief in the necessity of incentive-compatible designs to uncover authentic behavior. To evaluate this belief, a "true" benchmark is essential—in the case of the Allais problem, behavior under incentives that clearly matter to participants. So the second question we investigated was which of the following two approximations of decision-making is more accurate: responses to questions posed with high hypothetical stakes, or responses to questions posed with real but scaled-down stakes. Our findings indicate that decisions made under hypothetical high stakes serve as a better proxy than low real stakes. In a similar vein, Kühberger et al. (2002) find that the extent to which participants become more risk averse when stakes are multiplied by the factor 25 is much better approximated by hypothetical choices with the same amounts than scaled down real choices. The similarity of hypothetical choices and real choices in our data is also mirrored in the time participants took to make decisions, where decision times in hypothetical scenarios were more aligned with those in high-stakes real conditions than with scaled-down real incentives. However, notable systematic differences persist between decisions made in hypothetical scenarios and those with actual high stakes. Specifically, our study revealed that participants tend to underestimate their level of caution in scenarios involving high real stakes.

Based on our findings, we suggest a pragmatic approach for experimentalists examining high-stakes situations. The best option, of course, is to run experiments with real high stakes. However, if budget considerations do not allow for such experiments, we recommend piloting with real low stakes and high hypothetical stakes. If the responses vary, our research suggests that the outcomes of the real-low-stakes experiment might less accurately represent true behavior than the outcomes of the hypothetical high-stakes experiments. In such discrepant cases, our results suggest that further investigation into the reasons behind the divergence is required. Such an investigation might explore, for example, whether responses to high-stakes hypothetical questions are influenced by factors such as reduced cognitive effort and diminished attention, or by some psychological aspects of the situation that are better captured by a hypothetical high-stakes design.

References

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine. *Econometrica: journal of the Econometric Society*, pages 503–546.

Andersen, S., Ertaç, S., Gneezy, U., Hoffman, M., and List, J. A. (2011). Stakes matter in ultimatum games. *American Economic Review*, 101(7):3427–3439.

Arrow, K. (1971). *Essays in the theory of risk-bearing*, volume 121. North-Holland Amsterdam.

Baillon, A., Halevy, Y., and Li, C. (2022a). Experimental elicitation of ambiguity attitude using the random incentive system. *Experimental Economics*, 25(3):1002–1023.

Baillon, A., Halevy, Y., and Li, C. (2022b). Randomize at your own risk: on the observability of ambiguity aversion. *Econometrica*, 90(3):1085–1107.

Ballinger, T. and Wilcox, N. (1997). Decisions, error and heterogeneity. *The Economic Journal*, 107(443):1090–1105.

Becker, G. M., DeGroot, M. H., and Marschak, J. (1963). Stochastic models of choice behavior. *Behavioral science*, 8(1):41–55.

Binswanger, H. P. (1980). Attitudes toward risk: Experimental measurement in rural india. *American journal of agricultural economics*, 62(3):395–407.

Blavatskyy, P., Ortmann, A., and Panchenko, V. (2022). On the experimental robustness of the allais paradox. *American Economic Journal: Microeconomics*, 14(1):143–163.

Bouchouicha, R. and Vieider, F. M. (2017). Accommodating stake effects under prospect theory. *Journal of Risk and Uncertainty*, 55:1–28.

Butler, D. J. and Loomes, G. C. (2007). Imprecision as an account of the preference reversal phenomenon. *American Economic Review*, 97(1):277–297.

Cameron, L. A. (1999). Raising the stakes in the ultimatum game: Experimental evidence from indonesia. *Economic Inquiry*, 37(1):47–59.

Cerreia-Vioglio, S., Dillenberger, D., and Ortoleva, P. (2015). Cautious expected utility and the certainty effect. *Econometrica*, 83(2):693–728.

Cerreia-Vioglio, S., Dillenberger, D., and Ortoleva, P. (2020). An explicit representation for disappointment aversion and other betweenness preferences. *Theoretical Economics*, 15(4):1509–1546.

Chew, S. and MacCrimmon, K. (1979). Alpha-nu choice theory: an axiomatization of expected utility. *University of British Columbia Faculty of Commerce working paper*, 669.

Chew, S. H. (1983). A generalization of the quasilinear mean with applications to the measurement of income inequality and decision theory resolving the allais paradox. *Econometrica: Journal of the Econometric Society*, pages 1065–1092.

Cox, J. C., Sadiraj, V., and Schmidt, U. (2015). Paradoxes and mechanisms for choice under risk. *Experimental Economics*, 18:215–250.

Dekel, E. (1986). An axiomatic characterization of preferences under uncertainty: Weakening the independence axiom. *Journal of Economic theory*, 40(2):304–318.

Enke, B., Gneezy, U., Hall, B., Martin, D., Nelidov, V., Offerman, T., and Van De Ven, J. (2023). Cognitive biases: Mistakes or missing stakes? *Review of Economics and Statistics*, 105(4):818–832.

Freeman, D. J., Halevy, Y., and Kneeland, T. (2019). Eliciting risk preferences using choice lists. *Quantitative Economics*, 10(1):217–237.

Gul, F. (1991). A theory of disappointment aversion. *Econometrica: Journal of the Econometric Society*, pages 667–686.

Holt, C. A. (1986). Preference reversals and the independence axiom. *The American Economic Review*, 76(3):508–515.

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

Kachelmeier, S. J. and Shehata, M. (1992). Examining risk preferences under high monetary incentives: Experimental evidence from the people's republic of china. *The American economic review*, pages 1120–1141.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):363–391.

Karni, E. and Safra, Z. (1987). " preference reversal" and the observability of preferences by experimental methods. *Econometrica: Journal of the Econometric Society*, pages 675–685.

Kühberger, A., Schulte-Mecklenbeck, M., and Perner, J. (2002). Framing decisions: Hypothetical and real. *Organizational Behavior and Human Decision Processes*, 89(2):1162–1175.

Loomes, G. (2005). Modelling the stochastic component of behaviour in experiments: Some issues for the interpretation of data. *Experimental Economics*, 8:301–323.

Markowitz, H. (1952). The utility of wealth. *Journal of political Economy*, 60(2):151–158.

McFadden, D. (1974). Analysis of qualitative choice behavior. zarembka, p.(ed.): Frontiers in econometrics.

McGranaghan, C., Nielsen, K., O'Donoghue, T., Somerville, J., and Sprenger, C. D. (2024). Distinguishing common ratio preferences from common ratio effects using paired valuation tasks. *American Economic Review*, 114(2):307–347.

Munier, B. and Zaharia, C. (2002). High stakes and acceptance behavior in ultimatum bargaining. *Theory and Decision*, 53(3):187–207.

Plott, C. R. (1986). Laboratory experiments in economics: The implications of posted-price institutions. *Science*, 232(4751):732–738.

Pratt, J. W. (1964). Risk aversion in the small and in the large. *Econometrica*, 32(2):122–136.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of economic behavior & organization*, 3(4):323–343.

Rabin, M. (2000). Risk aversion and expected-utility theory: A calibration theorem. *Econometrica*, 68(5):1281–1292.

Rubinstein, A. (2013). Response time and decision making: An experimental study. *Judgment and Decision Making*, 8(5):540–551.

Samuelson, P. (1963). Risk and uncertainty: A fallacy of large numbers. *scientia*, 57(98):108–113.

Savage, L. J. (1972). *The foundations of statistics*. Courier Corporation.

Segal, U. (1988). Does the preference reversal phenomenon necessarily contradict the independence axiom? *The American Economic Review*, 78(1):233–236.

Slonim, R. and Roth, A. E. (1998). Learning in high stakes ultimatum games: An experiment in the slovak republic. *Econometrica*, pages 569–596.

Smith, V. L. (1982). Microeconomic systems as an experimental science. *The American economic review*, 72(5):923–955.

Smith, V. L. (1991). Rational choice: The contrast between economics and psychology. *Journal of Political Economy*, 99(4):877–897.

Svorenčík, A. and Maas, H. (2016). *The Making of Experimental Economics*. Springer.

Thaler, R. H. (1986). The psychology and economics conference handbook: Comments on simon, on einhorn and hogarth, and on tversky and kahneman. *The Journal of Business*, 59(4):S279–S284.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, 5:297–323.

Von Neumann, J. and Morgenstern, O. (1944). Theory of games and economic behavior, 2nd rev.

Appendix

A Tables

Table A1: Demographics

	Mean	Median	Min	Max	N
Age	24.0	24	18	40	955
Female	0.45	0	0	1	949
Income	14,259	10,000	0	300,000	950
Savings	13,324	2,000	0	2,300,000	953

Table A2: Demographics by Treatment

		Hypo Low	Hypo High	Real Low		Real High		Kruskal-Wallis
				Lottery set 1	Lottery set 2	Lottery set 1	Lottery set 2	test (p-value)
Age	Mean	23.5	25.2	24.1	24.1	23.8	23.6	
	Median	23	25	24	23	23	23	$< .001$
Female	Mean	.42	0.44	0.44	0.42	.49	0.50	
	Median	0	0	0	0	0	0	.581
Income	Mean	12,921	18,019	11,377	11,800	14,448	17,052	
	Median	6,000	10,000	6,750	7,000	10,000	10,000	< 0.001
Savings	Mean	10,655	30,599	6,564	13,143	7,872	11,200	
	Median	1,350	3,000	1,000	2,300	2,000	1,500	.018

Table A3: EU violations

	(1) Hypo Low	(2) Hypo Low	(3) Hypo High	(4) Hypo High	(5) Real Low	(6) Real Low	(7) Real High	(8) Real High
Lotteryset 1	0.169***	0.170***	0.308***	0.312***	0.098***	0.096***	0.241***	0.244***
	(0.036)	(0.036)	(0.048)	(0.049)	(0.033)	(0.033)	(0.054)	(0.055)
Age (years)		-0.007		-0.008		0.006		-0.009
		(0.012)		(0.007)		(0.007)		(0.011)
Female		0.010		0.062		-0.014		0.096*
		(0.050)		(0.051)		(0.034)		(0.056)
Income (in 10k)		0.003		0.002		0.020		-0.001
		(0.021)		(0.016)		(0.022)		(0.019)
Savings (in 10k)		-0.005		0.000		-0.003		-0.011
		(0.011)		(0.001)		(0.006)		(0.007)
Constant	0.106***	0.266	0.170***	0.337*	0.051***	-0.113	0.318***	0.502*
	(0.024)	(0.280)	(0.030)	(0.172)	(0.017)	(0.152)	(0.037)	(0.257)
Observations	320	318	318	314	320	315	316	313
R-squared	0.046	0.051	0.108	0.118	0.026	0.040	0.059	0.082

Notes: Robust standard errors in parentheses, clustered at the subject level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A4: Risk Differences by Treatment

DV: Option A (Safe choice)	(1) safe choice	(2) safe choice
Lotteryset 1	0.241***	0.248***
	(0.054)	(0.054)
Hypo Low	-0.212***	-0.213***
	(0.045)	(0.045)
Hypo High	-0.149***	-0.136***
	(0.048)	(0.049)
Real Low	-0.268***	-0.260***
	(0.041)	(0.042)
Lotteryset 1 X Hypo Low	-0.073	-0.078
	(0.065)	(0.065)
Lotteryset 1 X Hypo High (b_{13})	0.067	0.064
	(0.073)	(0.073)
Lotteryset 1 X Real Low (b_{14})	-0.144**	-0.151**
	(0.064)	(0.064)
Age (years)		-0.004
		(0.004)
Female		0.039
		(0.024)
Income (in 10k)		-0.000
		(0.008)
Savings (in 10k)		0.000
		(0.001)
Constant	0.318***	0.393***
	(0.037)	(0.105)
Test $\|b_{13}\| = \|b_{14}\|$ (p-value)	0.533	0.482
Observations	1,274	1,260
R-squared	0.148	0.153

Notes: Robust standard errors in parentheses, clustered at the subject level. *** p<0.01, ** p<0.05, * p<0.1.

Table A5: Approximation of safe choices in Hypo High and Real Low to Real High

	(1)	(2)	(3)	(4)
	Lotteryset 1		Lotteryset 2	
	safe choice	safe choice	safe choice	safe choice
Hypo High (a)	-0.082	-0.064	-0.149***	-0.149***
	(0.056)	(0.057)	(0.048)	(0.050)
Real Low (b)	-0.412***	-0.404***	-0.268***	-0.266***
	(0.048)	(0.049)	(0.041)	(0.042)
Age (years)		-0.010		0.004
		(0.007)		(0.006)
Female		0.113***		-0.021
		(0.043)		(0.034)
Income (in 10k)		0.003		0.002
		(0.013)		(0.013)
Savings (in 10k)		0.002		-0.002
		(0.001)		(0.001)
Constant	0.560***	0.738***	0.318***	0.233
	(0.039)	(0.173)	(0.037)	(0.147)
Test $\|(a)\| = \|(b)\|$ (p-value)	$< .001$	$< .001$	$< .001$	.002
Observations	480	476	474	466
R-squared	0.133	0.156	0.081	0.082

Notes: Omitted treatment is Real High. Robust standard errors in parentheses, clustered at the subject level. *** p<0.01, ** p<0.05, * p<0.1.

Table A6: Risk-differences by treatment (Robustness checks)

	% safe choices Lottery set 1	% safe choices Lottery set 2	Risk-difference (RD)	Test RD=0 (p-value)
Panel A: Full sample				
Hypo Low	27.5	10.6	16.9	$<$ .001
Hypo High	47.8	17.0	30.8	$<$ .001
Real Low	14.8	5.1	9.8	.004
Real High	56.0	31.8	24.1	$<$ .001
Panel B: First round				
Hypo Low	26.3	13.1	13.2	.035
Hypo High	44.4	20.5	23.9	.001
Real Low	14.8	5.1	9.8	.004
Real High	56.0	31.8	24.1	$<$ .001
Panel C: Response time $\geq$ 60s				
Hypo Low	24.6	11.0	13.6	.005
Hypo High	47.7	19.1	28.6	$<$ .001
Real Low	13.6	2.2	11.4	$<$ .001
Real High	56.3	31.1	25.2	$<$ .001

Notes: Safe choices are option A in each of the lottery sets. p-values in the last column are from tests of proportions. Panels A and C: Data for the treatments with hypothetical stakes are pooled over the rounds.

B Screenshots

Screenshots of the main pages of the low stakes treatments.

Consent

This Informed Consent Form has two parts:

- Information Sheet (to share information about the study with you)
- Certificate of Consent (for clicking Yes if you choose to participate)

Part I: Information Sheet

Why is this project important?

Busara Center for Behavioral Economics, in collaboration with University of Toronto, Harvard Business School, University of California and the University of Amsterdam. We are conducting a study to further understand how people think about decisions. The information below tells you important things you should think about before deciding to join the study. Please ask questions by writing to us via the email provided on this document about any of the information before you decide whether or not to participate.

Who can participate?

You are being invited to take part in this research project because you are at least 18 years and are a student at University of Nairobi and Strathmore University.

Voluntary participation

It is your choice whether or not to participate in this research. If you choose to participate, you may change your mind and leave the study at any time. Refusal to participate or stopping your participation will involve no penalty or loss of benefits to which you are otherwise entitled.

What is involved in this project?

As a participant, you will engage in decision making and answer survey questions.

How long will the project last?

You may leave whenever you finish the tasks. We have built in some time leniency, so you can spend up to 60 minutes if you want.

What are the risks?

If you choose to participate, the risks are no more than you would encounter in everyday life.

What are the benefits?

The benefit to society will be the generalized knowledge derived from the study as described in the description of the purpose of the research. You will also benefit from monetary compensation.

How will we protect your information and maintain confidentiality?

The data we collect will be kept confidential. We collect the following personal information: Your name, phone number, and ip-address. This information will be collected for payment purposes and not stored for longer than necessary. Your responses will be stored permanently in publicly available databases and might be used for (academic) publications. All data that are stored publicly or published will be anonymized and cannot be linked to your personal information.

Cookies policy

We only use cookies that are strictly necessary for the study to be able to function on your browser.

What will happen with the results

The results of this research study may be presented at scientific or professional meetings or published in scientific journals. Your privacy will be maintained in all published and written data resulting from the study.

Can I refuse to participate or withdraw from the study?

Participation is voluntary. Refusal to participate or withdrawing from the research will involve no penalty or loss of benefits to which you might otherwise be entitled.

Compensation

You will receive a flat fee of 400 KSh for participating in this study. You will also be given the opportunity to earn additional money, depending on your decision.

Study procedures

The study you are participating in today is being conducted by economists, and our professional standards don't allow us to deceive research subjects. Thus, whatever you will read in the instructions is all true. Everything will actually happen as we describe.

Who can I contact?

If you have questions about this study, among the researchers for this study is Theo Offerman who can be reached at t.j.s.offerman@uva.nl or any of the following:

- If you have questions, concerns, or complaints,
- If you would like to communicate to the research team,
- If you think the research has harmed you, or
- If you wish to withdraw from the study.

This research has been reviewed by the Committee on the Use of Human Subjects in Research at the University of Amsterdam. They can be reached at ebec@uva.nl for any of the following:

- If your questions, concerns, or complaints are not being answered by the research team,
- If you cannot reach the research team,
- If you have questions about your rights as a research participant.

Or

If you have questions about your rights as a research participant, you may contact:

The Committee Chairperson,
Strathmore University, Institutional Review Board,
P.O. BOX 59857-00200
Nairobi, Kenya
Tel: +254 703 034 375
Madaraka Estate, Ole Sangale Road.
Email: ethicsreview@strathmore.edu

Part II: Certificate of Consent

I have read the above information. I have had the opportunity to ask questions about it and any questions I have been asked have been answered to my satisfaction.

Do you consent to participation in this study? If you decide you want to be in this study, please tick the 'yes' box below.

☐ **I have read and understood the above terms and agreements and give my consent.**

Continue

Welcome

Welcome to the experiment. Please read all instructions very carefully.

The study takes about 15-20 minutes to complete. We have built in some time leniency, so you can spend up to 60 minutes if you want.

Do not rush, the time allowed is more than enough.

Make sure that there are no distractions and keep your focus on the study.

continue ›

[Treatment Real Low]

Instructions

You will be asked to make a single choice that will determine your payment in this experiment, followed by a short survey.

The possible amounts you can earn are 0 KSh, 100 (one hundred) Ksh, and 500 (five hundred) KSh. This is in addition to your participation payment of 400 KSh.

Your payment may depend on the outcome of a lottery. In that case, to guarantee that the lottery is fair, the random draw will be performed on an external website (random.org) over which the organizers of the study have no control. If you like to know more about this website, then go to www.random.org.

When you continue, there will be some quiz questions to test your understanding of the above instructions.

continue ›

[Treatment Hypo Low]

Instructions

You will be asked to answer 2 questions. The amounts of money in the questions are hypothetical, but we ask you to answer the questions as if they were real.

You will earn 400 KSh for answering all questions.

When you continue, there will be some quiz questions to test your understanding of the above instructions.

continue ›

Quiz

Please answer the following quiz questions. You can only continue if you answer them correctly.

1) What are the possible amounts you can earn on top of the participation payment?

Please select...

2) How many choices will you be asked to make (excluding the survey)?

Please select...

Submit

[Treatment real low]

Choice Description

Two boxes contain 100 balls each. Each ball is marked with a prize.

Box A contains 11 balls marked with a prize of 100 KSh, and 89 balls marked with the prize of 0 KSh (zero KSh).

Box B contains 10 balls marked with a prize of 500 KSh, and 90 balls marked with a prize of 0 KSh (zero KSh).

One ball will be randomly drawn from the box you will choose (A or B) and the prize marked on this ball will determine your payment in this experiment.

Thus,

- If you choose Box A, and the randomly drawn ball is marked with a prize of 100 KSh, then you will be paid 100 KSh.
- If you choose Box A, and the randomly drawn ball is marked with a prize of 0 KSh, then you will be paid 0 KSh.
- If you choose Box B, and the randomly drawn ball is marked with a prize of 500 KSh, then you will be paid 500 KSh.
- If you choose Box B, and the randomly drawn ball is marked with a prize of 0 KSh, then you will be paid 0 KSh.

Please choose the box you want to determine your payment in this experiment:

Box A Box B

Submit choice

[Treatment hypo low]

Choice Description

Two boxes contain 100 balls each. Each ball is marked with a prize.

Box A contains 11 balls marked with a prize of 100 KSh, and 89 balls marked with the prize of 0 KSh (zero KSh).

Box B contains 10 balls marked with a prize of 500 KSh, and 90 balls marked with a prize of 0 KSh (zero KSh).

One ball will be randomly drawn from the box you will choose (A or B) and the prize marked on this ball will determine your payment in this experiment.

Thus,

- If you choose Box A, and the randomly drawn ball is marked with a prize of 100 KSh, then you will be paid 100 KSh.
- If you choose Box A, and the randomly drawn ball is marked with a prize of 0 KSh, then you will be paid 0 KSh.
- If you choose Box B, and the randomly drawn ball is marked with a prize of 500 KSh, then you will be paid 500 KSh.
- If you choose Box B, and the randomly drawn ball is marked with a prize of 0 KSh, then you will be paid 0 KSh.

Suppose we offered you the above choice for real money, which box would you choose?

Box A Box B

Submit choice

[Treatment real low]

Your Choice

You chose Box A.

In this box, balls 1,..,11 have been marked with a prize of 100 KSh, balls 12,..,100 have been marked with a prize of 0 KSh.

Click on the button Generate in the widget below to get a random draw. The random draw is performed on the website random.org and determines the number of the ball drawn.

The number of the ball that was randomly drawn is: 42. Your payment is therefore: 0 KSh.

continue ›