Revealing Risky Mistakes through Revisions *

Zachary Breig †
University of Queensland

Paul Feldman ‡
Johns Hopkins University

First Draft: November 27, 2019

This Version: June 8, 2020

Abstract

We define a choice which is modified, absent any informational change, as a *mistake*. In an experiment, we allow subjects to choose from budgets over binary lotteries. To identify mistakes, we allow subjects to revise a subset of their original choices. These mistakes are prevalent: subjects modify over 75% of their original choices when given the chance. The revised decisions are closer to being consistent with several normative measures of decision quality. Subjects make mistakes more often when inexperienced and when choosing over small probabilities.

JEL classification: C91, D81, D91

Keywords: mistakes, risk preferences, uncertainty, revealed preference, expected utility, experiment.

*We are very grateful to David Arjona Rojo, Soo Hong Chew, Kenan Kalayci, Edi Karni, Jim Murphy, John Quah, Karl Schlag, and Tom Wilkening for helpful comments. This research was funded by the NSF's dissertation grant #1824121. *Mistakes* are ours.

†z.breig@uq.edu.au

‡pjfeldman@jhu.edu

1 Introduction

Mistakes are an important part of decision making. Parents tell their children to learn from their mistakes, and political leaders tell their constituents that "mistakes were made." In academic contexts researchers sometimes refer to failures to optimize some particular objective or adherence to a "biased" decision rule as a mistake. However, this approach goes against the basic approaches of revealed preference, and decision makers often may not even agree that their choice is a mistake. This, then, raises our research question: how can a researcher identify mistakes when correct choices are not identifiable *a priori*?

We propose and carry out a methodology to study mistakes. Specifically, we argue that if a choice is revised *without any new information or change in circumstances*, then either the original choice or the revision must be a mistake. This approach can be used in any choice environment and does not rely on the researcher's objective or subjective evaluation of what the correct choice is. We use this definition to study mistakes in a laboratory experiment. We find that when offered the chance to revise their earlier choices, subjects overwhelmingly do so. Subjects' revised choices are better according to every normative measure we employ, suggesting both that the original choices are mistakes and that stationary models of (stochastic) choice can explain their revisions. We then study how the characteristics of decision problems affect the likelihood of a mistake.

In our experiment, 181 undergraduates at the University of Queensland make choices over binary lotteries. Following Andreoni and Harbaugh (2009), subjects trade off the chance of a positive outcome p against the size of that positive outcome $\$x$. Feasible choices satisfy a linear budget constraint of the form $x + \frac{M}{m}p = M$ where M is the maximum outcome and m is the maximum chance. Our subjects choose over the same twenty-five choice sets twice. Subjects are informed about the complete set of budgets and that any of these fifty tasks can be chosen for payment. After choosing from these fifty budgets, subjects revise a random subset of thirty-six of their original choices.

We find that when given the chance, subjects consistently revise their earlier choices. Over 75% of choices are revised, and 176/181 of subjects make at least one revision. Moreover, these revisions are often not small: over 50% of revisions shift at least 10% of a subject's budget from one good to the other.

Comparing revised choices to original choices reveals the former are more consistent with a number of normative criteria. First, revisions decrease the number of violations of first order stochastic dominance (FOSD). Second, revised choices are closer to being rationalizable by an increasing utility function and an increasing utility function which satisfies FOSD. Third, this relationship is preserved over specific functional forms: expected utility and probability weighting. Fourth, revised choices are more likely to be consistent with risk aversion. Finally, stationarity across repetitions of the same budget increases for revised choices, although stationarity only increases when both choices on the same budget are revised on the same screen. The fact that revisions improve all of these metrics indicates that the original choices are mistakes and that revisions cannot be explained by indifference.

Given the evidence that a revision indicates that the original choice is a mistake, we study under what conditions these mistakes are made. First, we find that giving a subject a reminder about the choice they made earlier decreases the likelihood that they make a revision by 17 percentage points, while offering them the chance to revise two choices at once increases the chance of making a revision by just under 3 percentage points. Controlling for subject fixed effects, the amount of time spent making a choice is positively correlated with revising that choice, but this correlation is driven by the *negative* correlation between experience and making revisions. Finally, we find that subjects tend to make more and larger revisions when the budget set contains only lotteries with low probabilities.

The use of revealed preference for study of risk preferences in experiments is not unique to our study. Choi et al. (2007) uses revealed preferences to study consistency with rationality in a study where subjects choose between arrow securities using budgets. Halevy et al. (2018) employs the same data set and a separate experiment to correlate consistency with rationality to parametric fit using predicted behavior as a benchmark. Our revealed preference approach is closer to Polisson et al. (2020). They provide revealed preference tests for different functional specifications and use them to analyze the Choi et al. (2007) and the Halevy et al. (2018) data sets. We adapt their results to budgets over simple binary lotteries and use their finite-data revealed preferences' measures—adapted to various specifications—to reveal mistakes.

Prior research studies how violating specific norms is correlated with real outcomes and financial decisions. Jacobson and Petrie (2009) show that subjects who make choices that are inconsistent with a class of theories of choice under risk do not choose optimally over non-experimental financial instruments. Choi et al. (2014) find that experimental measures of rationality correlate with wealth and education. Rather than using predetermined normative criteria, our measure of a mistake is revealed by the decision makers themselves.

Other studies have considered choice behavior when choices can be objectively ranked, but these rankings must be determined by the decision maker through arithmetic calculation. Caplin et al. (2011) document departures from full rationality and towards a satisficing heuristic in search problems. Kalaycı and Serra-Garcia (2016) find that adding complexity leads to choices which decrease overall payoffs. Gaudeul and Crosetto (2019) find that adding this sort of complexity can induce the attraction affect in decision makers, but that they eventually make more informed decisions. Martínez-Marquina et al. (2019) find that adding uncertainty impedes subjects' ability to maximize their payoff. Our identification of mistakes does not rely on there being an optimal choice which the experimenter knows but the decision maker does not.

Recent work documents how decision makers reconcile potentially inconsistent prior choices. Benjamin et al. (2019) offers subjects hypothetical choices over retirement savings options and confronts them with choices which may be inconsistent. Nielsen and Rehbeck (2019) find that subjects report a desire for their decisions over lotteries to satisfy several axioms and that a majority of subjects revise their choices if they find that these choices violate the axioms. The majority of the revision opportunities in our experiment did not give any indication to the subject that there were inconsistencies in their choices.

The paper proceeds as follows: Section 2 describes the choice environment for binary lotteries. Section 3 describes the experimental procedures. Section 4 features our results contrasting sets of original choices and sets of revisions using several normative benchmarks. Section 5 explores the determinants of mistakes in the experiment. Section 6 provides our final remarks.

2 Choice Environment

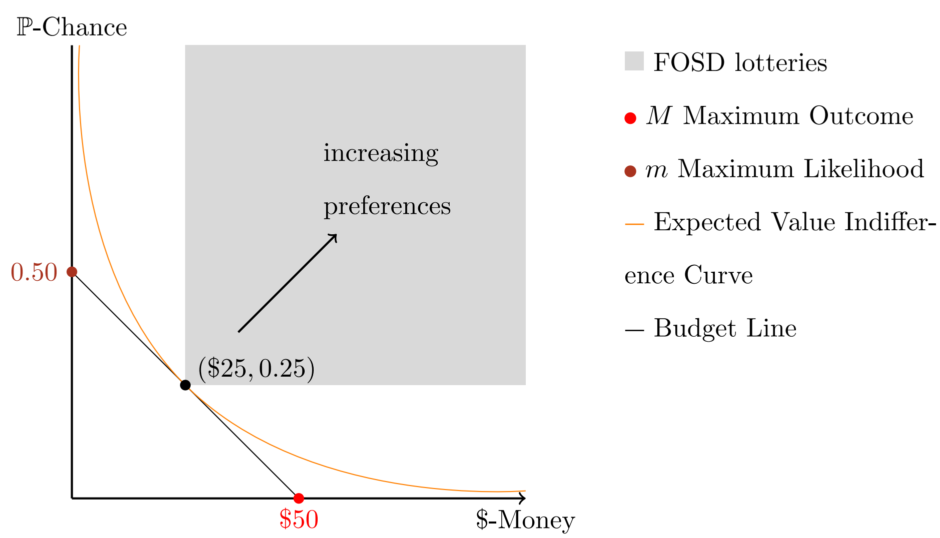
We begin this section by describing the choice environment and properties related to risk preferences. We then show how a decision maker with a canonical form of expected utility preferences makes choices in this environment. We conclude by discussing how we evaluate the concordance of sets of choices with various theories.

Preferences are defined over simple binary lotteries. A simple binary lottery is a lottery which has at most two outcomes, one positive outcome $\$x$ with probability p and $\$0$ with probability $1-p$.

Because one outcome is always \$0, we will abuse notation to represent each lottery by the pair $(\$x, p)$.

The choice problem involves a tradeoff between x and p using a linear budget. Each budget can be described with the maximum prize $M \in \mathbb{R}_{++}$ and the maximum probability $m \in (0, 1]$ which can be chosen from the budget. Thus, any choice from the budget must satisfy $x + \frac{M}{m}p = M$, such that $\frac{M}{m}$ is the "price" of increasing the likelihood of receiving the prize. With this construction, corner allocations on a budget line will always yield a certain outcome of \$0.

Figure 1: Two-goods Diagram for Binary Lotteries



Notes: The decision maker faces a single budget with endpoints $m = 0.5$ and $M = 50$. An expected value maximizer would choose the option $(\$25, 0.25)$, and the indifference curve that this point is on is given in orange.

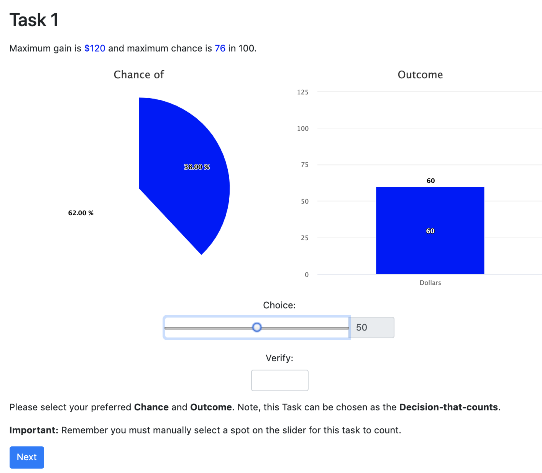
Figure 1 shows how we can depict lotteries, budgets, and increasing preferences using the familiar two-goods diagram. An expected value maximizer would maximize $p \cdot x$, leading to choices $\$x^* = .5M$ and $p^* = .5m$. This highlights two general features of expected utility: first, we may restrict attention to (x, p) without loss of generality, and second, any risk neutral agents devote half their budget to x. Consequently, any risk averse (risk-tolerant) expected utility maximizer will allocate a budget share of more (less) than one-half to the maximum probability.

Now, consider an expected utility maximizer with CRRA preferences given by $u(x) = x^{\alpha}$. An increasing transformation can be applied to the agent's objective function to obtain $p^{\frac{1}{1+\alpha}} x^{\frac{\alpha}{1+\alpha}}$. Thus, these preferences can be represented by a Cobb-Douglass utility function, and the budget shares the decision maker chooses will be constant across budgets.

In our results, we will opt for non-parametric revealed preferences tests. In particular, we will use Afriat's theorem first to determine whether an increasing, concave, and continuous function can rationalize our data. Second, we will use a generalization of Afriat's theorem (Nishimura et al., 2017; Polisson et al., 2020) that allows us to test for the ability of specific functional forms to rationalize our data and extend the standard measure of rationality. The functional forms we consider are expected utility ($p * u(x)$) and generalized probability weighting ($\pi(p) * v(x)$).

3 Experimental Design

Figure 2: Experimental Task Summary



(a) Sample Task

Maximum Chance	Maximum Outcome
50 in 100 chance	$30
60 in 100 chance	$30
50 in 100 chance	$40
80 in 100 chance	$40
100 in 100 chance	$50
60 in 100 chance	$60
75 in 100 chance	$60
100 in 100 chance	$60
40 in 100 chance	$80
80 in 100 chance	$80
100 in 100 chance	$80
15 in 100 chance	$100
20 in 100 chance	$100
100 in 100 chance	$100
15 in 100 chance	$120
30 in 100 chance	$120
60 in 100 chance	$120
30 in 100 chance	$150
20 in 100 chance	$160
40 in 100 chance	$160
80 in 100 chance	$160
25 in 100 chance	$200
30 in 100 chance	$200
40 in 100 chance	$200
50 in 100 chance	$200

(b) Full Set of Distinct Tasks

Notes: Panel A shows a sample choice task. Panel B summarizes the full set of budgets as it was presented to our subjects.

For each task, we elicit subjects' preferences over the set of binary lotteries—lotteries that give $\$x$ with probability p and $0 otherwise—in a linear budget with endpoints $\{M, m\}$. The ratio of M to m gives the tradeoff between the size of the outcome and its likelihood. We emphasize three advantages of using this method. First, because budgets are linear in the $(\$x, p)$ plane, standard consumer theory can be applied.[1,2] Second, because setting either $\$x$ or p equal to 0 is strictly dominated, choices will typically be interior. This is beneficial because corner choices pose identification issues for budget-based methods. Third, in contrast to other linear budgets over lotteries (for example Feldman and Rehbeck (2019) for probabilities or Choi et al. (2007) for outcomes), this method features variation in both the probabilities and the outcomes simultaneously. A sample task, as subjects saw it, appears in Figure 2a.

Subjects select their preferred lottery from each budget using a slider. Before making each choice, no information is displayed on the subject's screen other than the maximum outcome and the maximum chance. Once a subject interacts with the slider, a pie-chart is used to represent probabilities and a bar-chart represents the positive monetary amount. As the subject moves the slider to the right (left), the pie-chart increases (decreases) and the bar decreases (increases). Once a subjects has identified their preferred bundle, they confirm their selection by separately entering it in a box.

Figure 3 summarizes the budget sets used. The fact that the budgets cross allows for analysis of traditional rationality measures. The set also includes parallel budgets and pure price shifts to allow for analysis of income and substitution effects. A pre-analysis plan was submitted to the AER RCT registry (AEARCTR-0004572) prior to the experiment and the visual interface was coded using oTree (Chen et al., 2016).

3.1 Implementation

One hundred and eighty-one University of Queensland undergraduates read the instructions on their computer terminal while the experimenter read the instructions aloud. Before starting the main part of the experiment, subjects completed three sample tasks.[3] These examples familiarize the subjects with how the slider affects positive outcomes, chances, and the tradeoff between them. The experiment itself has two parts: repetitions and revisions.

In Part I of the experiment, subjects made choices in 50 tasks. The twenty-five different budgets that were used were described to subjects by presenting them with a list of the pairs of maximum

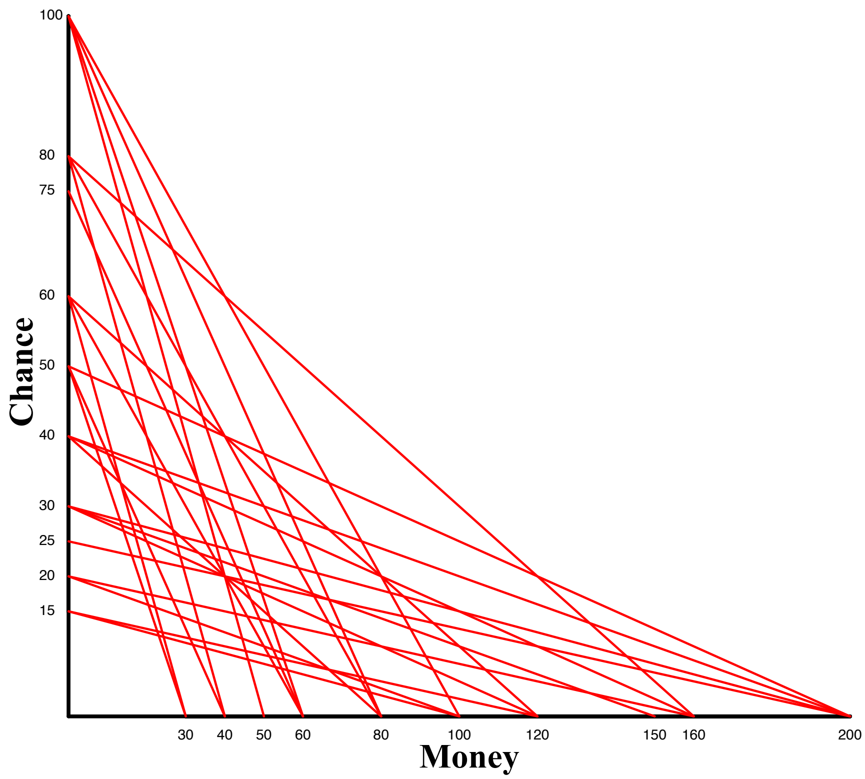
[1]Only compactness and downward comprehensiveness are necessary, see Nishimura et al. (2017) for a detailed explanation.

[2]This, of course, requires for preferences to be monotonic in money and the probability of receiving money. This is an assumption we maintain throughout the paper.

[3]Sample tasks and the complete instructions appear in Appendix B.

outcomes and chances. The information, as subjects saw it, is summarized in Figure 2b. Each subject chose from the twenty-five unique budgets followed by choosing from the same twenty-five budgets for a second time. However, the order across subjects and across the two blocks was random.

Figure 3: Budgets



Notes: This figure depicts the full set of our experimental budgets using a two-goods diagram. This figure was not displayed to subjects.

In Part II of the experiment, subjects revise a subset of the choices they made in these first 50 tasks. These revision tasks feature a 2×2 within-subject treatment which changes the presentation of the tasks (see Table 1). The first dimension of treatment is the number of revisions they make within a revision task. Each revision task is either a "single" (in which the subject can revise a single earlier choice) or a "double" (in which the subject can revise two earlier identical tasks on a single screen). The second dimension of treatment is whether or not subjects are given a reminder of the original choice they made.[4] The subject makes six revisions in each condition, leading to 36

[4]For revisions with reminders, subjects are shown a pie-chart and bar graph that matched their prior choice. The pie-chart and bar graph are replaced with representations of their current choices as soon as they click on the slider. For all other choices, the initial graph was empty.

choices, without replacement, being revised from 24 unique budgets. Thus, no single task is revised twice. The order of treatments is randomized at the subject level.

Table 1: Revisions by Type

	reminders	no reminders
single choice	6	6
double choices	12	12

Notes: Double choices featured the same choice problem twice—choices over the same budget. Appendix B contains samples for each type of revision.

To incentivize choices, one of the fifty choices was chosen at random from the revised set and implemented. Subjects made an average of 9.5 (19.5 s.d.) Australian dollars (AUD) and received a 10 AUD as a participation payment. Both experimental parts took around 30 minutes.

Table 2 provides summary statistics. Each of the 181 subjects made 50 choices in the first section of the experiment, for a total of 9050. Each choice is the portion of the budget (out of 100) which is allotted to increasing the probability of receiving the prize. The average choice was to devote just over 54% of their budget towards probability, indicating a mild level of risk aversion. Subjects spent an average of roughly 24 seconds per task on these first fifty tasks.

Each subject faced 36 revisions problems, for a total of 6516. We say that the subject made a revision if their revision choice differs from their original choice. Subjects make revisions roughly 75% of the time when given the chance. The size of the revision is the difference in the portion of the budget assigned to probability between the original choice and the revision. These revisions are on average near zero (indicating that revisions are not on average significantly more or less risky than the original choices). However, the average absolute value of the revision is nearly 12, indicating that subjects are on average shifting more than 10% of their budget from prize to probability (or vice-versa).

Table 2: Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Original Choice	9050	54.297	20.746	0	100
Seconds on Page	9050	24.024	17.661	3	375
Made Revision	6516	.752	.432	0	1
Revision	6516	.127	19.581	-100	100
Abs. Revision	6516	11.977	15.491	0	100

4 Do Mistakes Have Normative Content?

This section examines whether mistakes, as we have defined them, lead to "poor" choices. To decide whether choices are indeed worse, we evaluate them according to traditional normative benchmarks. The first benchmark is picking strictly dominated alternatives (violations of monotonicity), the second benchmark is rationalizability by an increasing utility function, the third benchmark is consistency with various functional forms (including expected utility), the fourth benchmark is consistency with risk aversion, and the fifth benchmark is stationarity of choice behavior (picking the same lottery across repetitions of the same choice task).

4.1 Monotonicity

We find that 32/181 subjects violate monotonicity by selecting a corner—a certain outcome of zero—on at least one budget for their original set of choices. In contrast, 17/181 subjects violate monotonicity when we look at their revised sets of choices. For each subject, the original set consists of their 50 initial choices while the revised set consists of 14 of their original choices and 36 revisions—the revisions that overwrite their initial choices.

The mean number of corners chosen in the original 50 budget sets is 0.768, while the mean number of corners in the revised set of 50 choices is 0.525. Furthermore, only three subjects increase the number of corners chosen in their revised set, while 29 subjects decrease the number of corners chosen.

4.2 Rationalizability with an Increasing Utility Function

The next benchmark which we use to compare choices to revisions is rationalizability. Following Afriat (1967) and Varian (1982), we define a set of choices to be rationalizable if there exists a utility function which the choices maximize. Because every data set can be rationalized by a utility function (e.g. the constant utility function), we further place the restriction that the utility function which is maximized must be increasing.

Because the simple test for rationalizability has a binary outcome, it is common to use a more continuous measure. The measure of rationalizability we employ is Afriat's index (AI), which is a number e between zero and one (Afriat, 1973). Mathematically, a lower index reduces the number of restrictions that a utility function has to satisfy: rather than a bundle (x_i, p_i) from budget $\{M_i, m_i\}$

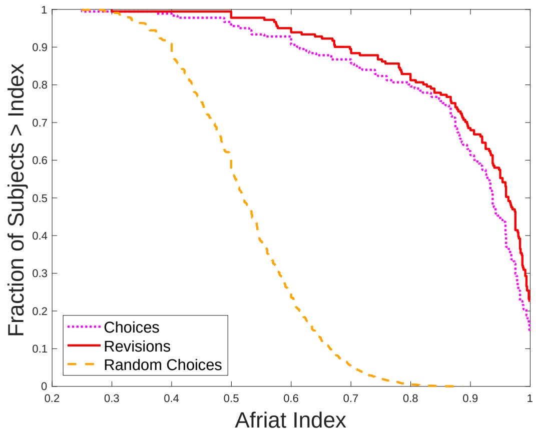
needing to have associated utility higher than all bundles which satisfy $X + \frac{M_i}{m_i} p \leqslant M_i$, they instead need to have utility higher than all bundles which satisfy $X + \frac{M_i}{m_i} p \leqslant eM_i$. The AI for a set of choices is the highest e for which the choices are rationalizable. This index has become a common measure for how far a set of choices is from being rationalizable (Andreoni and Miller, 2002; Choi et al., 2007; Polisson et al., 2020).

In our context, there are two relevant types of monotonicity. The first is monotonicity in the classic sense: the decision maker strictly prefers a bundle which is strictly higher in one dimension and no lower in any other dimension. In this case, we use the Afriat Index as it has been classically defined for any collections of choices from linear budget constraints. Our stronger notion of monotonicity is first order stochastic dominance. This places the same restrictions as standard monotonicity, but also requires that the decision maker never chooses on the endpoints of the budget line (because any interior choice first order stochastically dominates the endpoints, which guarantee a payoff of zero). When using FOSD as the notion of monotonicity, a set of choices is assigned an index of zero if it includes any choices on the endpoints of the budget line. Otherwise, it is equal to the standard Afriat Index.

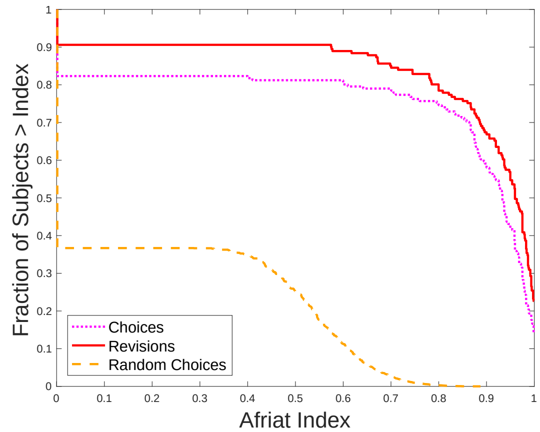
The Afriat indices and Afriat indices under FOSD can be found in Figures 4a and 4b, respectively. The figures also contain the Afriat Index for a uniform random choice rule that measures the power of our design to detect violations of rationality (Bronars, 1987).[5] Clearly, both the Afriat and Afriat FOSD indices of the revised sets of choices first order stochastically dominate the distributions from the original sets of choices. Revised decisions are closer to being rationalizable by a utility function, indicating that some of the original decisions may have been of poor quality.

[5]Choices on the budgets were discretized to 101 distinct choices that are equidistant on each budget. Our uniform random rule randomizes over the options on a budget subjects could make.

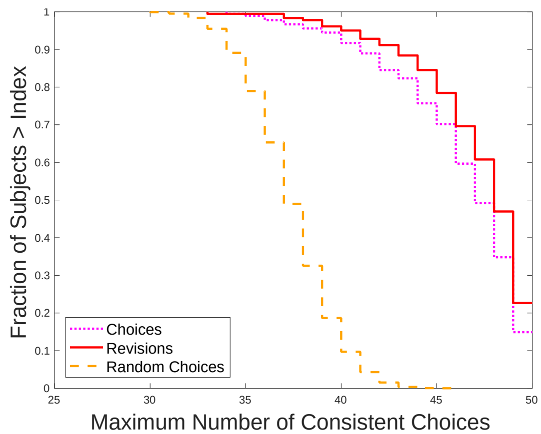
Figure 4: Rationalizibility for Original Choices and Revised Choices



(a) Afriat Index



(b) Afriat FOSD Index



(c) Maximal Acyclic Choices

Notes: These figures illustrate our main rationality results using Afriat's index (Panel A), Afriat's index under FOSD (Panel B), and maximal transitive relation (Panel C). Each panel contains the fraction of subjects whose rationality index is greater than the x-axis value for their original choices, their revised choices, and a uniformly random choice rule (n=10,000).

We also report another consistency measure for the maximum acyclic set—the maximum number of choices that could be rationalized by an increasing utility function (Houtman and Maks, 1985; Rehbeck, 2020). This measure appears in Figure 4c and does not alter the result that the consistency of revised choices is always higher for any fraction of subjects.

Our general rationalization results are as follows. For their original choices, 80 subjects have an Afriat Index$\geqslant$95%, 76 subjects have an FOSD consistent Afriat Index$\geqslant$95%, and 95 subjects have their maximum number of consistent choices>47. For their revised choices, the number of consistent subjects increases across all three benchmarks to 100, 99, and 113, respectively. Median consistencies for the original choices are 94%, 93%, and 47, compared to 96%, 96%, and 48 for the revised choices, across the three benchmarks. A signed rank test rejects ($p<.01$) equality of distributions between original choices and revised choices for the three benchmarks. Hence, the number of subjects whose choices can be rationalized by some utility function is unambiguously larger for revised choices as implied by these metrics and Figure 4.

4.3 Consistency with Common Utility Functions

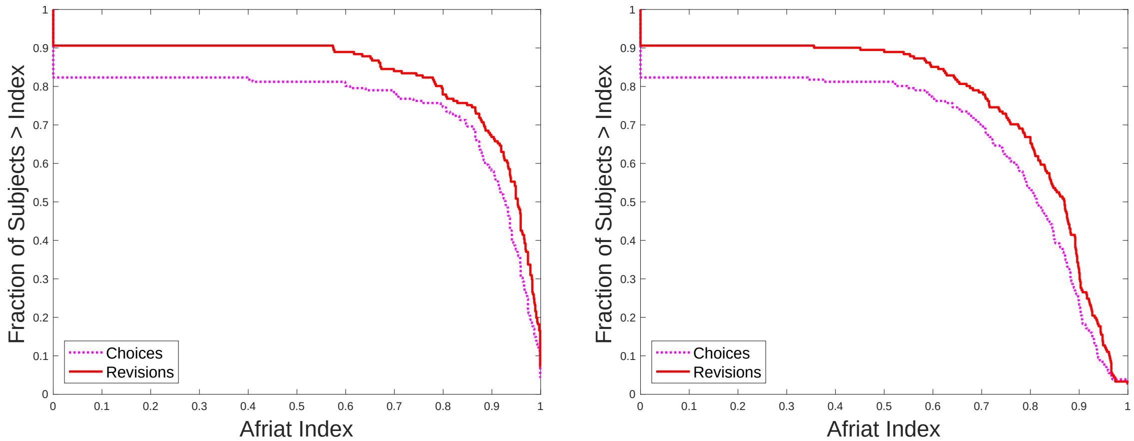
An additional means of evaluating a subject's choices is to establish whether those choices are consistent with utility functions which some have found normatively appealing, such as expected utility. Given recent developments in the theory of revealed preferences we can test these specific models of behavior. In particular a corollary of the results in Nishimura et al. (2017) is that any utility functional representation (because it induces a preorder on the set of choices) can be tested by checking for a cyclical mononotonicity condition under that same preorder. We can further adapt results from Polisson et al. (2020) to our context, allowing us to check for these cyclical monotonicity conditions over a a finite set of points induced by each sequence of choices. Formal details and results are collected in Appendix A.

The results of Nishimura et al. (2017) and Polisson et al. (2020) also show that we can use indices similar to Afriat's index to derive weaker tests of this cyclical monotonicity condition. Essentially, a set of choices will have an index of e if e is the minimum value such that there exists a utility function from the specified family which assigns a utility to each bundle (x_i, p_i) chosen from budget $\{M_i, m_i\}$ that is higher than all bundles that satisfy $X + \frac{M_i}{m_i}p \leqslant eM_i$.

The utility families we consider are cumulative probability weighting (PW) and expected utility (EU). Because each of these families places additional restrictions on the previous one and all must satisfy the restrictions from Afriat's theorem, the PW index is lower than the Afriat FOSD index and the EU index is lower than the PW index.

The results for the indices can be found in Figures 5a and 5b. The PW indices of the revised sets of choices first order stochastically dominate the PW indices of the original sets of choices. The EU indices of the revised sets of choices *almost* first order stochastically dominate the EU indices of the original sets of choices. Thus, when offered the chance, subjects revise their choices in a way that makes them closer to being consistent with commonly used families of utility functions.

Figure 5: Rationalizibility Using Common Utility Functions



(a) Probability Weighting Index (b) Expected Utility Index

Notes: These figures illustrate our main rationality for probability weighting ($\sum \pi(p_i)u(x_i)$, Panel A) and expected utility ($\sum p_i u(x_i)$, Panel B). Each panel contains the fraction of subjects whose rationality index is greater than x-axis value for the original choices and the revised choices.

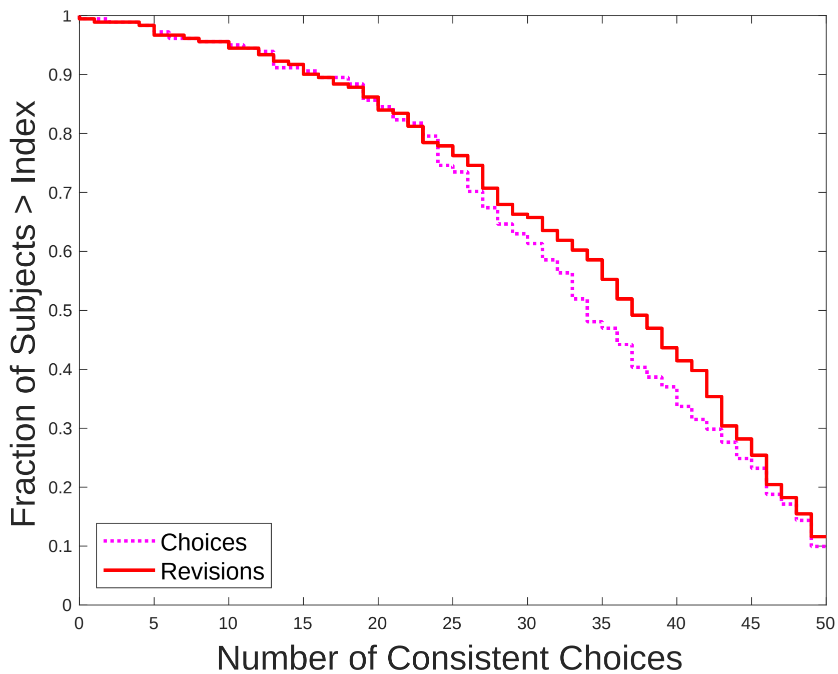
Our specific rationalization results are as follows. For their original choices, 68 subjects have a PW-consistent Afriat Index⩾95%, and 14 subjects have an EU-consistent Afriat Index⩾95%. For their revised choices, the number of consistent subjects increases across both specifications to 92 and 23, respectively. Median consistencies for the original choices are 93% and 81% compared to 95% and 87% for the revised choices, across the two specifications. A signed rank test rejects (p<.01) equality of distributions between original choices and revised choices for the two specifications. The number of subjects whose choices can be rationalized by either an expected utility or probability weighting function is larger for revised choices as implied by these metrics and Figure 5.

4.4 Risk Aversion

We also discuss a simple heuristic benchmark for risk aversion. Note that any allocation where the budget shares favor the outcome (x) over the (p) likelihood will be second order stochastically dominated by equal shares—the optimal allocation for an expected value maximizer. Therefore, any concave EU subject—or any risk averse subject—can never select an allocation the places a greater budget share on the outcome.[6] Our simple benchmark counts the number of choices that

[6]Note that for a subject to be risk averse it is not sufficient for U to be concave. For example, $U(x,p) = \log(p) + 2\log(x)$ is concave and it represents the same preferences as $V(x,p) = p * x^2$, a risk tolerant utility function. For probability weighting both u and π must be concave for preferences to be consistent with risk aversion (Hong et al., 1987).

Figure 6: Number of Choices that are Consistent with Risk Aversion Across Subjects



Notes: This figure plots the fraction of subjects whose number of choices that are consistent with risk aversion are greater than the x-axis value for their original choices and their revised choices.

are consistent with FOSD and that place a greater budget share on the probability. As depicted in Figure 6, this measure provides a benchmark for the maximum number of choices that can be consistent with risk aversion.

We find that 18/181 subjects do not violate risk aversion—on at least one budget—over their original choices. Revisions lead to a slight increase in the number of subjects that do not violate risk aversion 21/181 at all. 51 subjects increase the number of violations in their revisions, while 100 subjects decrease the number of violations. A signed rank test rejects the null hypothesis that the number of violations of risk aversion is the same across original choices and revisions ($p < 0.01$). Whether risk aversion is a normatively compelling criteria is a choice for the reader.

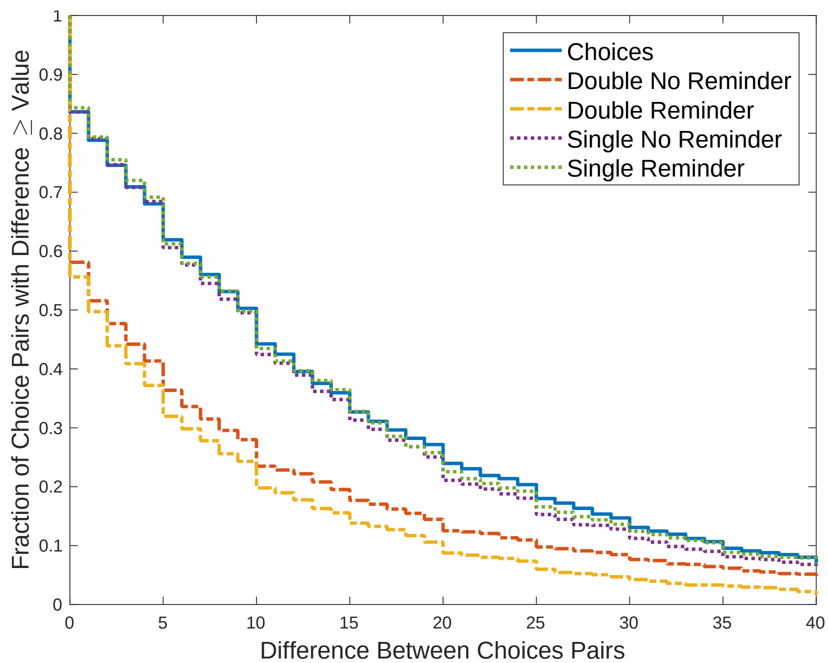
4.5 Stationarity

Only five subjects were stationary in all of their choices.[7] 16.35% of subjects' original pairs of choices were stationary. When pairing a revised choice in the single revision treatment with its

[7] These five subjects were all consistent with expected value maximization, which involves always choosing exactly in the middle of the budget line.

unrevised paired choice, the two are only equal to each other in 16.02% of cases. When two revisions are made at a single moment, they are equal to each other in 43.14% of all cases.

Figure 7: Non-Stationarity in Choice Behavior



Notes: This figure plots the fraction of choice pairs that are inconsistent with stationarity across our experimental treatments. The x-axis captures how far apart choices were across the repetitions.

Figure 7 plots the distributions of differences between pairs of decisions in these cases. It is immediately apparent that allowing for a single decision to be revised does not necessarily mean that this revised choice will be any closer to its paired choice than the original choice was—there is essentially no difference between the CDFs of differences between the original choices and the single revision problems. On the other hand, there is a clear shift to the left of the distribution of differences when two choices are made at once. Signed-rank tests for equality of distributions of differences between original sets and revised sets gives a $p = 0.02$ for single revisions and $p < 0.01$ for double revisions.

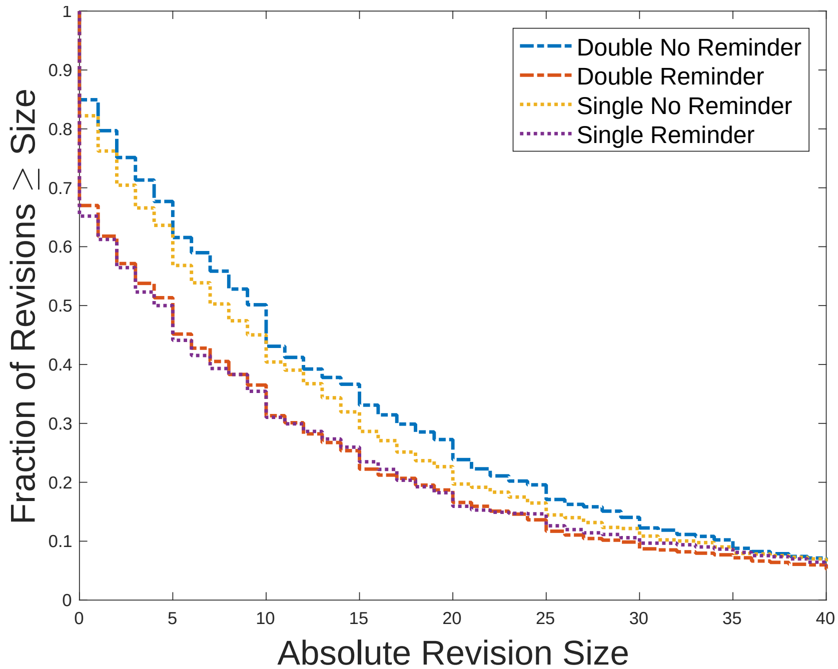
Repeated choices should only match under two criteria. First, preference over single budgets should have a unique maximizer. Second, preference need to be dynamically consistent and consistent with consequentialism (Machina, 1989). The latter criteria is satisfied by expected utility. The former is satisfied only if preferences over single outcome lotteries are strictly quasiconcave. For instance, Friedman-Savage expected utility preference can violate stationary.[8] Thus, stationarity is not a property of expected utility. Note further that non-stationarity is implied by *preference for randomization*. Often this behavior has been associated with quasiconcavity in the probabilities, but in our context quasiconvexity in the probabilities can also accommodate it.

5 Mistakes and Their Determinants

This section discusses the characteristics of the decision problems over which subjects made mistakes. As discussed above, we label a decision as a "mistake" if when given the chance to revise the decision without any new information, the subject decides to make a revision. Subjects were offered the chance to revise 36 of their 50 decisions. Just over 75% of the original choices were revised when subjects were offered the chance. These revisions could have made the decision less risky (a positive revision) or more risky (a negative revision). Revisions were on average near 0 (mean of 0.127 with clustered standard error 0.603), indicating that subjects did not on average revise their decisions towards probabilities or outcomes.

[8] An example can be provided upon request.

Figure 8: Absolute Size of Revisions



Notes: This figure showcases the relationship between the original choices and the revised choices by measuring the distance between them. The curves represent the fraction of choices whose distance was greater than the x-axis value across the experimental treatments.

Despite subjects not revising towards one direction or the other on average, the mean absolute value of revisions was 11.977 (clustered s.e. 0.634). This represents over 10% of subjects' budgets. This is not the result of a few outliers: Over 30% of choices had an absolute revision of at least 15.

5.1 Treatments and the Likelihood of Revisions

Figure 8 graphically represents the effects that treatments have on revisions. It shows the distribution of absolute revision size for each of the treatments. Offering subjects a reminder of their previous decision tends to make it less likely that they will revise that decision.

Table 3 shows the effects that treatments have on revisions in regression form. Columns (1) and (2) report how the likelihood of making a revision changes with treatments, while columns (3) and (4) show how the absolute value of revisions change with treatments. The treatment effects are consistent in all cases. Reminding subjects of what they chose previously both makes the subject less likely to revise and makes the average absolute revision smaller. Giving the subject two revisions at once makes subjects slightly more likely to revise and increases the size of revisions. The interaction of these treatments makes revisions less likely and the absolute size of revisions smaller, but only the latter of these effects is significant at the 10% level.

Table 3: Treatment Effects

	(1) Made Revision	(2) Made Revision	(3) Abs. Revision	(4) Abs. Revision
Reminder	-0.17***	-0.17***	-2.27***	-2.21***
	(0.022)	(0.022)	(0.63)	(0.63)
Double	0.027**	0.028**	1.19**	1.24**
	(0.013)	(0.013)	(0.60)	(0.61)
Reminder × Double	-0.0092	-0.0097	-1.30*	-1.41*
	(0.022)	(0.022)	(0.74)	(0.75)
Subject FE	No	Yes	No	Yes
Task FE	No	Yes	No	Yes
Observations	6516	6516	6516	6516

Notes: Linear regression clustered at the subject level. Each column represents a different regression, with the column head specifying the dependent variable. Significance indicated by: *** p<0.01, ** p<0.05, * p<0.1.

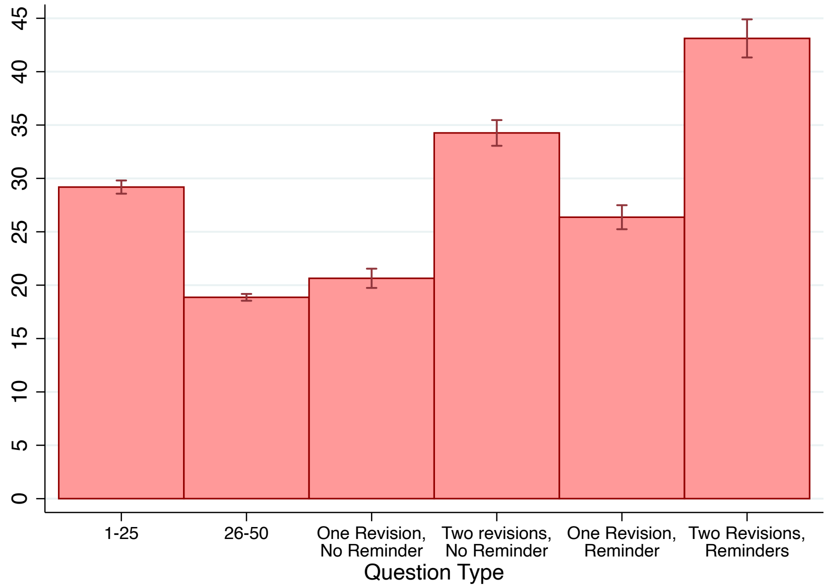
5.2 Decision Times

The amount of time that subjects took to complete each type of problem can be found in Figure 9. Single choices take less time than double choices over the same budget and on the same screen. Earlier choices and choices with reminders also take more time. The average time taken on the first portion of the experiment was just over 24 seconds per task.

The time taken to make a decision does vary with the likelihood that the decision is later revised. This can be seen in Figure 10. The relationship appears to be nonlinear: decisions that are taken very quickly are revised less often, but outside of this range time taken is negatively correlated with revision rates. However, this relationship is not causal. Because subjects are not randomly assigned to time taken, unobservable characteristics of the subject or decision problem may be driving the relationship between decision time and mistake rates.

The relationship between decision time and revisions is further explored in Table 4. The dependent variable in this table is the absolute size of revisions. Column (1) shows that over all observations, the amount of time spent on making a decision is uncorrelated with the amount that this decision is revised. However, Column (3) demonstrates that after controlling for both subject and task (i.e. budget set) fixed effects, there is a positive correlation between between time taken

Figure 9: Time Taken by Decision Type

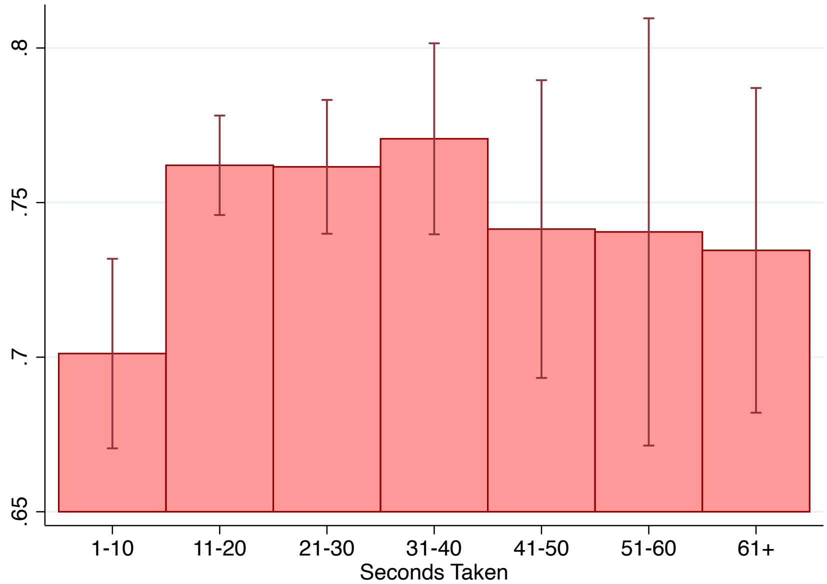


Notes: This figure shows average time spent on a task's page for various decision types. The height of the bar gives the sample mean for each category of decision and the thinner lines give the 95% confidence interval for the mean.

and revision size.[9] This suggests that subjects who make decisions faster are less likely to revise their decisions later, but that conditional on the subject, spending more time on a decision is associated with a higher likelihood of revising the decision later.

[9]The difference in coefficients from time taken is due almost entirely to the addition of subject fixed effects rather than task fixed effects.

Figure 10: Revision Rates by Time Taken



Notes: This figure shows how the likelihood of a revision varies with the amount of time spend on the original choice. The height of the bar gives the sample mean for each time window and the thinner lines give the 95% confidence interval for the mean. Decisions which were made very quickly were less likely to be revised, but outside of that range the time taken on a decision is negatively correlated with revision rates.

Columns (2) and (4) of Table 4 additionally control for the round the decision is made in, which varies between 1 and 50. When controlling for the round, the relationship between time taken and the size of revision is both small and statistically insignificant. After controlling for individual fixed effects, the relationship between time taken and the size of revisions is driven by the fact that subjects both take longer and make more mistakes when they are less experienced.

5.3 Budget Characteristics

Tables 5 and 6 study how the characteristics of the budgets the subject faces affect revisions. Since a budget is completely characterized by its endpoints, we regress revision rates and revision size on these endpoints.

Table 5 shows the linear relationship between the size of budgets and the size and likelihood of making a revision. The coefficient for both regressions on the maximum prize is near zero. Thus, the potential size of the prize does not affect the likelihood that the decision maker makes a mistake. This

Table 4: Decision Time

	(1) Abs. Revision	(2) Abs. Revision	(3) Abs. Revision	(4) Abs. Revision
Seconds on Page	0.00031	-0.025	0.033***	0.0069
	(0.020)	(0.023)	(0.012)	(0.013)
Round		-0.083***		-0.068***
		(0.018)		(0.017)
Subject FE	No	No	Yes	Yes
Task FE	No	No	Yes	Yes
Observations	6516	6516	6516	6516

Notes: Linear regression clustered at the subject level. Each column represents a different regression, but all columns use the absolute value of the revision as the dependent variable. Significance indicated by: *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Budget Characteristics

	(1) Made Revision	(2) Made Revision	(3) Abs. Revision	(4) Abs. Revision
Max Prize	-0.00012		-0.0023	
	(0.00011)		(0.0039)	
Max Probability	-0.095***		-2.32**	
	(0.025)		(1.00)	
Round	-0.00085**	-0.00087**	-0.071***	-0.071***
	(0.00034)	(0.00034)	(0.015)	(0.015)
Subject FE	Yes	Yes	Yes	Yes
Task FE	No	Yes	No	Yes
Observations	6516	6516	6516	6516

Notes: Linear regression clustered at the subject level. Each column represents a different regression, with the column head specifying the dependent variable. Significance indicated by: *** p<0.01, ** p<0.05, * p<0.1.

contrasts with the coefficient on the maximum likelihood of receiving the prize, which is significantly negative. This implies that subjects have a harder time making choices when the probabilities that they are choosing between are small.

Table 6 repeats the analysis of Table 5 in a more flexible way. In particular, it uses dummy variables for each maximum prize when estimating the effect of changing the maximum probability, and it uses dummy variables for each maximum probability when estimating the effect of changing the maximum prize. These results largely confirm the results from Table 5: the maximum size of the prize does not affect the likelihood of revisions, but a smaller maximum probability makes revisions larger and more likely.

Table 6: Robustness of Budget Characteristics

	(1) Made Revision	(2) Made Revision	(3) Abs. Revision	(4) Abs. Revision
Max Prize		-0.00011		-0.0023
		(0.00012)		(0.0040)
Max Probability	-0.093***		-2.01*	
	(0.025)		(1.03)	
Subject FE	Yes	Yes	Yes	Yes
Max Prize FE	Yes	No	Yes	No
Max Probability FE	No	Yes	No	Yes
Observations	6516	6516	6516	6516

Notes: Linear regression clustered at the subject level. Each column represents a different regression, with the column head specifying the dependent variable. Significance indicated by: *** p<0.01, ** p<0.05, * p<0.1.

6 Conclusion

Do revisions reveal mistakes? We find that indeed revised choices improve welfare according to all our normative benchmarks. Revealed preference analysis suggests further that these revisions are closer to being generated by a strictly increasing utility function. Revised behavior is therefore more consistent with models that assume individuals have complete and transitive preferences over all alternatives. Thus, choices which are later revised are likely to be mistakes.

What lessons can we learn from detecting mistakes? One lesson is that mistakes are common, meaningful, and potentiality make it more difficult to observe preferences. Fortunately, adherence to how we believe individuals *ought* to behave improves with a simple prompt to revise. A second lesson is that mistakes are made when the outcomes are unlikely and when the environment is unfamiliar. Choice sets with these characteristics may be more difficult for decision makers to choose from. A third lesson is that reminders make revisions less likely, highlighting a potential tradeoff between the desire for consistency and choosing what one prefers in the moment. Whether demand effects, status quo bias, or memory is behind this discrepancy remains an open question.

Our results should not be read as a refutation of the core revealed preference hypothesis—that individuals have stable preferences. Mistakes are made, but identifying them is possible. Properly accounting for these inconsistencies improves the ability of utility functions to summarize observed behavior *as if* it is consistent with this hypothesis. Future applications can benefit from detecting and limiting these types of mistakes in order to draw more robust inferences about economic models.

References

Afriat, S. N. (1967). The construction of utility functions from expenditure data. *International economic review 8*(1), 67–77.

Afriat, S. N. (1973). On a system of inequalities in demand analysis: An extension of the classical method. *International economic review*, 460–472.

Andreoni, J. and W. Harbaugh (2009). Unexpected utility: Five experimental tests of preferences for risk. *Unpublished Manuscript*.

Andreoni, J. and J. Miller (2002). Giving according to garp: An experimental test of the consistency of preferences for altruism. *Econometrica 70*(2), 737–753.

Benjamin, D., M. Fontana, and M. Kimball (2019). Reconsidering risk aversion. *Presentation: Economic Science Association North American Meeting*.

Bronars, S. G. (1987). The power of nonparametric tests of preference maximization. *Econometrica: Journal of the Econometric Society*, 693–698.

Caplin, A., M. Dean, and D. Martin (2011). Search and satisficing. *American Economic Review 101*(7), 2899–2922.

Chen, D. L., M. Schonger, and C. Wickens (2016). otree: an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance 9*(1), 88–97.

Choi, S., R. Fisman, D. Gale, and S. Kariv (2007). Consistency and heterogeneity of individual behavior under uncertainty. *American Economic Review 97*(5), 1921–1938.

Choi, S., S. Kariv, W. Müller, and D. Silverman (2014). Who is (more) rational? *American Economic Review 104*(6), 1518–50.

Feldman, P. and J. Rehbeck (2019). Revealing a preference for mixing: An experimental study of risk. *Unpublished Manuscript*.

Gaudeul, A. and P. Crosetto (2019). Fast then slow: A choice process explanation for the attraction effect.

Halevy, Y., D. Persitz, and L. Zrill (2018). Parametric recoverability of preferences. *Journal of Political Economy 126*(4), 1558–1593.

Hong, C. S., E. Karni, and Z. Safra (1987). Risk aversion in the theory of expected utility with rank dependent probabilities. *Journal of Economic theory 42*(2), 370–381.

Houtman, M. and J. Maks (1985). Determining all maximal data subsets consistent with revealed preference. *Kwantitatieve methoden 19*(1), 89–104.

Jacobson, S. and R. Petrie (2009). Learning from mistakes: What do inconsistent choices over risk tell us? *Journal of risk and uncertainty 38*(2), 143–158.

Kalaycı, K. and M. Serra-Garcia (2016). Complexity and biases. *Experimental Economics 19*(1), 31–50.

Machina, M. J. (1989). Dynamic consistency and non-expected utility models of choice under uncertainty. *Journal of Economic Literature 27*(4), 1622–1668.

Martínez-Marquina, A., M. Niederle, and E. Vespa (2019). Failures in contingent reasoning: The role of uncertainty. *American Economic Review 109*(10), 3437–74.

Nielsen, K. and J. Rehbeck (2019). When choices are mistakes. *Available at SSRN 3481381*.

Nishimura, H., E. A. Ok, and J. K.-H. Quah (2017). A comprehensive approach to revealed preference theory. *American Economic Review 107*(4), 1239–63.

Polisson, M., J. K.-H. Quah, and L. Renou (2020). Revealed preferences over risk and uncertainty. *American Economic Review 110*(6), 1782–1820.

Rehbeck, J. (2020). How to compute the largest number of rationalizable choices. *Available at SSRN 3542493*.

Varian, H. R. (1982). The nonparametric approach to demand analysis. *Econometrica: Journal of the Econometric Society*, 945–973.

[Supplemental Material]

A Revealed Preference Computation and Results

In this section we show why our revealed preference tests are valid and how we compute them. Our results use the basic intuition from Polisson et al. (2020). However our results are different as our choice environment is different. In particular, subjects in our experiment choose both p and x which differs from the choice environment they consider, outcomes x_1 and x_2 with fixed likelihoods. In their environment both consumption goods are measured in the same units and both both enter the bernoulli utility function in the same manner.

For the general existence of a utility function and its Afriat Index, we implement the tests as described in Nishimura et al. (2017). Although Afriat's theorem only assumes local non satiation our results extend trivially to first order stochastic dominance in our environment. First, if all choices are at the interior of a budget, then our strictly revealed preference relation is the same as in Afriat's original theorem. In the case of a corner choice, the subject is encoded as having an Afriat Index of 0. Figure 11 illustrates the distinction between AIs using a violation of WARP.

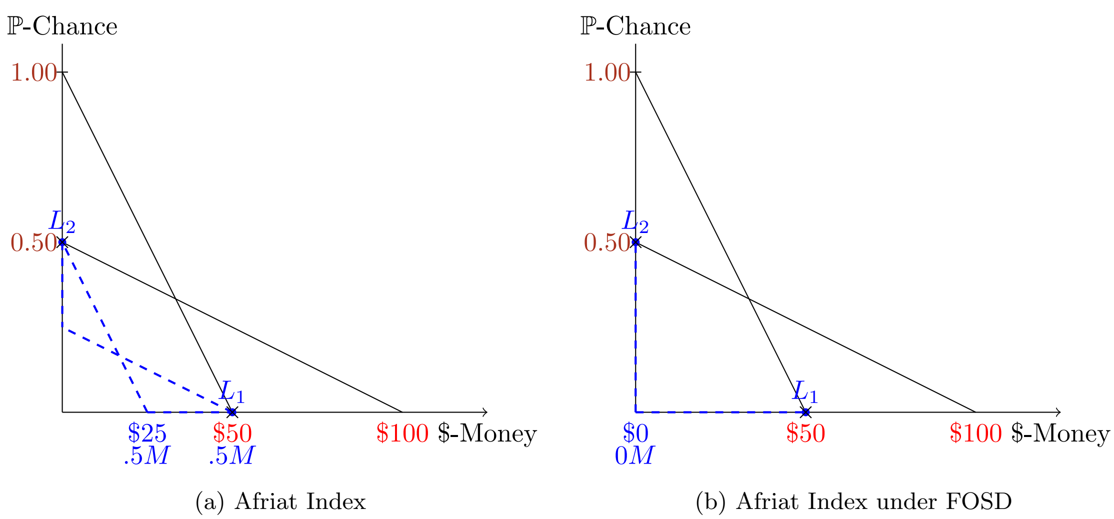


Figure 11: Violations of WARP and their Afriat Indices

To simplify exposition we first describe the validity of the proofs in terms of two abstract commodities x_1 and x_2 and for some utility specification $U(x_1, x_2) = u_1(x_1) + u_2(x_2)$. All of the utility specifications we test, are of the form $U(p, x) = \pi(p) * u(x)$, which is ordinally equivalent to

$U(p,x) = log(\pi(p)) + log(u(x))$. Hence, $x_1 = p$, $x_2 = x$, $u_1 = log(\pi)$ and $u_2 = log(u)$. Our results are organized by decreasing generality, we first discuss probability weighting and then expected utility, which imposes further restrictions on π being an identity function.

Let the $\mathcal{X} \subseteq \mathbb{R}^2_+$ be the consumption space. Define the set of observation, choices and budgets, to be $\mathcal{O} = \{x^t, B^t\}_{t=1}^T$. Now define the downward closure of a budget set by $\underline{B}^t = \{y \in \mathbb{R}^2_+ : y \leqslant x \text{ for some } x \in B^t\}$. Also let $x_i(x_j, B^t) : \mathbb{R}_+ \times \mathcal{X} \to \mathbb{R}_+$ be the i coordinate of element x_j on budget B^t or the origin if $(x_j, 0)$ is not in $\underline{B}^t$.

Our generalized restriction of infinite domains (GRID) consists of

$$\left\{ x^t \cup \bigcup_{s=1}^{T} \left(x_1^t, x_2(x_1^t, B^s)\right) \cup \bigcup_{s=1}^{T} \left(x_1(x_2^t, B^s), x_2^t\right) \right\}_{t=1}^{T} \cup \{0\}.$$

Theorem A.1. ***(Sufficiency of $\mathcal{G}$)*** *There exist a strictly increasing and continuous function that rationalizes $U(a,b) = u_1(a) + u_2(b)$ that rationalizes $\mathcal{O}$ on $\mathcal{X}$ if and only if there exists $\bar{U}(a,b) = \bar{u}_1(a) + \bar{u}_2(b)$ increasing function that rationalizes $\mathcal{O}$ on $\mathcal{X} \cap \mathcal{G}$.*

Proof. Clearly, if U rationalizes $\mathcal{O}$ on $\mathcal{X}$ and is strictly increasing, then it also rationalizes $\mathcal{O}$ on $\mathcal{X} \cap \mathcal{G}$.

For the converse, let be strictly increasing functions $\bar{u}_1$ and $\bar{u}_2$ that rationalize $\mathcal{O}$ on $\mathcal{X} \cap \mathcal{G}$. Suppose that x_1' and x_1'' are both numbers such that elements of the grid have these numbers as their first dimension, $x_1' < x_1''$ and no element of the grid has first dimension which is between these numbers. We define $\hat{u}_1$ as an extension of $\bar{u}_1$ such that for ε near zero,

$$\hat{u}_1(x) = \begin{cases} \bar{u}(x_1') + \varepsilon(x - x_1') \text{ for } x \in [x_1', x_1'' - \varepsilon] \\ \bar{u}(x_1'') + \left(\frac{\bar{u}(x_1'') - \bar{u}(x_1') - \varepsilon(x_1'' - \varepsilon - x_1')}{\varepsilon}\right)(x - x_1'') \end{cases}.$$

$\hat{u}_1$ is a continuous and increasing piecewise linear extension of $\bar{u}_1$ which approaches the "step-function" extension of $\bar{u}_1$ as $\varepsilon \to 0$. Define $\hat{u}_2$ as a similar piecewise linear extension of $\bar{u}_2$, and $\hat{U}(x) = \hat{u}_1(x_1) + \hat{u}_2(x_2)$. For ε small enough, $\hat{U}$ rationalizes $\mathcal{O}$ on $\mathcal{X}$. To see this, note that at all points which are on the budget line but not in the grid, the marginal rate of substitution approaches either zero or infinity as $\varepsilon \to 0$. Thus, there will always be a point on the grid which is preferred to a point which is not on the grid. Since $\hat{U}$ extends $\bar{U}$ (which rationalizes $\mathcal{O}$ on the grid), $\hat{U}$ must rationalize $\mathcal{O}$ on $\mathcal{X}$. □

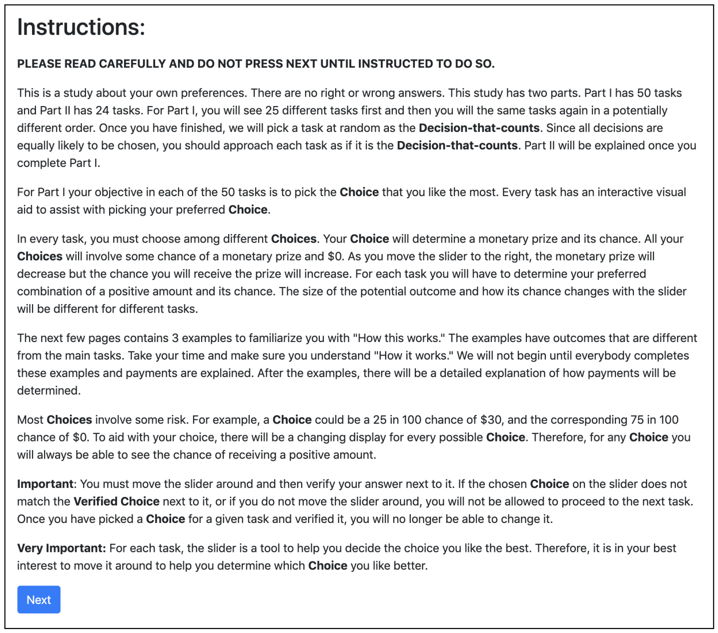
We use two additional observations for the results given in the paper. First, it is straightforward to extend these results to budgets "scaled" by the index e. In this case, the generalized restriction of infinite domains (GRID) consists of

$$\left\{ x^t \cup \bigcup_{s=1}^{T} \left(x_1^t, x_2(x_1^t, e \times B^s) \right) \cup \bigcup_{s=1}^{T} \left(x_1(x_2^t, e \times B^s), x_2^t \right) \right\}_{t=1}^{T} \cup \{0\}.$$

Second, to test the expected utility model rather than the probability weighting model, it is sufficient to restrict $\pi(p) = p$ (and thus $u_1(p) = log(p)$).

B Experimental Instructions

The full set of instructions appears below.



Instructions:

PLEASE READ CAREFULLY AND DO NOT PRESS NEXT UNTIL INSTRUCTED TO DO SO.

This is a study about your own preferences. There are no right or wrong answers. This study has two parts. Part I has 50 tasks and Part II has 24 tasks. For Part I, you will see 25 different tasks first and then you will the same tasks again in a potentially different order. Once you have finished, we will pick a task at random as the **Decision-that-counts**. Since all decisions are equally likely to be chosen, you should approach each task as if it is the **Decision-that-counts**. Part II will be explained once you complete Part I.

For Part I your objective in each of the 50 tasks is to pick the **Choice** that you like the most. Every task has an interactive visual aid to assist with picking your preferred **Choice**.

In every task, you must choose among different **Choices**. Your **Choice** will determine a monetary prize and its chance. All your **Choices** will involve some chance of a monetary prize and $0. As you move the slider to the right, the monetary prize will decrease but the chance you will receive the prize will increase. For each task you will have to determine your preferred combination of a positive amount and its chance. The size of the potential outcome and how its chance changes with the slider will be different for different tasks.

The next few pages contains 3 examples to familiarize you with "How this works." The examples have outcomes that are different from the main tasks. Take your time and make sure you understand "How it works." We will not begin until everybody completes these examples and payments are explained. After the examples, there will be a detailed explanation of how payments will be determined.

Most **Choices** involve some risk. For example, a **Choice** could be a 25 in 100 chance of $30, and the corresponding 75 in 100 chance of $0. To aid with your choice, there will be a changing display for every possible **Choice**. Therefore, for any **Choice** you will always be able to see the chance of receiving a positive amount.

Important: You must move the slider around and then verify your answer next to it. If the chosen **Choice** on the slider does not match the **Verified Choice** next to it, or if you do not move the slider around, you will not be allowed to proceed to the next task. Once you have picked a **Choice** for a given task and verified it, you will no longer be able to change it.

Very Important: For each task, the slider is a tool to help you decide the choice you like the best. Therefore, it is in your best interest to move it around to help you determine which **Choice** you like better.

Next

Figure 12: General Instructions

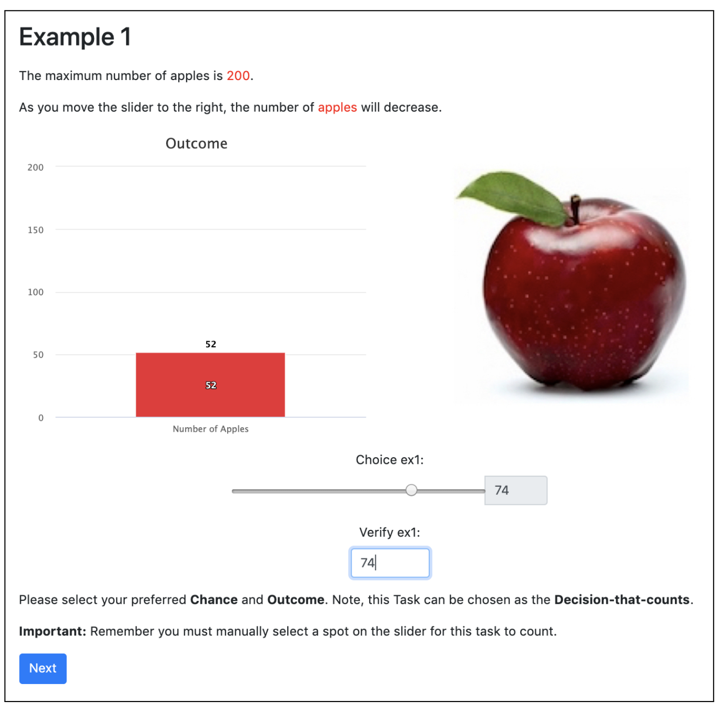


Figure 13: First Example

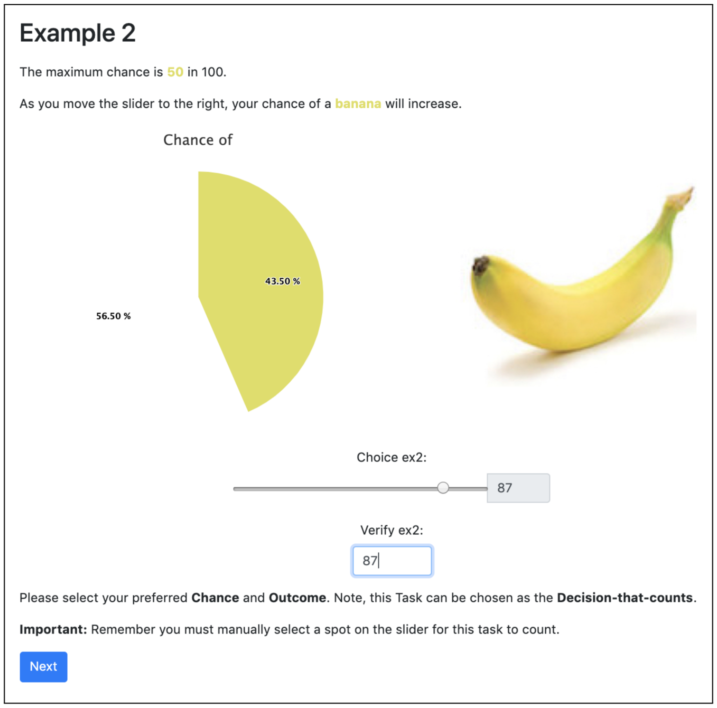


Figure 14: Second Example

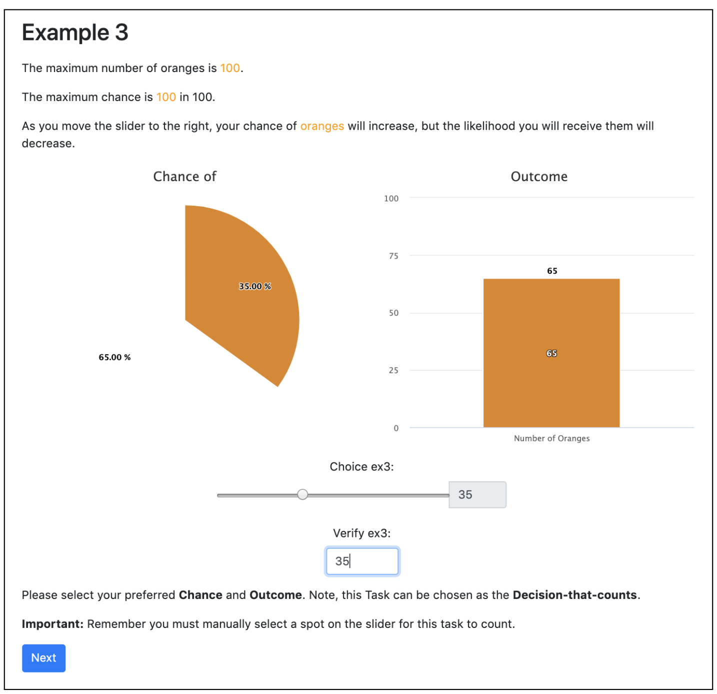
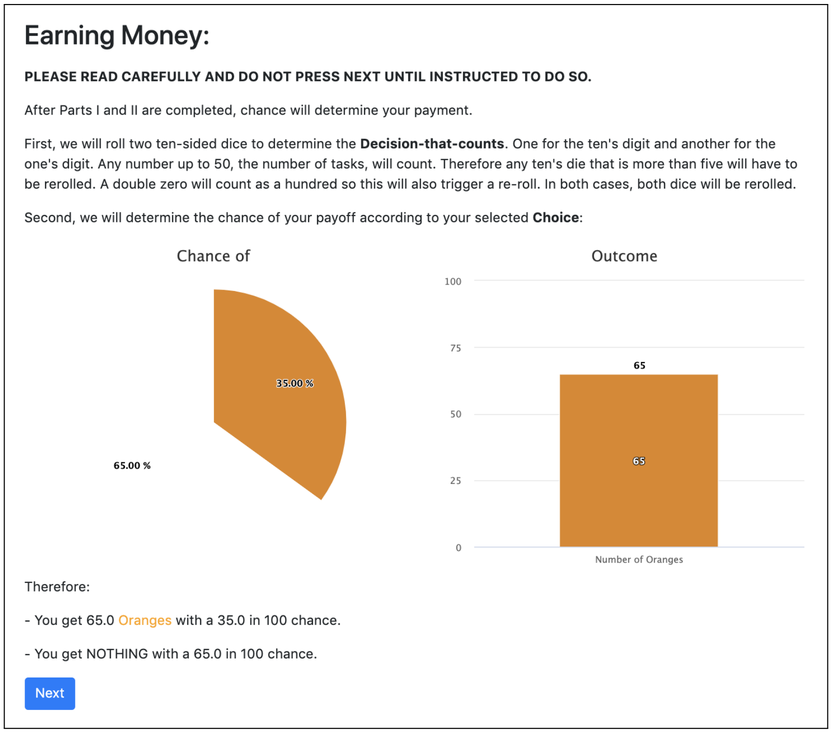


Figure 15: Third Example



Earning Money:

PLEASE READ CAREFULLY AND DO NOT PRESS NEXT UNTIL INSTRUCTED TO DO SO.

After Parts I and II are completed, chance will determine your payment.

First, we will roll two ten-sided dice to determine the **Decision-that-counts**. One for the ten's digit and another for the one's digit. Any number up to 50, the number of tasks, will count. Therefore any ten's die that is more than five will have to be rerolled. A double zero will count as a hundred so this will also trigger a re-roll. In both cases, both dice will be rerolled.

Second, we will determine the chance of your payoff according to your selected **Choice**:

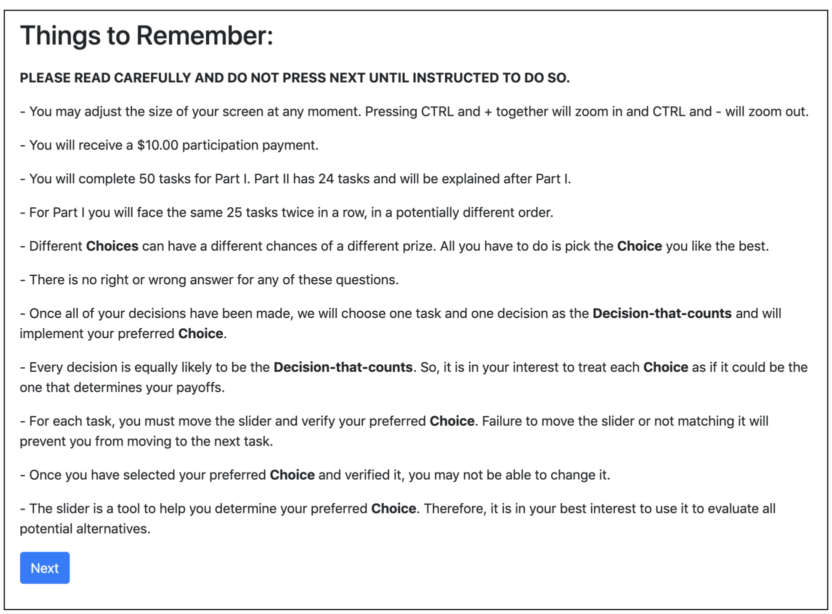
Therefore:

- You get 65.0 Oranges with a 35.0 in 100 chance.

- You get NOTHING with a 65.0 in 100 chance.

Next

Figure 16: Earnings



Things to Remember:

PLEASE READ CAREFULLY AND DO NOT PRESS NEXT UNTIL INSTRUCTED TO DO SO.

- You may adjust the size of your screen at any moment. Pressing CTRL and + together will zoom in and CTRL and - will zoom out.

- You will receive a $10.00 participation payment.

- You will complete 50 tasks for Part I. Part II has 24 tasks and will be explained after Part I.

- For Part I you will face the same 25 tasks twice in a row, in a potentially different order.

- Different **Choices** can have a different chances of a different prize. All you have to do is pick the **Choice** you like the best.

- There is no right or wrong answer for any of these questions.

- Once all of your decisions have been made, we will choose one task and one decision as the **Decision-that-counts** and will implement your preferred **Choice**.

- Every decision is equally likely to be the **Decision-that-counts**. So, it is in your interest to treat each **Choice** as if it could be the one that determines your payoffs.

- For each task, you must move the slider and verify your preferred **Choice**. Failure to move the slider or not matching it will prevent you from moving to the next task.

- Once you have selected your preferred **Choice** and verified it, you may not be able to change it.

- The slider is a tool to help you determine your preferred **Choice**. Therefore, it is in your best interest to use it to evaluate all potential alternatives.

Next

Figure 17: Reminders

Set of Unique Tasks:

PLEASE READ CAREFULLY AND DO NOT PRESS NEXT UNTIL INSTRUCTED TO DO SO.

- The set of tasks you will complete appears below.

- The column on the left gives the maximum amount of money available for each task

- The column on the right gives the maximum amount of chance available for each task

- Your **Choice** in each task will be some combination of this maximum outcome and chance

Maximum Chance	Maximum Outcome
8 in 100 chance	$140
8 in 100 chance	$160
8 in 100 chance	$180
8 in 100 chance	$200
24 in 100 chance	$120
24 in 100 chance	$140
24 in 100 chance	$160
24 in 100 chance	$180
24 in 100 chance	$200
50 in 100 chance	$80
50 in 100 chance	$100
50 in 100 chance	$120
50 in 100 chance	$140
50 in 100 chance	$160
50 in 100 chance	$180
50 in 100 chance	$200
76 in 100 chance	$80
76 in 100 chance	$100
76 in 100 chance	$120
76 in 100 chance	$140
76 in 100 chance	$160
92 in 100 chance	$80
92 in 100 chance	$100
92 in 100 chance	$120
92 in 100 chance	$140

Next

Figure 18: Full Set of Budgets

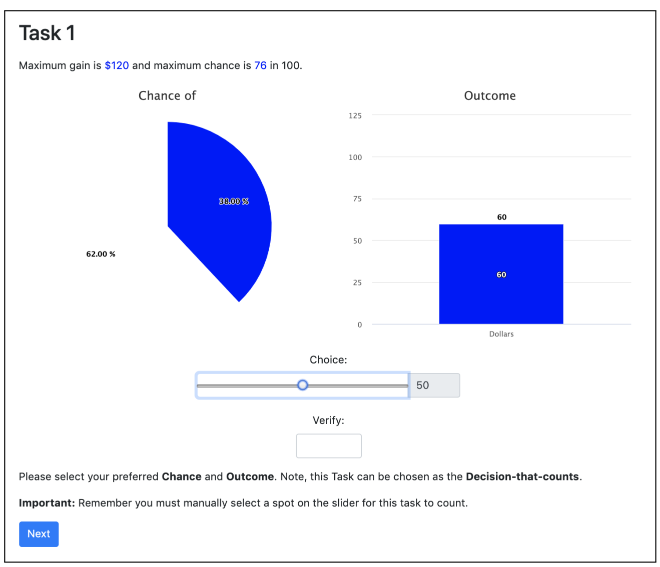
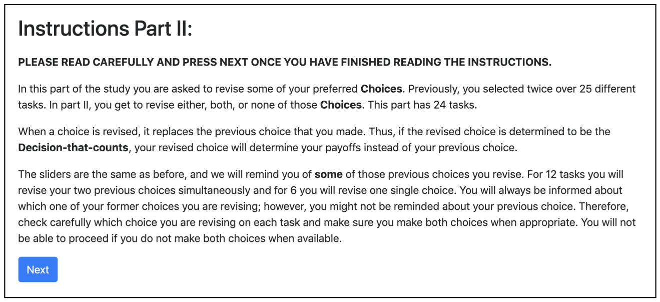


Figure 19: Sample Task



Instructions Part II:

PLEASE READ CAREFULLY AND PRESS NEXT ONCE YOU HAVE FINISHED READING THE INSTRUCTIONS.

In this part of the study you are asked to revise some of your preferred **Choices**. Previously, you selected twice over 25 different tasks. In part II, you get to revise either, both, or none of those **Choices**. This part has 24 tasks.

When a choice is revised, it replaces the previous choice that you made. Thus, if the revised choice is determined to be the **Decision-that-counts**, your revised choice will determine your payoffs instead of your previous choice.

The sliders are the same as before, and we will remind you of **some** of those previous choices you revise. For 12 tasks you will revise your two previous choices simultaneously and for 6 you will revise one single choice. You will always be informed about which one of your former choices you are revising; however, you might not be reminded about your previous choice. Therefore, check carefully which choice you are revising on each task and make sure you make both choices when appropriate. You will not be able to proceed if you do not make both choices when available.

Next

Figure 20: Instructions Part 2

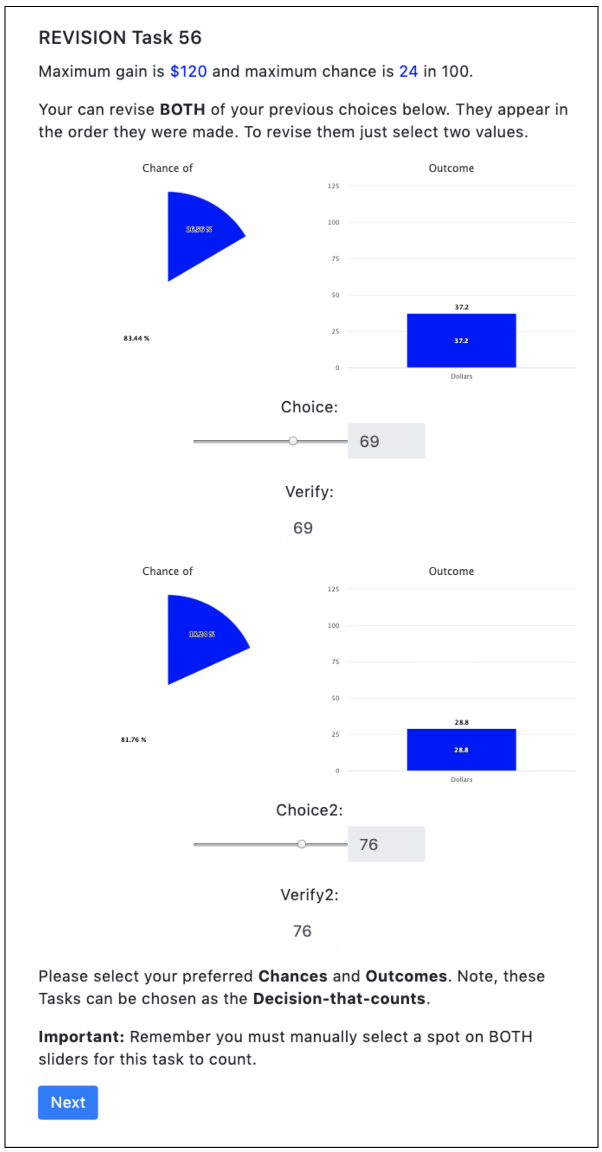


Figure 21: Revisions without Reminders

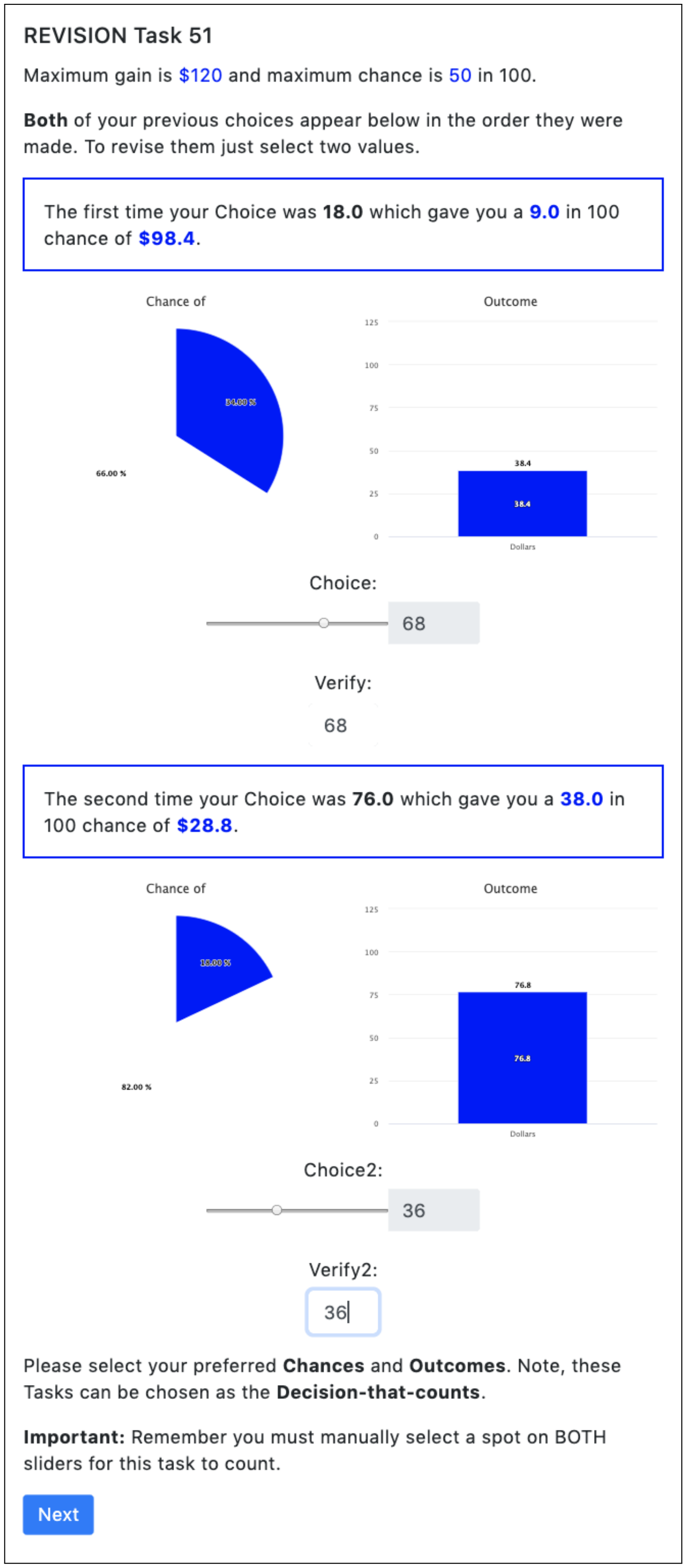


Figure 22: Revisions with Reminders

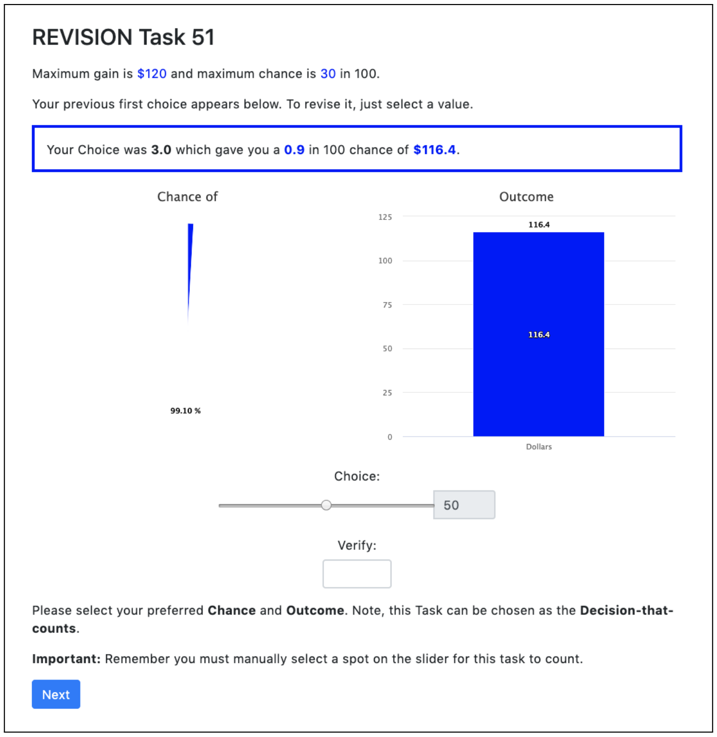


Figure 23: One Revision with Reminders

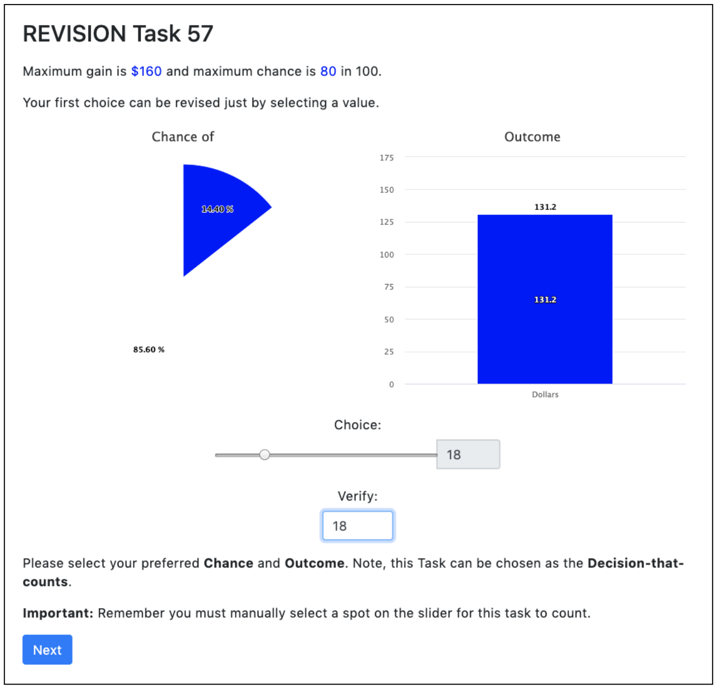


Figure 24: One Revision without Reminders