When Choices Are Mistakes*

Kirby Nielsen† John Rehbeck‡

Draft: April 6, 2022

Abstract

Using a laboratory experiment, we identify whether decision-makers consider it a mistake to violate canonical choice axioms. To do this, we incentivize subjects to report axioms they want their decisions to satisfy. Then, subjects make lottery choices which might conflict with their axiom preferences. In instances of conflict, we give subjects the opportunity to re-evaluate their decisions. We find that many individuals want to follow canonical axioms and revise their choices to be consistent with the axioms. In a shorter online experiment, we show correlations of mistakes with response times and measures of cognition.

KEYWORDS: mistakes; rationality; normative versus descriptive theory; procedural decision making

*We thank Marina Agranov, Dan Benjamin, Doug Bernheim, Christopher P. Chambers, Josh Dean, Paul Feldman, Shane Frederick, Tzachi Gilboa, Ben Grodeck, Yoram Halevy, Paul J. Healy, Miles Kimball, Jennifer Manly, Muriel Niederle, Ryan Oprea, Lee Pang, Collin Raymond, Frank Schilbach, Joel Sobel, and Colin Sullivan for many helpful comments and suggestions. We also thank the editor, coeditor, and three anonymous referees for their very helpful comments and suggestions. A previous version of this paper was circulated with the title "Are Axioms Normative? Eliciting Axiom Preferences and Resolving Conflicts with Lottery Choices." This study was approved by the Institutional Review Boards at Stanford University and The Ohio State University. We thank the National Science Foundation for research support; this work was sponsored by NSF-2049749 and NSF-2049748.

†kirby@caltech.edu; Division of the Humanities and Social Sciences, Caltech

‡rehbeck.7@osu.edu; Department of Economics, The Ohio State University

> In reversing my preference... I have corrected an error. There is, of course, an important sense in which preferences, being entirely subjective, cannot be in error; but in a different, more subtle sense they can be.
>
> — Leonard Savage (1954)

I. Introduction

An enormous experimental literature—spanning at least six decades—has shown that individuals consistently violate canonical axioms in decision theory.[1] However, the literature has remained relatively silent on whether these violations are intentional deviations from the axioms or are simply "mistakes." When an individual violates an axiom but would not have done so had they *known* they were violating the axiom, we call the violation a mistake.[2] If violations of canonical axioms stem mainly from mistakes rather than from intentional deviations, then we can maintain confidence in the *normative* content of the theory, despite the fact that the theory is not *descriptively* accurate. However, if an individual violates the axioms because they do not *want* to follow them, then one should look for other "behavioral axioms" that the individual agrees with. In this paper, we develop an incentivized experimental framework designed to detect the "subtle sense" in which individuals make mistakes in risky choice, as mentioned by Savage (1954).

Empirically identifying a mistake requires three pieces of information, reflected in the three main parts of our experiment. First, we elicit the axioms an individual wants their choices to satisfy. Eliciting preferences over axioms directly allows us to identify when an individual prefers the axiom as a principle governing *all* of

[1] Examples include May (1954); MacCrimmon (1968); Tversky (1969); Slovic and Tversky (1974); Kahneman and Tversky (1979); Huber et al. (1982); Segal (1988); Loomes et al. (1991); Wedell (1991); Loomes et al. (1992); Camerer (1995); Birnbaum and Chavez (1997); Seidl (2002); Birnbaum and Martin (2003); Birnbaum and Schmidt (2008); Regenwetter et al. (2011); Birnbaum et al. (2016), among many others.

[2] Note that by "intentional deviation" we do not necessarily mean that the deviation was conscious. For example, individuals may view an axiom as a description of internal judgement that is independent of their choices. Mistakes may also occur from random errors as in Thurstone (1927), Luce (1959), or McFadden (1973). We remain agnostic on the source of mistakes, but define a mistake as a violation of an axiom that would not be maintained after the decision-maker understands the full implications of the axiom and their choices.

their choices, not just in specific instances. We incentivize this decision by asking individuals whether they would prefer to make a choice on their own or instead have the axiom choose for them. Second, we present decision problems where the individual is likely to violate an axiom they wanted to satisfy. This part is most similar to standard choice experiments. Finally, we observe how individuals perceive this inconsistency in their choices, and whether/how they reconcile their conflicting preferences.[3] Since we elicit both the preference for the axiom and the related lottery choices, we can present subjects with inconsistencies in *their own* preferences which mitigates experimenter demand effects. This reconciliation opportunity provides individuals with strictly more information about the implications of the axioms; they see their axiom preference and their lottery choices that violate the axiom at the same time. Given this, we take the position that these reconciled choices are more reflective of the preferences an individual *wants* to express.

We examine six fundamental axioms in the domain of risk—independence of irrelevant alternatives, first order stochastic dominance, transitivity, independence, branch independence, and consistency. We focus attention on the domain of lotteries, and on these axioms in particular, since many papers have shown violations of these axioms and some papers suggest that violations are a mistake while others suggest that they reflect underlying preferences. While we start in this simple domain, we emphasize that our methods could be applied to most axioms or decision-making procedures. We describe some examples from different environments in Section VII.C.

We find that subjects want to follow these axioms at high rates—around 85% of subjects desire an axiom to make choices on their behalf. This gives strong ex-ante evidence that individuals view axioms as normative principles. However, as in previous experiments, subjects often violate these axioms in their lottery choices. We find that subjects who prefer the axiom to make choices on their behalf violate the axiom at similar rates as subjects who prefer to choose on their own. This implies that wanting choices to satisfy an axiom does not predict adherence to the axiom.

When subjects' axiom and lottery choices are inconsistent, we give them the op-

[3] Relative to existing research, we collect data for each stage above and all choices are incentivized. Furthermore, we examine preferences over axioms as global preferences rather than preferences in a specific choice problem. MacCrimmon (1968); Moskowitz (1974); Slovic and Tversky (1974); MacCrimmon and Larsson (1979) are the closest papers to ours in the existing literature. These papers only collect some of this information, or only in specific choice problems, or are not incentivized. Further discussion of these papers is in Section VI.

portunity to reconcile this inconsistency. Subjects are not required to reconcile their choices, but if they choose to do so, then they can reconcile their choices by changing their lottery decisions, by declaring they no longer want their choices to obey the axiom, or by doing a combination of these. Aggregating across axioms, we find that individuals change their lottery choices to be consistent with the axiom in 47% of violations, while they renounce the axiom in only 13% of violations. We interpret this 47% of violations as subjects treating the axioms as normative and viewing their lottery choices as mistakes. Just over one-third of violations are kept inconsistent, with subjects maintaining their lottery choices while still stating a desire to follow the axiom. We discuss this puzzle and possible interpretations in Section IV.

A major concern in this type of experiment is that of experimenter demand effects or other psychological concerns pushing subjects toward selecting axioms. To isolate this concern, we include "control axioms," which are the "opposite" of each of our axioms of interest. For example, we present subjects with the rule $c-$transitivity (control-transitivity), which says, "If A is preferred to B and B is preferred to C, then C is preferred to A." We designed these axioms to be intentionally normatively unappealing so that we can cleanly identify the extent to which demand effects and other motivations drive axiom selection.

Subjects are much less likely to select the control axioms, doing so only about 10% of the time (compared to 85% for the axioms). This suggests that subjects are not simply agreeing with all axioms presented to them. Furthermore, we find in aggregate that subjects are much more likely to renounce the control axioms than axioms in the reconciliation stage. Further details on the role of the control axioms and alternative design choices are discussed in Section VI. We also discuss how our approach of identifying mistakes relates to other approaches in the literature in Section VI.

While our results suggest that individuals do prefer to follow these fundamental axioms and that violations are often mistakes, we exercise modesty in generalizing our results. We do not make general conclusions that violations of the axioms in question are, *definitively*, mistakes. Just as it has taken decades to show where axioms are violated, it will take much more work to show where and when these violations are mistakes. Our results are suggestive of the interpretation that violations of canonical axioms *can be* mistakes, and we provide a framework by which to detect these mistakes. We view this paper as one step in a much larger research agenda identifying mistakes and preferences for following choice rules. We describe how these may be welfare-relevant in Section VII.

Furthermore, we are agnostic about how mistakes occur. For example, mistakes might result from decision costs or inattention which are ameliorated in our reconciliation stage. Alternatively, individuals could have a preference for their choices to be consistent with logical principles, even when their organic decisions are not. Whatever the source of the mistakes, our results suggest that choices violating canonical axioms are not necessarily welfare-maximizing since the observed violations could be mistakes. We view our paper as contributing to the literature that identifies principles an individual feels *should* guide their choices and identifies when it is difficult for individuals to follow these principles. This is in a similar spirit to Oprea (2020) who studies what makes a rule complex for individuals to implement.

While we exercise modesty in the conclusions from the experiment, we also view this paper as a methodological contribution and proof of concept that opens the door to a number of future research directions. For example, researchers can use our experimental paradigm to elicit the normative appeal of—and identify mistakes in implementing—axioms, strategies, social choice rules, and many other objects of interest. We purposefully chose simple axioms to study, but one could easily use a similar procedure to study more complicated axioms such as reduction of compound lotteries, the weak axiom of revealed preference, dynamic consistency, or time stationarity, among many others. We view the methods here as a paradigm that can be transplanted to inform other areas of economics. In the Discussion, we highlight other interesting domains where this approach could be informative.

In addition to our laboratory study, we ran a shorter online experiment that focuses on the independence axiom. Our online study demonstrates that it is feasible to include a shorter module at the end of a study to elicit attitudes towards axioms or decision rules.[4] In the online experiment, we collect response times and cognitive reflection test (CRT) scores (Frederick, 2005) to study how these measures of individual cognition interact with axiom preferences, lottery choices, and revisions. We find that individuals with lower CRT scores are more likely to make their choices consistent with the axiom in the reconciliation stage. Individuals who make choices consistent with the axiom also do so very quickly, which could indicate strength of preference. This analysis is only suggestive, and we discuss interpretations in Section V.

[4]We are grateful to the editor and referees for suggesting this experiment.

II. THEORETICAL FRAMEWORK

Before outlining the experimental design, we define the theoretical framework underlying the experiment. We presented all questions and axioms in the domain of non-negative monetary lotteries. We considered lotteries with US dollars as prizes, with potential outcomes in $X = [0,30]$. We represent the set of lotteries with prizes in X by $\Delta(X)$, with strict preferences $\succ$ defined over $\Delta(X)$.[5] We denote generic prizes in X by x, y, z, and denote generic lotteries in $\Delta(X)$ by p, q, r, s. We represent the degenerate lottery giving \$$x$ for sure as δ_x. Lastly, for a set of lotteries, S, we denote the set of lotteries chosen from S as $C(S)$. We write $p \succ q$ to mean $p = C(\{p,q\})$, or p is chosen from the set of $\{p,q\}$.

Throughout the experiment, we study six fundamental axioms:

1. Independence of Irrelevant Alternatives (IIA): $p = C(\{p,q,r\}) \Rightarrow p = C(\{p,q\})$
   IIA states that if a lottery p is chosen from the set of lotteries p, q and r, then it is also chosen from the subset p and q.

2. First Order Stochastic Dominance (FOSD):[6] $\forall x \quad 1-P(x) \geq 1-Q(x) \Rightarrow p \succ q$
   FOSD states that if the probability of winning a prize greater than x is higher in p than in q, for all prizes, then p will be chosen over q.

3. Transitivity (TRANS): $p \succ q$ and $q \succ r \Rightarrow p \succ r$
   TRANS states that if a lottery p is chosen over lottery q, and q is chosen over r, then p will be chosen over r.

4. Independence (IND): $\forall \lambda \in [0,1] \quad p \succ q \Rightarrow \lambda p + (1-\lambda)r \succ \lambda q + (1-\lambda)r$
   IND states that if p is chosen over q, then the mixture of p with any lottery r will be chosen over the equivalent mixture of q with r.[7]

5. Branch Independence (BRANCH): $\lambda p + (1-\lambda)r \succ \lambda q + (1-\lambda)r \Rightarrow \lambda p + (1-\lambda)s \succ \lambda q + (1-\lambda)s$
   BRANCH states that if the mixture of p and r is chosen over the mixture of q and r, then the preference will not change when r is swapped out for a different lottery, s.

6. Consistency (CONS): $p \succ q \Rightarrow p \succ q$
   CONS states that if p is chosen over q, then p always will be chosen over q.

[5] Indifference and other factors such as preference for randomization are important elements of choice, and we cannot identify these in our experiment. We leave this for future work.

[6] Where $P(x)$ and $Q(x)$ are the cumulative distribution functions to x of p and q respectively. For example, $P(x) = \sum_{y \leq x} p(y)$ where $p(y)$ is the probability of winning prize y.

[7] We study mixture independence rather than compound independence (Segal, 1990). This means that $\lambda p + (1-\lambda)r$, for example, is a reduced one-stage lottery in our lottery questions.

In addition to these six main axioms, we included the "opposite" of each axiom (denoted as "control axioms"). The control axioms reverse the preference relation in the consequent of the implication for each of the six main axioms. The control axioms were intentionally unappealing and have the same structure as the corresponding axiom.

Formally, we included the following six control axioms:

1. $c-$Independence of Irrelevant Alternatives ($c-$IIA): $p = C(\{p,q,r\}) \Rightarrow q = C(\{p,q\})$

2. $c-$First Order Stochastic Dominance ($c-$FOSD): $\forall x \quad 1-P(x) \geq 1-Q(x) \Rightarrow q \succ p$

3. $c-$Transitivity ($c-$TRANS): $p \succ q$ and $q \succ r \Rightarrow r \succ p$

4. $c-$Independence ($c-$IND):$\forall \lambda \in [0,1] \quad p \succ q \Rightarrow \lambda q + (1-\lambda)r \succ \lambda p + (1-\lambda)r$

5. $c-$Branch Independence ($c-$BRANCH): $\lambda p + (1-\lambda)r \succ \lambda q + (1-\lambda)r \Rightarrow \lambda q + (1-\lambda)s \succ \lambda p + (1-\lambda)s$

6. $c-$Consistency ($c-$CONS): $p \succ q \Rightarrow q \succ p$

We also designed six meaningless *distractor rules*, which were over unrelated lotteries. For example, one distractor rule is $p \succ q \Rightarrow r \succ s$ where the lotteries p, q, r, and s are unrelated. This rule essentially implements a random choice. We used the distractor rules as a buffer so that subjects were less likely to notice the relationships between the axioms and control axioms. The full list of the distractor rules is in the Supplemental Appendix. When we refer to the axioms, control axioms, or distractor rules as general choice objects, we refer to them as *rules*, which is the language used in the experimental instructions.

We make no assumptions on preferences over simple lotteries except for dominance in degenerate lotteries, i.e. $\delta_x \succ \delta_y$ if and only if $x > y$. In using the random problem selection payment mechanism, we also assume a form of monotonicity in the space of two-stage lotteries (Azrieli et al., 2019).[8] Brown and Healy (2018) give evidence that this condition is met in a risky choice experiment similar to ours.

[8]This is referred to as compound independence in Segal (1990). This is not the same as the independence axiom over monetary lotteries that is elicited from subjects. Thus, even when a subject violates independence on monetary lotteries, this does not have any implications on whether the incentive mechanism is valid over the state-space induced by the questions in the experiment. However, it is possible these preferences are correlated which may induce bias as suggested by Baillon et al. (2021), but further research is needed to understand whether this is an issue in practice.

III. Experimental Design

Identifying a mistake under our definition requires three pieces of information: eliciting an individual's preference over axioms, observing violations of these axioms, and studying how discrepancies in these preferences are reconciled. Our experiment consists of three main blocks to elicit these three pieces of information.[9] First, we overview these blocks and discuss the underlying design choices in each block. We present more details for each block in the following subsections.

III.A. Overview

We summarize the most important design choices and brief reasoning below. We discuss our design choices in light of forgone alternative methods in Section VI.

1. All decisions, including the choice to follow an axiom, are incentivized.
2. We directly elicit an individual's preference over decision rules.
3. Control axioms capture demand effects, confusion, and other latent tendencies to follow rules.
4. The opportunity to reconcile choices is neutral and voluntary, so that one can make changes to axiom choice, lottery choice, both, or have choices remain inconsistent.

In Block 1, we elicit an individual's preferences over decision rules. Eliciting preferences over rules presents many challenges, such as presenting the rules in a clear way, incentivizing subjects' responses, and controlling for demand effects. To help subjects understand the rules, we explained them using simple colored circles rather than using their mathematical expressions. To incentivize selection of a rule we presented them akin to "algorithms" that would make a relevant choice on a subject's behalf. For example, a subject who prefers transitivity, and who chooses A over B and B over C, would have the choice of A over C *automatically made for them* when the transitivity axiom is chosen for payment. If this subject did not select the transitivity axiom, then they would make the choice between A and C on their own. We view this as a high bar for axiom preferences to overcome since individuals are generally

[9] We included an additional module to elicit rankings over axioms and the willingness to pay for the opportunity to reconcile choices. We defer explanation of this to Appendix Section E.

averse to having choices made for them (Owens et al., 2014; Agranov and Ortoleva, 2017).

Finally, to control for experimenter demand effects and other motivations for selecting rules, we include the “opposite” of each axiom, which we refer to as our “control axioms” (denoted “c-axioms”). The purpose of including this is not to conclude that the axioms are more normatively appealing than the c-axioms, since this is fairly straightforward. Instead, we include the c-axioms to demonstrate that rule selection is not driven by blind rule following as a result of experimenter demand, using rules to reduce effort costs, or other considerations outside the ones we induce with our experimental incentives. Differences between selection rates of the axioms and c-axioms suggest that axiom selection cannot be explained merely by experimenter demand effects, subjects not wanting to make choices on their own, responsibility aversion, and so-on since the c-axioms are presented and incentivized in the same manner as the main axioms. The axioms and c-axioms were presented in an ex-ante random order to ensure order effects did not drive choices.

After eliciting preferences over decision rules, in Block 2 we present lottery choices designed to offer the possibility of individuals violating an axiom they wanted to satisfy. Finally, in Block 3, we observe how individuals perceive inconsistencies in their choices, and whether/how they reconcile their conflicting preferences. We assume that the decisions in Part 3 are more reflective of the preferences an individual wishes to express, since individuals have strictly more information about the implications of the rules and can directly observe the rules underlying their lottery choices.

To mitigate experimenter demand effects, we provide subjects with a neutral reconciliation opportunity. That is, subjects could make their choices internally consistent by renouncing the axiom or by changing their lottery choices to be consistent with the axiom. There is no default direction for this reconciliation opportunity. Subjects are also allowed to keep their choices inconsistent if they do not wish to reconcile. This not only allows us to identify a mistake, but we can see, from the subject’s own perspective, whether the mistake was in the axiom choice or in the lottery choice. We also allow subjects to revise inconsistencies with any control axioms they selected.

We describe the different Blocks and payment mechanisms in detail below. To overview the payment mechanism, subjects could be paid for one of four possibilities: original rule choices (Block 1), original lottery choices (Block 2), revised rule choices

(Block 3), or revised lottery choices (Block 3). The incentivization procedures are the same for original and revised rules, and are the same for original and revised lotteries. Rules are incentivized by applying them on a set of lotteries and paying subjects what the rule prescribes selecting. If individuals do not want to follow a rule, then they make the lottery choices themselves. Original and revised lottery choices are incentivized in the standard manner by paying subjects a realization from the lottery they selected. All payment uncertainty was resolved using physical randomization devices, in particular two ten-sided dice. The choice of which question would be paid is based on random chance. Subjects were paid at the end of the experiment, regardless of which decision was selected for payment.

III.B. Block 1: Rule Choices

The objective in Block 1 was to elicit a subject's preferences over canonical choice axioms. The first challenge is incentivizing the rule choice so that subjects select all of the rules they view as desirable and do not select any others. We did this by asking subjects to decide whether they prefer the rule to make a choice for them or whether they would rather make the relevant choice themselves.[10]

If the subject preferred a rule to make decisions for them and the rule was selected as the payoff-relevant decision, then we applied the rule to a set of lotteries where it has implications. The subject was paid a realization of the lottery prescribed by the implications of the rule. If the subject did not select a rule to make decisions for them and the rule was selected as the payoff-relevant decision, then they would make the relevant choice on their own.[11]

For example, if IIA were chosen for payment, then we would present the subject with a choice set $\{p, q, r\}$ and would ask them to choose their most preferred lottery. Denote the chosen lottery by p. The subject would be paid from the binary decision problem involving the chosen lottery and some other lottery, e.g., $\{p, q\}$. If the subject chose IIA to make decisions on their behalf, then we would automatically implement the choice of p over q for them, as prescribed by IIA, and would pay them a realization

[10]Subjects were not allowed to choose between these two options until at least 30 seconds had passed. This design feature encourages subjects to consider the rules carefully before deciding.

[11]Note, this means that subjects who do not select a rule must make one additional decision, and subjects may wish to avoid this. This additional decision is true for both our axioms and control axioms, so while it could lead to increased rule selection, it should not affect the difference between axiom and control axiom selection rates.

of the lottery p. If the subject did not choose IIA to make decisions on their behalf, then we would present them with the choice set $\{p, q\}$ and would pay them whichever lottery they choose from this set.[12]

Individuals made independent decisions across the axiom and c-axioms. For example, a subject decided whether to have IIA make a choice for them or instead make the choice on their own; they separately decided whether to have c-IIA make a choice for them or instead make the choice on their own. As a result, a subject could follow both IIA and c-IIA, neither IIA nor c-IIA, or could follow exactly one of them. This implies that a subject who wishes to follow IIA sometimes, but not always, would choose to follow *neither* rule. This way, they could make their own decision rather than having either IIA or c-IIA decide for them. Given our independent incentivization of the rules, it is not the case that a desire to violate an axiom implies a desire to adhere to the corresponding c-axiom, or vice versa.[13]

Under our incentive scheme, a subject is incentivized to select the rule as long as the cost of making a decision is not greater than their expected loss in utility from following the rule in situations where they would not actually like to follow the rule. Under the assumption of no decision costs, an individual would select the rule only in the event that they want to follow it in *all* possible instances. If decisions are costly, however, then selecting a rule instead could indicate that the subject views the rule as mostly—but not always—true. While we believe that decisions are not cognitively costless, the difference in selection rates between our axioms and c-axioms reported below suggests that this is not a main factor in rule selection.

A subject who selects a rule reveals that they want to make decisions according to the rule, since it can be applied over any lotteries in the domain. However, the interpretation is less clear for subjects who do not select a rule. A subject who agrees with a rule but believes their choices will align with the rule anyway has no strict incentive to select the rule, aside from the time and effort cost of making choices on

[12]When a subject was paid for their rule choice at the end of the experiment, they were not told which rule was being implemented. If we had told them which rule was being implemented, then they could answer the initial choices "opposite" their true preferences for the control axioms and still receive their truly-preferred alternative. For example, a subject who truly prefers p over q but knows that they are being paid for $c-$Consistency could pick q over p, knowing that the rule picks the other lottery to determine their payment.

[13]We clearly communicated this to subjects: "If you think the rule should describe your choices, you should select it... If you think there are situations where the rule would not give you your favorite option, you should not select it." Furthermore, the axiom and c-axiom were presented on separate screens in random order.

their own. In Appendix Section D, we present results from another treatment where subjects had to pay a small cost, \$1, to make the choice on their own.[14] We find that the rules are selected slightly more often in this treatment, responding to the incentives, but all qualitative results remain unchanged.

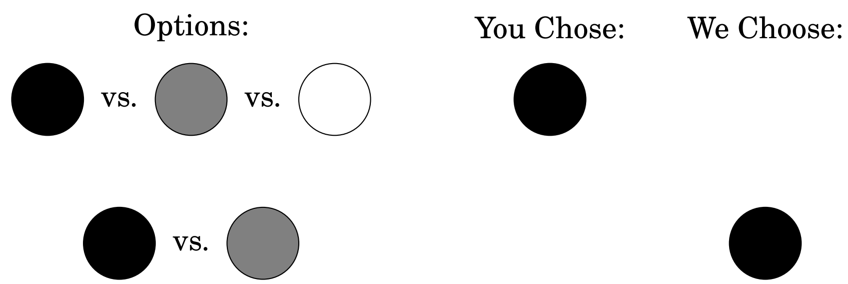


Figure I: Rule representation of IIA

Notes: We represent rules as above. Colored circles represent any possible lotteries with payoffs from \$0–\$30. We also included a written description of the rule on the subjects' screens under the abstract depiction. Subjects choose whether to have this rule make choices for them or instead make choices on their own.

The second challenge to elicit preferences over rules lies in making the domain of the rules accessible and easy for subjects to understand while retaining their broad implications on choices. We presented the decision rules using simple pictorial logic statements with lotteries represented by colored circles. Subjects were told that the colored circles represent monetary lotteries but they did not know the exact lotteries associated with each rule. We inform subjects that the lotteries could have payoffs from \$0–\$30, with any probabilities from 0%–100%. Again, we use IIA as an example to show how we present the rules to subjects in Figure I. In Appendix Section F, we show how we represent the other five axioms in rule format. We explained mixtures of lotteries to subjects using examples. Subjects made eighteen total axiom, control axiom, and distractor rule decisions in Block 1, and the order of these decisions was randomized ex-ante.

Our instructions, included in the Appendix, included many examples of rules. None of the rules used in the experiment were included in the instructions in order to

[14]This makes it strictly costly to not select a rule, eliminating this concern. Here, however, the interpretation is less clear for subjects who do select a rule. A subject who selects a rule does not necessarily indicate they *always* want to follow the rule. It could be that they want to follow the rule "most" of the time, so in expectation, they believe it is not worth paying \$1 to make choices on their own. One could also interpret this as an additional \$1 bound on decision-making costs.

avoid introducing any bias in subjects' choices over the rules of interest. In addition, we clearly communicated to subjects that there were no right or wrong answers in their rule selection choices (and in all other decisions throughout the experiment).

III.C. Block 2: Lottery Choices

Given that our main interest is in studying how individuals reconcile inconsistent choices, we selected lottery questions from previous papers that found violations of the axioms. We do not focus on the specifics of the lotteries, but we picked questions to maximize axiom violations. Our intention is not to compare violation and reconciliation rates across axioms since violations can differ in magnitude. The full set of questions and descriptions can be found in Appendix C.

We displayed the lotteries simply by reporting the probabilities and payoffs of each possible outcome, as shown in Figure II. Subjects saw the lotteries on their screens as below and made their choices by selecting the button corresponding to their preferred option. Altogether, subjects make choices from 33 binary or trinary decision problems in Block 2. The order of these choices was randomized ex-ante.

Option A:	**Option B:**
50% chance of $3	25% chance of $5
50% chance of $15	75% chance of $12
Option A	Option B

Figure II: Representation of lotteries

We chose lotteries so that we did not use any lottery to target more than one axiom. For example, the lotteries used in FOSD1 are completely distinct from lotteries used in any other question. This allows us to study violations of a given axiom in isolation without considering the joint implications of the axioms taken altogether.[15]

[15] The one caveat is that some of our IIA questions involve "decoy" lotteries which are related by FOSD. This decoy lottery was selected by only one subject, and we did not include a reconciliation stage to explain this as an FOSD violation.

III.D. Block 3: Reconciliation

After completing the two earlier blocks, we presented subjects with every inconsistency between their lottery choices and selected rules. For example, a subject who selected IIA in Block 1 but violated IIA with their lottery choices in Block 2 saw these choices side by side on their screen.

On subjects' screens, we highlighted the rule that the subject selected and the decisions that they made in relevant lottery questions. We match the subject's lottery choices to the colored circles of the rule when presenting the reconciliation opportunity, so the subject could better understand how the rule mapped onto their choices. We also include a written explanation of why choices violated the rule and how the rule would choose instead. We used neutral language in describing the violations. We phrase any inconsistency in rule and lottery choice by saying that the rule would have chosen something different for the subject than what the subject chose for themselves. We provide a screenshot in the Appendix A and reproduce an example below in Figure III. The language used to describe violations with the c-axioms is identical to the language used to describe violations with the main axioms. We also provide a c-axiom reconciliation screenshot in Appendix A.

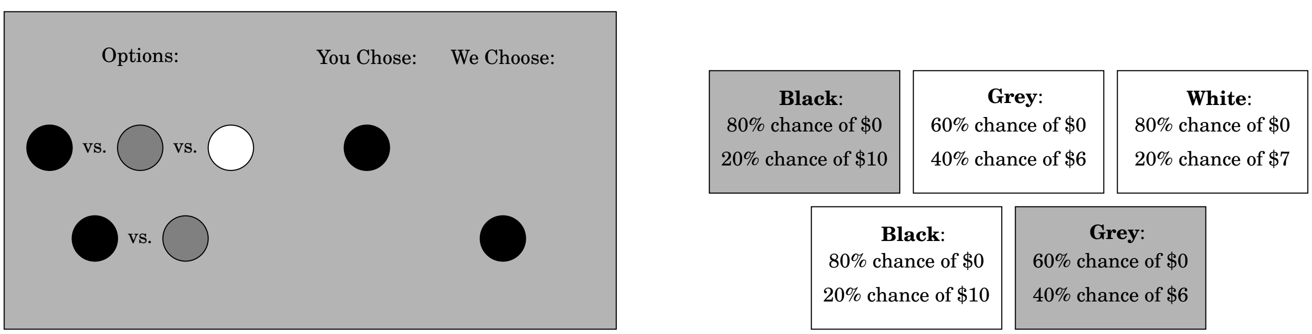


Figure III: Example of Reconciliation Screen
Notes: The options highlighted in grey indicate subjects' original choices in Blocks 1 and 2. For example, this subject selected IIA in Block 1, but chose "black over grey and white" in one question and chose "grey over black" in another question. Below this, subjects saw an explanation of why the rule would have selected something different than what they chose for themselves. In the actual experiment, the circles and highlighting were shown in colors rather than grayscale.

Subjects could change *any* of their choices, or could leave them as they were. We impressed upon subjects that they could change any of their lottery choices, could unselect the rule, could do both, or could leave choices inconsistent. For example, suppose, as in Figure III, an individual selected IIA as a decision rule and then chose lottery p from $\{p,q,r\}$ and q from $\{p,q\}$. The individual could unselect the rule, could change their selection from $\{p,q,r\}$, could change their selection from $\{p,q\}$, could do combinations of these, or could do nothing. As a result, there was no default direction for any potential experimenter demand effect, which is an important feature in our design.

Our key assumption is that when an individual revises their axiom and lottery choices to be consistent, this reveals that the original choice was a mistake. We believe this is a reasonable assumption since the revision opportunity provides the individual with strictly more information about the implications of the axiom and their previous decisions. Thus, we interpret the decisions in Part 3 as better revealing the preferences that an individual wishes to express.

While the reconciliation opportunity occurs on a single screen, any choice on the screen has an independent chance of being selected for payment. For example, consider the reconciliation opportunity for IIA. A subject could be paid for their revised rule choice, their revised choice from $\{p,q,r\}$, or their revised choice from $\{p,q\}$. Each choice on the screen is paid in the same manner as the original choices were paid in Blocks 1 and 2. Furthermore, the subject's original choices from Block 1 and Block 2 were not overturned by this reconciliation opportunity and still could be chosen for payment.[16]

Subjects had the opportunity to reconcile choices inconsistent with each of the six axioms and the six control axioms.[17] Subjects reconciled each violation independently. That is, a subject who selected IIA and violated it on two separate occasions had two separate opportunities to reconcile the violations rather than reconciling all choices together.[18] We did this to encourage subjects to analyze each choice in isola-

[16]Choices that did not violate a rule also could be paid again as reconciliation choices to maintain equal probability of all rules and lotteries being paid. In this case, we paid the subject based on their original rule or lottery choice.

[17]We did not have subjects reconcile $c-$TRANS with the price list, as we could not explain how to make the price list completely intransitive. We did not have subjects reconcile the meaningless distractor rules given that there is no natural way to present the violating choices.

[18]This also means that the reconciliation was not dynamic. That is, a subject who selected both IIA and $c-$IIA in Block 1, and then violated IIA in Block 2, may have reconciled these choices to be consistent with IIA. In doing so, this might lead them to be inconsistent with $c-$IIA! We did not present them this subsequent reconciliation. The reconciliation opportunities were fixed at the beginning of Block 1, as determined by their choices in Blocks 1 and 2.

tion, and to reduce cognitive demand in the reconciliation stage. The reconciliation opportunities for the axioms and control axioms were randomized together ex-ante to minimize any systemic order effects.

Subjects also had the opportunity to reconcile inconsistencies when they chose both the axiom and control axiom. For example, a subject who chose both IIA and $c-$IIA in Block 1 would also see these rules side by side on their screen and choose which, if any, to keep selected. The subjects were not presented with their lotteries during this reconciliation opportunity. Again, the language in these decisions was neutral and simply said that these two rules make opposite choices. These decisions were incentivized in the same way as other revised rule choices.

The number of reconciliation opportunities varied per subject, based on number of violations and on number of axioms and control axioms selected. On average, subjects had six reconciliation opportunities. The number of the reconciliations ranged from 0 to 22.

Our main results analyze data from 110 subjects, primarily undergraduate students at the Ohio State University where the sessions took place. We programmed the experiment using z-Tree (Fischbacher, 2007), and recruited subjects using ORSEE (Greiner, 2015). Sessions lasted about one hour, and subjects earned about $14 on average, including a $7 show-up payment. Subjects were paid after the last subject finished the experiment, and subjects were not able to leave early if they finished quickly. Instructions are included in a Supplemental Appendix.

IV. Main Results

Figure IV shows the percentage of subjects who selected each axiom in Block 1, broken down by whether a subject selected the axiom only, the axiom and the control axiom, only the control axiom, or neither. In aggregate, FOSD is the most popular axiom, selected by 90% of subjects. For the remaining axioms, 85% of subjects select Consistency, 83% select Transitivity, 83% select IIA, and 83% select Independence, and 82% select Branch Independence. Given that the alternative to selecting an axiom is to make one's own choice, these high axiom selection rates indicate a strong ex-ante normative appeal; a vast majority of individuals would rather have the axiom make a choice for them than choose on their own.

One could worry that these axiom selection rates instead reflect a general aversion

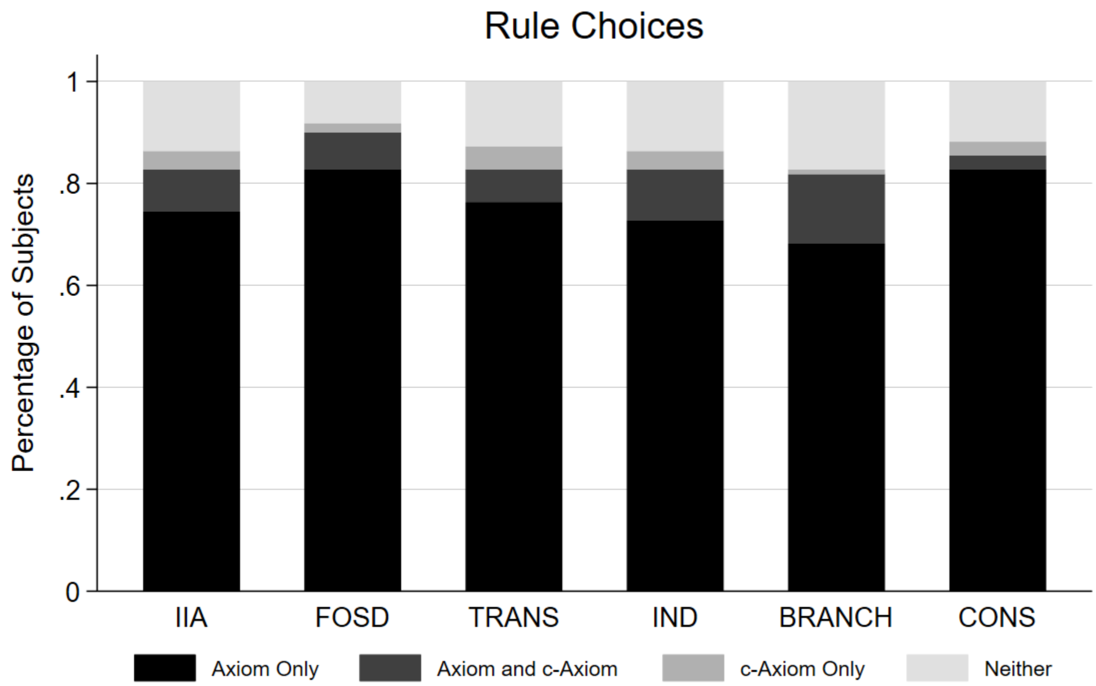


Figure IV: Percentage of Subjects Selecting Each Rule in Block 1

to making decisions, perceived pressure from the experimenter to select the rules, or other external forces masquerading as endorsement of the axioms. Our control axioms confirm that this is not the case, since they are selected by only 11% of subjects. In particular, 15% selected $c-$Branch Independence, 14% selected $c-$Independence, 12% selected $c-$IIA, 11% selected $c-$Transitivity, 9% selected $c-$FOSD, and 5% selected $c-$Consistency.

Our aggregate results are reflected in the individual-level rule selection rates. We find that 60% of subjects selected all six axioms and 65% of subjects never selected a control axiom. Among individuals who ever select a control axiom, it is most common for individuals to select only one (23% of subjects). Therefore, we have confidence that subjects understand the decision rules and incentivization procedure, and generally select only the rules they see as desirable. We report the full distribution of number of axioms and control axioms selected on an individual level in Appendix B Table IV.

Interestingly, FOSD is the most popular axiom while $c-$FOSD is among the least popular control axioms. Similarly, Branch Independence is the least popular axiom while $c-$Branch Independence is the most popular control axiom. This might indicate that there are some patterns to how subjects perceive the axioms. FOSD is most "obviously" desirable, and therefore $c-$FOSD is obviously not desirable. The opposite is true for Branch Independence. This suggests some features of axioms might be more compelling to individuals, or alternatively certain aspects of a rule might be particularly complex. It would be interesting for future work to identify these.[19] It is also interesting that both FOSD and Branch Independence involve "mixing," so it is not that case that individuals are simply averse to, or confused by, mixing.

Overall, we conclude that individuals view these axioms as desirable rules, since they overwhelmingly preferred the axiom to choose on their behalf rather than make the choice on their own.

Result 1. *Nearly all individuals reveal a preference for their choices to satisfy canonical choice axioms. These axioms are selected at higher rates ($\approx 85\%$) than their "opposites" ($\approx 10\%$).*

Given that subjects prefer to satisfy these axioms, a natural question is whether these individuals *do* satisfy the axioms in their choices. Among those who select the respective axiom, 85% of subjects violated FOSD, 75% violated Independence, 46% violated Consistency, 43% violated Transitivity, 38% violated IIA, and 24% violated Branch Independence.[20] In aggregate, over 85% of subjects who violate an axiom selected the axiom to make choices on their behalf in Block 1. This means that "wanting" to follow a rule does not ensure that a subject can or will follow the rule.

Indeed, individuals who select the axiom are no less likely to violate it than those who do not select the axiom. Aggregating across all questions, those who selected an axiom violated it 30% of the time, and those who did not select an axiom violated it 24% of the time (Fisher exact p=0.131).

Result 2. *Preferring an axiom to make choices does not predict adherence to the axiom. Individuals who reveal a preference for their choices to satisfy a canonical choice*

[19]This is in a similar vein to Oprea (2020), who analyzes features of rules that make them complex to implement. Additionally, Kendall and Oprea (2021) find that complexity is highly correlated with individuals' ability to formulate mental models from data, which could be related to understanding of decision rules.

[20]One should not interpret these violation rates as reflecting general comparative likelihood of violating the axioms. We did not have the same number of questions for each axiom (as outlined in the Appendix) and the likelihood of violating each axiom varied across axioms.

axiom are just as likely to violate the axiom as those who preferred to choose on their own.

Given that we observe inconsistencies between an individual's ex-ante preferences over axioms and their own lottery decisions, we analyze whether and how individuals reconcile this discrepancy. Table I reports the main results. In column two, we report the percentage of instances in which subjects maintained inconsistent choices (37% in aggregate). In the remaining columns, we report the direction in which individuals change their inconsistencies. In column three, we report the percentage of instances in which subjects unselected the axiom and kept their lottery choices as they had been (13% in aggregate). In column four, we report the percentage of instances in which subjects kept the axiom selected and changed their lottery choices to be consistent with it (47% in aggregate). In the last column, we report the minority of instances in which subjects both unselected the axiom and changed their lottery choices, or kept the axiom selected but changed their lottery choices in such a way that they were still inconsistent with the axiom (3% in aggregate). Note, the sample sizes vary widely across axioms as individuals violated some axioms more than others, and some axioms had more related questions than others.[21]

Aggregating across our main axioms, we see that just over one-third of violations are left inconsistent, which we discuss below. However, of those who do change their choices, it is far more common for individuals to change their lottery choices to be consistent with the axiom than to unselect the axiom. In 47% of violations, individuals change their lottery choices to be consistent with the axiom. In contrast, in only 13% of instances do they unselect the axiom. Our interpretation is that these 60% of violations reveal mistakes: 79% (47/60) of these are mistaken lottery choices, while only 22% are mistaken axiom choices.

We observe heterogeneity in the tendency to revise inconsistencies. It is interesting to note that FOSD, IND, and BRANCH are the least likely to be revised, and these are the three axioms which involve "mixing."[22] From our data, we cannot say whether this reflects subjects' preferences related to these axioms, or whether it is simply harder for subjects to understand why their choices violate axioms that in-

[21]For example, there were four FOSD questions, and 85% of subjects violated FOSD at least once. On the other hand, there was only one Branch Independence question and 24% of subjects violated the axiom.

[22]We thank Yoram Halevy for this observation.

Axiom	Keep Inconsistent	Unselect Axiom	Change Lotteries	Change and Still Inconsistent
Total (n=468)	37%	13%	47%	3%
IIA (n=63)	19%	2%	78%	2%
FOSD (n=194)	49%	21%	29%	1%
TRANS (n=41)	17%	5%	66%	12%
IND (n=96)	47%	16%	34%	3%
BRANCH (n=22)	41%	0%	55%	5%
CONS (n=52)	13%	0%	79%	8%
Total (n=124)	33%	35%	20%	11%
c-IIA (n=42)	38%	43%	14%	5%
c-FOSD (n=16)	38%	19%	44%	0%
c-TRANS (n=22)	23%	50%	0%	27%
c-IND (n=29)	38%	28%	24%	10%
c-BRANCH (n=8)	38%	38%	25%	0%
c-CONS (n=7)	0%	14%	43%	43%

Table I: Percentage of Violations Revised and Direction of Reconciliation
Notes: The second column gives the percentage of violations that were left inconsistent. The third column reports the percentage instances where subjects revised their axiom selection, the next column reports the percentage instances where subjects revised their lottery choices to be consistent with the axiom, and the final column reports instances where subjects did both or changed their lottery choices in such a way that they were still inconsistent with the axiom. The sample reported is all subjects who both selected and violated a given rule.

volve mixtures. This is especially interesting since FOSD is the most frequently selected axiom in Block 1. This shows that even though individuals may want to follow an axiom, this may not translate to them making choices consistent with it even when given an explanation of how the axiom applies to a decision problem. We leave further investigation of reconciliation properties for specific axioms to future research.

The bottom rows in Table I present the same breakdown of revised choices conditional on subjects selecting the control axioms. On aggregate, when subjects revised their choices to be internally consistent (55% of all inconsistencies), they changed their lottery choices to be consistent with the $c-$axiom only 36% (20/55) of the time, while in the remaining 64% of revisions, subjects renounced the control axiom and kept their lottery choices as they were. This is significantly lower than the 79% of instances where subjects reconcile in favor of the main axioms (Wilcoxon ranksum $p < 0.0001$). The 36% of revisions that change lottery choices to be consistent with the the control axioms could capture any latent tendency to follow rules, and our online data give more insight into these decision makers.

One might still worry that individuals who select the control axioms are systematically different from those who select the axioms. We can look at individuals who selected both the axiom and the control axiom to control for the potential confound that those who do not choose control axioms might be more likely to revise in favor of the lottery. We conduct the same analysis as above restricted to the sub-sample of subjects who choose both an axiom and its corresponding control axiom. We find the results unchanged. When reconciling violations of the axioms, these individuals revise their choices to be internally consistent in two-thirds of violations; in 40% of violations they change their lottery choices, while they unselect the rule in 23% of violations. In contrast, when reconciling violations of the control axioms, they revise their choices to be internally consistent 62% of the time; they change their lottery choices 19% of the time and unselect the rule 43% of the time. These results mimic the aggregate results on the full sample.

Furthermore, Figure V shows the rule reconciliation pattern for these individuals. That is, we look to see how individuals reconcile their rule choices when they selected both the axiom and corresponding control axiom. They had the opportunity to unselect the axiom, unselect the control axiom, both, or neither. Within that sample, we find individuals still favor the main axioms. Among individuals who unselect only one of the rules (70% of individuals), over 89% of them unselect the control axiom. That is, when individuals are faced with two decision rules that prescribe opposite choices, they realize this and abandon the less-sensible rule.

We conclude that about half of individuals who wanted to follow an axiom but violated it made a mistake in their lottery choices. Some violations are kept inconsistent, as we will discuss below. However, among reconciliations, the axioms are usually followed.

Result 3. *Individuals violating canonical axioms often change their choices to be consistent with the axiom (≈ 79% of revisions). Individuals violating control axioms are less likely to do so (≈ 36% of revisions).*

There is a sizable minority of violations that are not reconciled. Table I and Figure V show that about one third of subjects keep their choices inconsistent across these revision opportunities. Inconsistencies with the axioms are revised 63% of the

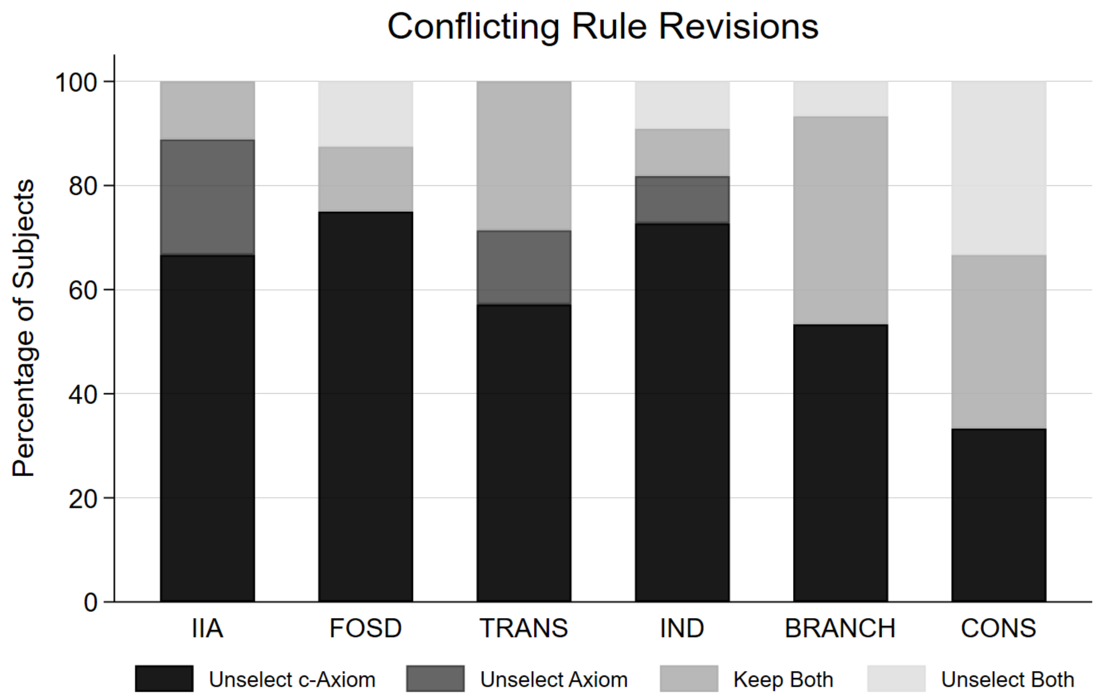


Figure V: Percentage of Subjects Revising Choices in Block 4, Conditional on Selecting Axiom and Control Axiom

Notes: Here, "unselect c-axiom" means the individual kept the axiom selected but unselected the control axiom, "unselect axiom" means they kept the control axiom selected but unselected the axiom, "keep both" means they kept both the axiom and control axioms elected, and "unselect both" means they unselected both the axiom and control axiom. The sample reported is all subjects who selected both an axiom and corresponding c-axiom.

time and inconsistencies with the control axioms are revised 67% of the time (Fisher exact, $p = 0.402$).

While this might seem odd at first blush, there are a few reasons why individuals might keep their choices inconsistent. The most obvious to us is simple effort cost. Subjects have already thought about these decisions and chosen what they prefer. Revising choices is costly in terms of time and cognitive effort, and individuals may view the cost as too high. To test this hypothesis, we look at the first and last revision opportunity that subjects faced. Averaged across all subjects, we find that choices are left inconsistent 31% of the time in the first reconciliation opportunity for a given axiom, while they are left inconsistent 40% of the time in the last opportunity (Fisher exact, $p = 0.148$). This is even stronger in our $1 cost treatment, where first revisions are left inconsistent 33% of the time and last revisions are left inconsistent 65% of the time (Fisher exact, $p < 0.001$). Individuals had more revision opportunities in this treatment, on average, since they select axioms more often due to the $1 cost of not selecting the axiom. The fact that choice fatigue seems to increase in this treatment where there are more revision opportunities supports the hypothesis that inconsistencies are due to attention and effort costs.

In addition, it is likely that mistakes sometimes result from cognitive, time, or attention costs. These same mechanisms would result in maintaining inconsistent choices. We cannot directly test this in our data, but it is plausible that the source of initial mistakes could simultaneously introduce additional "mistakes" in the form of maintained inconsistencies in choices. We leave this for future work to investigate.

Result 4. *Individuals keep their choices inconsistent in about one third of all reconciliation opportunities. We find suggestive evidence that choice fatigue contributes to subjects' willingness to maintain inconsistencies in their choices.*

While we could look at detailed comparisons in original and revised lottery choices, for example whether revised choices become more or less risk averse, our experiment is not designed to answer these questions. We chose the lottery questions in order to maximize violations of the axioms, and therefore the questions are in no sense representative of the violations and revisions we might see more generally. However, we believe our methodology could be very useful in answering these types of questions in future research. For a step in this direction, see Benjamin et al. (2019) who study risky investment decisions before and after reconciliation opportunities.

V. An Online Module

Our main results present evidence that individuals prefer their choices to adhere to normative axioms. We also find suggestive evidence that axiom complexity affects individuals' understanding of their mistakes and likelihood of revising their choices. Given this, it is natural to better understand the relationship between rule preferences and other observable information (e.g., measures of cognition, understanding, response times, risk preference, personality traits, et cetera). There are many open questions, but as a first step, we conduct a short exploratory follow-up experiment to examine whether there is any relationship between mistakes and scores on the

Cognitive Reflection Test (CRT) (Frederick, 2005) or individual response times.[23]

The online experiment also serves three additional purposes. First, the experiment online is presented in a streamlined module targeting a single axiom. Here, we present subjects with simplified reconciliation opportunities that might identify mistakes in a more transparent way, which would be beneficial in less attentive samples such as online participants. Second, the simplified setting serves as a proof of concept for how to implement rule elicitation methods as an add-on module. For example, a researcher studying risk preferences may want to add on this reconciliation opportunity following a more thorough set of tasks. Lastly, this experiment allows us to reach a larger sample of subjects to better understand the robustness of our in-person laboratory results. We discuss the design details of the online experiment below.

In Part 1 of the online experiment, subjects made choices over axioms just as described in Section III. To simplify our decision environment, we focus on a single axiom—independence—and its control.[24] In Part 2, subjects made lottery choices. We include the same six relevant lottery questions as we had in the lab, constituting three potential violations of independence.[25] In Part 3, subjects were given an opportunity to reconcile any inconsistencies in their decisions. We made subtle changes to simplify the reconciliation portion of the design, as we describe below. Finally, in Part 4, subjects answered 10 questions designed to measure cognitive reflection and ability. These questions included the original three questions from the cognitive reflection test (Frederick, 2005) as well as seven additional questions similarly designed to measure cognitive reflection (Meyer and Frederick, 2021).[26]

We changed our presentation of Part 3 to adapt to the online subject population

[23]Both the cognitive reflection task and response times are often thought to be associated with intuitive/heuristic processes (low CRT individuals/fast response times) or reflective/rational processes (high CRT individuals/slow response times). However, the correlation between these two measures depends on the type of question asked and how the question is framed (Alós-Ferrer et al., 2016; Stupple et al., 2017). For these reasons, we refrain from attributing any relation with rule preference to heuristics or reflective choices.

[24]We chose to focus on independence as a "stress test" of our methodology online, since independence is arguably the most complex of our axioms.

[25]We included consistency and a distractor axiom to have a larger set of rules so that subjects did not immediately see the relationship between IND and c-IND. We also included four additional lottery questions that were not related by independence so that similarities between the relevant lotteries were not apparent.

[26]These questions were selected in consultation with Shane Frederick via personal correspondence and we thank him for the suggestions.

which tends to be less attentive and demonstrate lower understanding (Gupta et al., 2021). On the reconciliation decision screen, subjects saw two questions: "Do you still want to keep this rule selected? (Yes/No)" and "Do you want to keep the lottery choices that you originally made, or would you like the lottery choices that the rule would make for you? (My original lottery choices/Choices that the rule would make for me)." If a subject selects to have the choices that the rule would make, an additional question appears and asks them to choose among the set of lottery pairs that are consistent with the rule. Subjects are allowed to change their mind after this is revealed. We discuss the motivation for these changes in our discussion of alternative design choices in Section VI.

We recruited 500 participants through the online platform Prolific.[27] Each participant received a $7 completion payment, equivalent to our show up fee in the lab. We randomly selected one out of every ten participants to receive a bonus payment determined by one randomly selected decision in the experiment.

V.A. Results

Just as in our lab data, we find a large majority of subjects selecting IND, with fewer selecting c-IND. Overall, 75% of individuals select IND and 25% select c-IND in Part 1. This demonstrates a clear preference for IND over c-IND (Wilcoxon signed-rank, $p < 0.0001$), though this difference is weaker than in our lab data. This difference between online and in the lab is driven both by fewer subjects selecting IND online than in the lab and more subjects selecting c-IND online than in the lab (IND: 75% vs. 83%, Fisher exact $p = 0.084$; c-IND: 25% vs. 14%, Fisher exact $p = 0.009$). This suggests that rule selection rates can be attenuated by lower attention and understanding.

Across all three questions, individuals violated IND in 42% of instances, significantly higher than the 35% in the lab (Fisher exact, $p = 0.046$). This is also consistent with noisier decisions online. Nonetheless, we again find that individuals who select IND are no less likely to violate it than those who do not select IND (42% vs. 39%, Fisher exact $p = 0.305$).

Table II reports the comparison between reconciliation behavior online and in the lab. Neither the IND nor c-IND distribution differs significantly online from in the

[27] We targeted college educated individuals in the United States for comparisons with our lab data.

lab (Fisher exact tests: IND $p = 0.230$, c-IND $p = 0.596$). However, the minor distribution changes result in distributions for IND and c-IND that do not differ from one another online (Fisher exact, $p = 0.788$). For both IND and c-IND, individuals are marginally more likely to make their choices consistent with the rule than they are to unselect the rule (Wilcoxon signed-rank IND: $p = 0.0550$, c-IND: $p = 0.0502$). This is perhaps not surprising seeing how IND and c-IND looked more similar to one another in the lab than some of our other axioms and noisy choices online result in data closer to uniform.[28]

Axiom	Keep Inconsistent	Unselect Axiom	Change Lotteries	Change and Still Inconsistent
Lab IND (n=96)	47%	16%	34%	3%
Online IND (n=471)	40%	24%	31%	5%
Lab c-IND (n=29)	38%	28%	24%	10%
Online c-IND (n=216)	41%	22%	31%	6%

Table II: Percentage of Violations Revised and Direction of Reconciliation
Notes: The second column gives the percentage of violations that were left inconsistent. The third column reports the percentage instances where subjects revised their axiom selection, the next column reports the percentage instances where subjects revised their lottery choices to be consistent with the axiom, and the final column reports instances where subjects did both or changed their lottery choices in such a way that they were still inconsistent with the axiom. The samples reported are for subjects who both selected and violated IND or c-IND.

We focus the rest of our analysis in this section on understanding the relationship between rule selection/adherence and CRT scores, our measure of cognition. We create an index, ranging from zero to ten, that indicates the number of correct CRT responses from a given subject. Additionally, we create an understanding index, ranging from zero to eight, that indicates the number understanding questions that a given subject answers correctly throughout the online experiment.[29] We find weak evidence that a higher CRT score is positively correlated with selecting IND (Spearman rank correlation: 0.0878, $p = 0.0498$) and negatively correlated with selecting c-IND (Spearman rank correlation: -0.0774, $p = 0.0840$). However, CRT scores are highly correlated with understanding measures (Spearman rank correlation: 0.221,

[28]In particular, the distributions of reconciliation behavior of IND and c-IND were not significantly different from one another in the lab (Fisher exact $p = 0.139$).

[29]Subjects answered three understanding questions about rules in general in Part 1 and answered five understanding questions about revising choices in Part 3.

$p < 0.0001$). After controlling for understanding, we find no significant relationship between CRT score and selecting IND or c-IND. The only significant relationship we find is that those with higher understanding scores are more likely to select IND and those with lower understanding scores are more likely to select c-IND.[30] Thus, we find that selecting suboptimal rules primarily results from lack of understanding and/or attention, but is independent of the CRT score.

	Keep Inconsistent	Change Lotteries	Change and Still Inconsistent
CRT Score	-0.0975	-0.141	-0.242
	(0.0515)	(0.0467)	(0.0965)
Understanding Score	-0.231	0.150	-0.336
	(0.115)	(0.116)	(0.148)
Constant	2.515	1.960	1.622
	(0.823)	(0.816)	(0.982)

Table III: Relationship Between CRT and Understanding Score on IND Reconciliation Decision
Notes: This reports results from a multinomial logistic regression. The omitted category is those who keep their lottery choices and unselect the axiom. We report standard errors in parentheses; standard errors are clustered at the subject-level.

Finally, we look at the relationship between CRT and understanding scores with reconciliation behavior from Part 3. Table III reports results from a multinomial logistic regression to assess the relationship between revision behavior, CRT, and understanding scores, where the omitted category is individuals who keep their original lottery choices and unselect the axiom. As one might expect, we find that both CRT and understanding scores are negatively associated with both keeping choices inconsistent and changing choices in such a way that they are still internally inconsistent. Perhaps more surprisingly, we find that individuals with lower CRT scores are more likely to change their lottery choices to be consistent with the axiom than unselect the axiom. Specifically, among individuals with a below-average CRT score, 36% of subjects change lotteries to be consistent with the axiom while only 17% select the axiom (signed-rank $p = 0.0002$). In contrast, individuals with above-average CRT scores are equally likely to change their lottery choices as they are to unselect the axiom (30% unselect the axiom, 27% change lottery choices; signed-rank $p = 0.519$).

[30] We report regression results in Appendix Table V.

Running the same regression as below on c-IND reconciliation decisions, we find no significant relationships between CRT or understanding with the c-IND reconciliation decisions (Appendix Table VI).

Thus, individuals who have lower CRT scores are those who are more likely to change their choices to align with the independence axiom than unselect the axiom. In contrast, those with high CRT scores are equally likely to change the choices as to unselect the independence axiom. We find no obvious relationships between the CRT and revisions for c-IND.

Next, we analyze subjects' decision times to give additional insight into the decision-making process in revising inconsistent choices. We consider time to first click rather than total decision time since individuals who change their lottery choices need to make an additional decision, which mechanically increases decision times. We find that those who make their lottery choices consistent with IND click significantly faster than those who decide to unselect the axiom (28 vs. 41 seconds, ranksum $p = 0.0027$). They do not click significantly faster than those who keep choices inconsistent (28 vs. 27 seconds, ranksum $p = 0.115$). It is often documented that faster response times reveal stronger preferences (Konovalov and Krajbich, 2019). This could suggest that lower CRT individuals more strongly prefer to adhere to the axiom. On the other hand, those who unselected the axiom may have been near indifferent (Mosteller and Nogee, 1951) or sufficiently confused since they spent much longer with the question.

Another interpretation is that some decision makers might use the axioms to help them make decisions when these decisions are difficult (Gilboa et al., 2012). One participant perfectly expressed this in our post-experiment questionnaire: "It is interesting that the (lotteries) I chose that were inconsistent (with the rule) were the ones that troubled me most to choose, and I ended up switching them all back to what the rule would pick for me."

However, we want to be clear that this is correlational evidence that is not straightforward to interpret. An additional interesting avenue of research might investigate whether overconfidence plays a role in these reconciliation and how this reacts with cognitive reflection and decision times. For example, those who change their lottery choices quickly might be the least confident about their lottery choices.

Result 5. *Individuals who score lower on the cognitive reflection test are more likely to make their choices consistent with IND than unselect the rule. Individuals who* *make their choices consistent with IND do so more quickly than those who unselect the rule.*

VI. Discussion of Alternative Design Choices

We carefully designed our experiment to allow for a clear interpretation of mistakes with minimal complexity for subjects. We discuss how our design relates to other designs in the literature. In addition, we discuss alternative design choices and the trade offs involved. We believe this discussion will be particularly useful for researchers who wish to transport our framework to other choice domains.

VI.A. Eliciting Rule Preference

We chose to elicit subjects' preferences over rules directly in order to identify mistakes. There are other approaches to identifying mistakes in the literature. These other experiments either explain to subjects that their choices violate a given rule without eliciting subjects' preferences over the rule, or they give the opportunity to revise conflicting choices without explaining the underlying rule. The choice to elicit preferences over axioms directly is a key difference between our approach and other approaches in the literature, so we discuss the trade offs in detail in the context of these related papers.

On one extreme, it is possible to elicit revised decisions without mentioning the underlying axiom at all. Papers such as Crosetto and Gaudeul (2019)—studying the asymmetric dominance effect—and Breig and Feldman (2019)—studying risky convex budget sets—take this approach. These papers simply present subjects with their previous decisions and allow them to revise these choices. This approach avoids any potential experimenter demand effect related to presenting axioms since the axiom is never made explicit.

One downside to this approach is that, in refraining from making the axiom explicit, it becomes less clear that the "revised" choice is a more informed measure of an individual's preferences. Since the subject does not know the decision rule associated with a given question, the researcher cannot say whether the initial choice or revised choice is the one more favored by the individual; we can only see that the choices are potentially different. In contrast, because we elicit the individual's preference for decision rules and present this rule alongside their decisions, we can more reliably interpret the later choices as the subject's preferred choices since the axioms give individuals strictly more information to form their own preferences.

One step around this, as exemplified in Benjamin et al. (2019), is to make the inconsistency in choices explicit without directly mentioning an axiom. In a survey on retirement savings decision, Benjamin et al. (2019) have subjects make decisions under different frames, where the decisions converge under various axioms of interest. This allows them to present and explain inconsistencies across frames, which gives subjects more information than simply asking them to reconsider their decisions. However, the subject never sees the axiom explicitly presented or explained.

Our approach is on the opposite extreme. We directly present and elicit preferences over axioms, and show subjects these axioms to explain inconsistencies in choices. This approach is similar to the studies of MacCrimmon (1968), Moskowitz (1974), and Slovic and Tversky (1974) who first have subjects make decisions, and then present them with arguments related to the axioms that their decisions violate. One key difference between our paper and these studies is that the arguments in the studies above are never for an axiom in *general*, but only argue whether the axiom should apply in *specific decisions*.

For example, MacCrimmon (1968) asked subjects to make decisions designed to induce violations of normative principles, and then discussed these violations verbally with participants and allowed them to change their choices. Moskowitz (1974) studied Allais-type violations and the effect of presenting discussion of adherence to and deviation from independence in responses. Slovic and Tversky (1974) asked lottery questions related to the Allais and Ellsberg paradoxes, then presented subjects with "advice" in the form of explained arguments for and against the earlier previous decisions. The "advice" given to subjects relates to the independence axiom and the sure thing principle. In all of these studies, the discussions of the axioms were in the context of a particular decision problem, rather than presenting the axioms as general principles.

In contrast to the above studies, we elicit subjects' preferences over axioms outside of any individual decision problem. This is most similar to MacCrimmon and Larsson (1979), who ask subjects to rank their agreement with various rules on a scale from zero to ten.[31] The rules were presented as written sentences and were not accompa-

[31] Slovic and Tversky (1974) also have a second experiment where subjects express how much they agree with the "advice." However, the advice does not explain the axioms in full generality.

nied by any specific decision problem.[32] We believe eliciting a subjects' preferences over axioms in general has additional benefits compared to learning about the axiom in the context of a single specific decision problem, which we describe below, though this approach is not without drawbacks.

First, eliciting which axiom an individual prefers allows us to know what rules (ex-ante) a subject wants to follow. This is separate from knowing their preference for decision rules "ex-post" after some intervention. In our paper, the "intervention" was to show subjects their own decisions that violated the rule. One could imagine other interventions using this approach, as well. For example, one could teach individuals about the implications of a rule by showing groups of choices that are consistent or inconsistent with the rule, and let the individuals choose to follow the rule's prescriptions or not.

Second, if a decision rule is only explained by the experimenter as it relates to a subject's choice, then this never gives the subject a chance to voice approval or disapproval of the decision rule in abstract. We felt this would be more likely to lead to subjects changing their choices out of "embarrassment" since the subject is explained the rule by an authority on decision making (i.e., the experimenter).[33] In contrast, eliciting a preference for the decision rule from the subject allows the subject to effectively "give themselves advice" when we later present them with their lottery choices related to the rule.

Finally, eliciting preferences over decision rules gives us a richer data set on subjects' preferences. For example, by selecting a rule, a subject reveals that they prefer *all* of their choices to be consistent with the decision rule on the relevant domain. Without eliciting the axiom preference, we cannot make a claim about an "overall" preference for following the axiom. For any of the alternative schemes above, even when a subject reconciles inconsistent lottery choices to be consistent with a rule, we could only interpret this as wanting to follow the rule *for those particular questions*. In contrast, our design allows us to elicit global information about preferences for a given domain, and it allows us to benchmark the appeal of an axiom against making choices for oneself.

While the above are advantages, this elicitation method also has drawbacks. For

[32]MacCrimmon and Larsson (1979) did not ask subjects to compare their rule choices and lottery choices.

[33]However, Slovic and Tversky (1974) find few revisions after explicitly explaining violations to subjects. This suggests that demand effects might not be a major factor in these types of decisions.

example, one drawback to this approach is that we cannot reliably disentangle subjects' failure to endorse a rule from their failure to understand a rule. We chose a pictorial representation to assist in understanding, but it would be interesting to test different presentation methods and how these interact with rule selection and adherence. Additionally, we chose to use a neutral framing of the rules and present them as global statements, rather than giving explicit detail on how a rule relates to specific decision problems. This also leaves open the possibility that subjects' endorsement, or lack thereof, stems from a failure of understanding.

We believe these approaches all have their own benefits and drawbacks. It is an interesting area for future work to investigate how these design choices influence subjects' understanding and endorsement of rules. Nonetheless, the evidence across all these experiments shows that subjects often change their decisions to become consistent with normative axioms, regardless of the design choice. This gives us reassurance that no single design choice is responsible for the main conclusions we draw.

VI.B. Reconciliation Opportunities

We carefully designed our reconciliation opportunity to be neutral for subjects. In particular, there was no default direction to reconciliation; we presented a subject simultaneously with both their rule choice and lottery choice to reduce experimenter demand effects. A subject could unselect a rule, change their lottery choices, a combination of these changes, or they could doing nothing.

We did not include "placebo" reconciliation opportunities. A "placebo" reconciliation opportunity would allow a subject to change choices that were already consistent with a rule. This design choice was made since we expected individuals would experience choice fatigue from facing a large number of reconciliation opportunities. Indeed, we find evidence that individuals revise their choices less in later reconciliation opportunities. Thus, including placebo reconciliation opportunities would have only increased the cognitive load on subjects and reduced our ability to detect mistakes.

Furthermore, we did not allow individuals to select a rule that had not been chosen originally, even when they satisfied the rule in their lottery decisions. Since we interpret selecting an axiom as a global preference for satisfying it, seeing a single set of choices that are consistent with the axiom should not affect an individual's global preference for satisfying the axiom elsewhere. This remains to be tested empirically.

Finally, we had subjects reconcile each question that violates a rule independently, rather than doing "batch" reconciliations for a given rule. This allows for subjects to make exceptions to the rule based on "what is more rational to do in this instance," as discussed by Gilboa et al. (2009). Moreover, we felt that allowing batch reconciliations while maintaining neutrality of the reconciliation opportunities would place more cognitive demands on the subjects. An interesting open question is whether batch reconciliations change how subjects evaluate the rule. It is also interesting to study whether reconciliation decisions would change or converge over multiple rounds, as in Benjamin et al. (2019).

VI.C. Framing of the Reconciliation Opportunities

As mentioned in Section V, the reconciliation decisions were more transparent in our online experiment. Subjects made active choices of whether to keep following the rule or not. They also had to decide whether they wanted to keep their original lottery choices or have the choices that the rule would make.

We made these changes for three reason. In our lab experiment, a subject's previous decisions were the default, and they could keep these decisions with no additional effort. Given lower attention and a stronger incentive to make fast decisions online, we changed the online version to require active choice. Second, we made it more transparent for subjects to understand how to make their choices consistent with the rule. Given that some mistakes could come from inattention or unwillingness to exert cognitive effort, we designed the online version so that subjects who wanted to follow a given rule could do so with lower cognitive cost. Finally, we eliminated the possibility for subjects to change their lottery choices in a manner inconsistent with the rule. We saw very little of this in the lab, so eliminating the possibility was rather innocuous and allowed us to simplify the decision problem.

We find no significant differences in the distribution of reconciliation choices between the lab and online. This suggests that future work can use this simpler decision framing without worrying about systematically affecting decisions.

VI.D. Control Axioms

Recall, the control axioms are an “opposite" of the axioms of interest. The control axioms are intentionally normatively unappealing. Our purpose in including them is to isolate the role of any mechanical effects from our design that could cloud our interpretation of the results, including experimenter demand effects, using the rules to reduce choice effort, confusion, etc. These control for experimenter demand effects since any argument that the axioms are chosen because they come from an authority also applies to the control axioms. Furthermore, a blanket preference for rule-following would manifest in selection of the axioms as well as the c-axioms.

While there are many possible rules that are unattractive, we chose the control axioms to be the opposite of our main axioms for three reasons. First, this provides a standardized form for the benchmark across all six axioms. The control for each axiom is its opposite, rather than choosing different types of controls for different axioms. Second, in doing so, the control axioms control for any “axiom-specific” confusion or other bias. For example, if one believes that our visual representation of IND is driving subjects’ preferences for following it, then this would also be true for the control axiom, c-IND, since it is displayed in a similar form. Finally, the control axioms always can be applied to the same questions as our main axioms. This allows us to compare violations of rules using the same lottery questions.

That said, it might be desirable to have control axioms that are entirely neutral rather than our *c*-axioms that are intentionally unappealing. Entirely neutral axioms would be difficult to construct in general and would be impossible given the constraints above. However, it might be feasible—and therefore desirable—to design neutral benchmarks in some cases, and we believe future work can explore this in other contexts.

Since the c-axioms are intentionally unappealing in our context, it is hard to interpret the level of axiom selection except to say that it is large in the context of the outside option of making one’s own decision. To provide more context on axiom selection rates, one could include rules that are similar to one another but involve key trade offs in their normative appeal. For example, one could include relaxations of the axioms, heuristics which are “mostly” true, etc. Expanding analysis to these likely would introduce complications and require developing additional machinery to represent domain restrictions and relevant subsets of lotteries. We believe this agenda opens interesting questions for future research to investigate.

VI.E. Incentivization Scheme

We believe it is important to incentivize decisions, but the method used to incentivize preference for decision rules is non-trivial. We chose to elicit an individual's preferences to follow the decision rule relative to making a decision on their own. We felt this was an intuitive benchmark for subjects, and also functions as a relatively high bar against which to interpret axiom selection. Furthermore, this avoids interpretation issues that would arise with different incentive schemes, such as implementing a random choice or the opposite choice when subjects do not want to follow a rule.

We do not elicit willingness to pay to follow a decision rule. In our main treatment, there is no cost for the rule to make a choice on the subject's behalf. In our robustness treatment, there is a $1 cost associated with *not* using the rule. Eliciting whether individuals are willing to pay to have a rule make decisions for them could reveal whether they think they cannot implement the rule themselves. We think this is an interesting open question which is in the spirit of Oprea (2020), who identifies positive willingness to pay to avoid implementing complex rules.

VI.F. Choice and Number of Lotteries

We chose to use a few questions per axiom, based on classic violations in the literature. For details on how questions were chosen, see Appendix C. It would be interesting to do more exhaustive analysis on each axiom to get an overview on where mistakes occur most often, but we leave this for future work. Furthermore, we believe it would be interesting for future work to compare across axioms, which we do not do explicitly in this paper. This presents unique challenges, since violations of some axioms might be "bigger" in utility terms than violations of other axioms, so it is not trivial how to make these comparisons. We believe this to be a fruitful avenue of study.

VII. Discussion

We present incentivized experimental evidence supporting the view that canonical choice axioms have normative content and that violations of axioms can represent mistakes. In directly eliciting preferences over axioms, we find that individuals view them as rules that they want their choices to follow. When lottery choices conflict with stated axiom preferences, individuals often change their choices to be consistent with the axiom, rather than inferring from their choices that the axiom is not desirable.

Our experiment takes a step towards identifying individuals' choices as "preferences" versus "mistakes," but also highlights the difficulties in doing so. The evidence suggests that most subjects do view these axioms as desirable and many subjects change choices accordingly, leading us to interpret their inconsistent lottery choices as mistakes. Nevertheless, a substantial minority of individuals do not change their choices despite wanting to follow the axiom. In this case, it is not obvious how to declare either the axiom or lottery decisions as preferences or mistakes in these cases. However, these situations might not be surprising. For example, Gilboa et al. (2009) argue that it is natural to encounter situations where a preference for a decision rule conflicts with preferences over a single decision problem. Sometimes individuals resolve a conflict by adhering to the rule and other times by adhering to their decisions, but neither needs to be abandoned in general. Subjects in our experiment who conflict in their rules and choices demonstrate that these cases occur in practice.

Our experiment also highlights the importance of understanding the normative content of economic models and axioms for making welfare statements. Assessing the welfare implications of a policy intervention requires adopting a normative model. Without understanding individuals' normative preferences, policy interventions might make individuals worse off. As a simple example, consider expected utility when an individual makes decisions that violate the independence axiom. If an individual *wants* their choices to follow the independence axiom, then a researcher or policy-maker might make the individual better off by giving them something they might not have chosen but that is consistent with independence. While this is a simple example, there are many important situations where these questions are relevant. For example: Can a retirement advisor improve an individual's 401k account by enforcing expected utility assumptions? Can a financial advisor improve individual saving behavior by enforcing a constant discount rate? Can a health coach improve welfare by imposing consistency across different menus?

VII.A. Implications for Theory

Our results suggest a role for economic theory to model preferences over axioms alongside modeling the choices individuals make. Our experiment elicits two revealed preferences—one over axioms and one over lotteries. These preference relations do not always align in practice, resulting in violations of the axioms. However, results suggest that, in many of these cases, the preference over axioms supersedes the lottery preference. Our results suggest that individuals do have preferences over axioms directly, so it might prove fruitful to incorporate these preferences into theoretical models. This would provide structure to exploring the interaction between axiom preferences and choices. There is little theoretical work that explicitly models different types of preferences that are related. One notable example is Gilboa et al. (2010) which models the relation between objective and subjective preferences.

Our results also contribute to an interesting discussion on the role of decision theory as outlined in Gilboa (2010), who writes:

> "We are equipped with the phenomenally elegant classical decision theory and faced with the outpour of experimental evidence á la Kahneman and Tversky, showing that each and every axiom fails in carefully designed laboratory experiments. What should we do in face of these violations? One approach is to incorporate them into our descriptive theories, to make the latter more accurate. This is, to a large extent, the road taken by behavioral economics. Another approach is to go out and preach our classical theories, that is, to use them as normative ones... In other words, we can either bring the theory closer to reality (making the theory a better descriptive one) or bring reality closer to the theory (preaching the theory as a normative one). Which should we choose?"

Our results demonstrate a role for the latter and suggest that individuals already view the classical theory as normative in many instances. For many individuals, violating canonical axioms is revealed a mistake by their own choices. We help individuals make better decisions, according to their own preferences, when we assist them in satisfying these axioms. This is not to diminish the role of descriptive theories, but to draw attention to the different roles that descriptive and normative theories may play.

VII.B. Implications for Experiments

Experimenters often present subjects with decisions like the ones we explore in order to estimate preference parameters. Given that we find individuals making mistakes, it's unclear how estimated preferences would change after individuals are given the opportunity to "correct" their choices. For example, are revised choices systematically more/less risk averse than original choices? As mentioned above, we chose specific questions that would result in violations of the axioms, so our experiment was not designed to answer these questions. However, we think this is an interesting direction of work.

In this vein, it might be the case that there are other features of the environment or experimental design that cause axiom preferences and choice preferences to align. For example, experimental interfaces could notify subjects when they violate consistency or independence of irrelevant alternatives.[34] If we want to elicit subjects' "rational" preferences, then this might require more study into the structure and design of experiments.

More generally, our results suggest caution in designing and interpreting experimental tests of axioms. While it is possible to design choice environments to induce violations of nearly any axiom, researchers (ourselves included) should think carefully about the information this reveals about preferences. Our results suggest that in many instances, these questions do not reveal fundamentally "behavioral" preferences, but instead identify the situations in which individuals have difficulty implementing their normative preferences. These situations are valuable to document descriptively, and future work can develop ways to assist individuals in implementing their preferences in these settings.

VII.C. Directions for Future Research

We view our experiment as one in a line of experiments in procedural choice. We see many interesting directions in which to take this agenda in addition to the open questions we have noted in the previous sections, and we outline a few below.

In our experiment, people tend to follow "rules" over following choices. It would be

[34] For example, many subjects who exhibited multiple switches on a price list (TRANS3) changed their decisions to be consistent with transitivity in the reconciliation stage. This suggests that enforcing a single switching point might actually help subjects express their underlying desire for transitivity.

interesting to understand more about when and where this is true. It is also interesting to identify what aspects of the environment (e.g., framing) alter an individual's perception of decision rules. In a related study, Oprea (2020) analyzes aspects of the decision environment that make rules more complex to implement. It would be interesting to understand more about how these measures of complexity interact with the questions we answer in our paper. For example, what features make axioms more complex to understand? Are more complex axioms less appealing? Does the complexity of the environment (here, relatively simple lottery choices) affect the rules one wishes to implement in that environment? More generally, we believe it fruitful to study when and why it is difficult for people to implement the principles they feel should guide their choices.

Though the environment differs, our paper is also related to the literature studying strategies in repeated games. Romero and Rosokha (2018), Cason and Mui (2019), and Dal Bó and Fréchette (2019), among others, allow subjects to design comprehensive strategies in indefinitely repeated prisoner's dilemma games, rather than choosing actions each period. While our subjects do not design their own "axioms" to follow, our paper can be thought of as a similar *procedural* experiment where subjects choose rules to implement decisions for them. This is similar to the distinction between substantive and procedural rationality as outlined in Simon (1976), who calls for economists to "become interested in the procedures—the rational processes—that economic actors use to cope with uncertainty" (Simon, 1976, p.81). Halevy and Mayraz (2020) take a step in that direction, allowing subjects to design procedures to carry out their investment decisions. They find that subjects prefer using procedures to making decisions on their own, which is similar to our subjects' preference for following the axioms.

There are many other environments in which our methodology could prove useful. In strategic games, one could use this methodology to elicit whether individuals view obeying dominant strategies and best-responding to beliefs as normative principles for different games, even if they fail to implement this principle in their actions. Researchers could also elicit attitudes toward fairness or aggregation rules in the domain of social preferences. In the case of impossibility theorems (e.g., Arrow, 1950), these methods could be used to identify which axioms are the most desirable to relax or abandon. Finally, researchers often have hypotheses about competing heuristics that are difficult to test. One could use our methodology to elicit the desirability of these heuristics directly. In short, we could elicit what individuals “want to” or think they “should” do, in addition to or instead of eliciting what they actually do. Naturally these methodologies are complementary, and we believe this to be a fruitful avenue for future research.

References

Agranov, M. and P. Ortoleva (2017): "Stochastic choice and preferences for randomization," *Journal of Political Economy*, 40–68.

Alós-Ferrer, C., M. Garagnani, and S. Hügelschäfer (2016): "Cognitive reflection, decision biases, and response times," *Frontiers in psychology*, 7, 1402.

Arrow, K. (1950): "A Difficulty in the Concept of Social Welfare," *Journal of Political Economy*, 328–346.

Azrieli, Y., C. P. Chambers, and P. J. Healy (2019): "Incentives in Experiments: A Theoretical Analysis," *Journal of Political Economy*, 126, 1472–1503.

Baillon, A., Y. Halevy, and L. Chen (2021): "Randomize at your own Risk: On the Observability of Ambiguity Aversion," .

Benjamin, D. J., M. Fontana, and M. Kimball (2019): "Reconsidering Risk Aversion," *Presentation: Economic Science Association North American Meeting*.

Birnbaum, M. H. and A. Chavez (1997): "Tests of theories of decision making: Violations of branch independence and distribution independence," *Organizational Behavior and human decision Processes*, 71, 161–194.

Birnbaum, M. H. and T. Martin (2003): "Generalization across people, procedures, and predictions: Violations of stochastic dominance and coalescing," *Emerging perspectives on decision research*, 84–107.

Birnbaum, M. H., D. Navarro-Martinez, C. Ungemach, N. Stewart, E. G. Quispe-Torreblanca, et al. (2016): "Risky decision making: Testing for violations of transitivity predicted by an editing mechanism," *Judgment and Decision Making*, 11, 75–91.

Birnbaum, M. H. and U. Schmidt (2008): "An experimental investigation of violations of transitivity in choice under uncertainty," *Journal of Risk and Uncertainty*, 37, 77.

Breig, Z. and P. Feldman (2019): "Risky Mistakes and Revisions," *Presentation: Economic Science Association North American Meeting*.

BROWN, A. L. AND P. J. HEALY (2018): "Separated decisions," *European Economic Review*, 101, 20–34.

CAMERER, C. (1995): "Individual decision making," *Handbook of experimental economics*.

CASON, T. N. AND V.-L. MUI (2019): "Individual Versus Group Choices of Repeated Game Strategies: A Strategy Method Approach," *Games and Economic Behavior*, 114, 128–145.

CROSETTO, P. AND A. GAUDEUL (2019): "Fast then Slow: A Choice Process Explanation of the Attraction Effect," *Working Paper*.

DAL BÓ, P. AND G. FRÉCHETTE (2019): "Strategy Choice in the Infinitely Repeated Prisoners' Dilemma," *American Economic Review*, forthcoming.

FISCHBACHER, U. (2007): "z-Tree: Zurich toolbox for ready-made economic experiments," *Experimental Economics*, 10, 171–178.

FREDERICK, S. (2005): "Cognitive Reflection and Decision Making," *Journal of Economic Perspectives*, 19, 25–42.

GILBOA, I. (2010): "Questions in Decision Theory," *Annual Review of Economics*, 2, 1–19.

GILBOA, I., F. MACCHERONI, M. MARINACCI, AND D. SCHMEIDLER (2010): "Objective and subjective rationality in a multiple prior model," *Econometrica*, 78, 755–770.

GILBOA, I., A. POSTLEWAITE, AND D. SCHMEIDLER (2009): "Is it always rational to satisfy Savage's axioms?" *Economics & Philosophy*, 25, 285–296.

——— (2012): "Rationality of belief or: why savage's axioms are neither necessary nor sufficient for rationality," *Synthese*, 187, 11–31.

GREINER, B. (2015): "Subject pool recruitment procedures: organizing experiments with ORSEE," *Journal of the Economic Science Association*, 1, 114–125.

GUPTA, N., L. RIGOTTI, AND A. WILSON (2021): "The Experimenters' Dilemma: Inferential Preferences over Populations," *Working Paper*.

HALEVY, Y. AND G. MAYRAZ (2020): “Modes of Rationality: Act versus Rule-Based Decisions,” *Working Paper*.

HUBER, J., J. W. PAYNE, AND C. PUTO (1982): “Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis,” *Journal of consumer research*, 9, 90–98.

JAIN, R. AND K. NIELSEN (2019): “A systematic test of the independence axiom near certainty,” *Working Paper*.

KAHNEMAN, D. AND A. TVERSKY (1979): “Prospect Theory: An Analysis of Decision under Risk,” *Econometrica*, 47, 263–291.

KENDALL, C. AND R. OPREA (2021): “On the Complexity of Forming Mental Models,” *Working Paper*.

KONOVALOV, A. AND I. KRAJBICH (2019): “Revealed strength of preference: Inference from response times,” *Judgment and Decision Making*, 14, 263–291.

LOOMES, G., C. STARMER, AND R. SUGDEN (1991): “Observing violations of transitivity by experimental methods,” *Econometrica: Journal of the Econometric Society*, 425–439.

——— (1992): “Are preferences monotonic? Testing some predictions of regret theory,” *Economica*, 17–33.

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

MACCRIMMON, K. R. (1968): “Descriptive and normative implications of the decision-theory postulates,” in *Risk and uncertainty*, Springer, 3–32.

MACCRIMMON, K. R. AND S. LARSSON (1979): “Utility Theory: Axioms versus ‘Paradoxes’,” in *Expected Utility Hypotheses and the Allais Paradox*, 333–409.

MAY, K. O. (1954): “Intransitivity, utility, and the aggregation of preference patterns,” *Econometrica: Journal of the Econometric Society*, 1–13.

MCFADDEN, D. (1973): “Conditional logit analysis of qualitative choice behavior,” .

MEYER, A. AND S. FREDERICK (2021): "Forming and Revising Intuitions," *Working Paper*.

MOSKOWITZ, H. (1974): "Effects of problem representation and feedback on rational behavior in Allais and Morlat-type problems," *Decision Sciences*, 5, 225–242.

MOSTELLER, F. AND P. NOGEE (1951): "An experimental measurement of utility," *Journal of Political Economy*, 59, 371–404.

OPREA, R. (2020): "What Makes a Rule Complex," *American Economic Review*, 110, 3913–3951.

OWENS, D., Z. GROSSMAN, AND R. FACKLER (2014): "The Control Premium: A Preference for Payoff Autonomy," *American Economic Journal: Microeconomics*, 6, 138–161.

REGENWETTER, M., J. DANA, AND C. P. DAVIS-STOBER (2011): "Transitivity of preferences." *Psychological review*, 118, 42.

ROMERO, J. AND Y. ROSOKHA (2018): "Constructing Strategies in the Indefinitely Repeated Prisoner's Dilemma Game," *European Economic Review*, 104, 185–219.

SAVAGE, L. J. (1954): *The foundations of statistics*, Courier Corporation.

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

——— (1990): "Two-Stage Lotteries without the Reduction Axiom," *Econometrica*, 58, 349–377.

SEIDL, C. (2002): "Preference reversal," *Journal of Economic Surveys*, 16, 621–655.

SIMON, H. A. (1976): "From Substantive to Procedural Rationality," *25 Years of Economic Theory*, 65–86.

SLOVIC, P. AND A. TVERSKY (1974): "Who accepts Savage's axiom?" *Behavioral science*, 19, 368–373.

STUPPLE, E. J., M. PITCHFORD, L. J. BALL, T. E. HUNT, AND R. STEEL (2017): "Slower is not always better: Response-time evidence clarifies the limited role of miserly information processing in the Cognitive Reflection Test," *PloS one*, 12, e0186404.

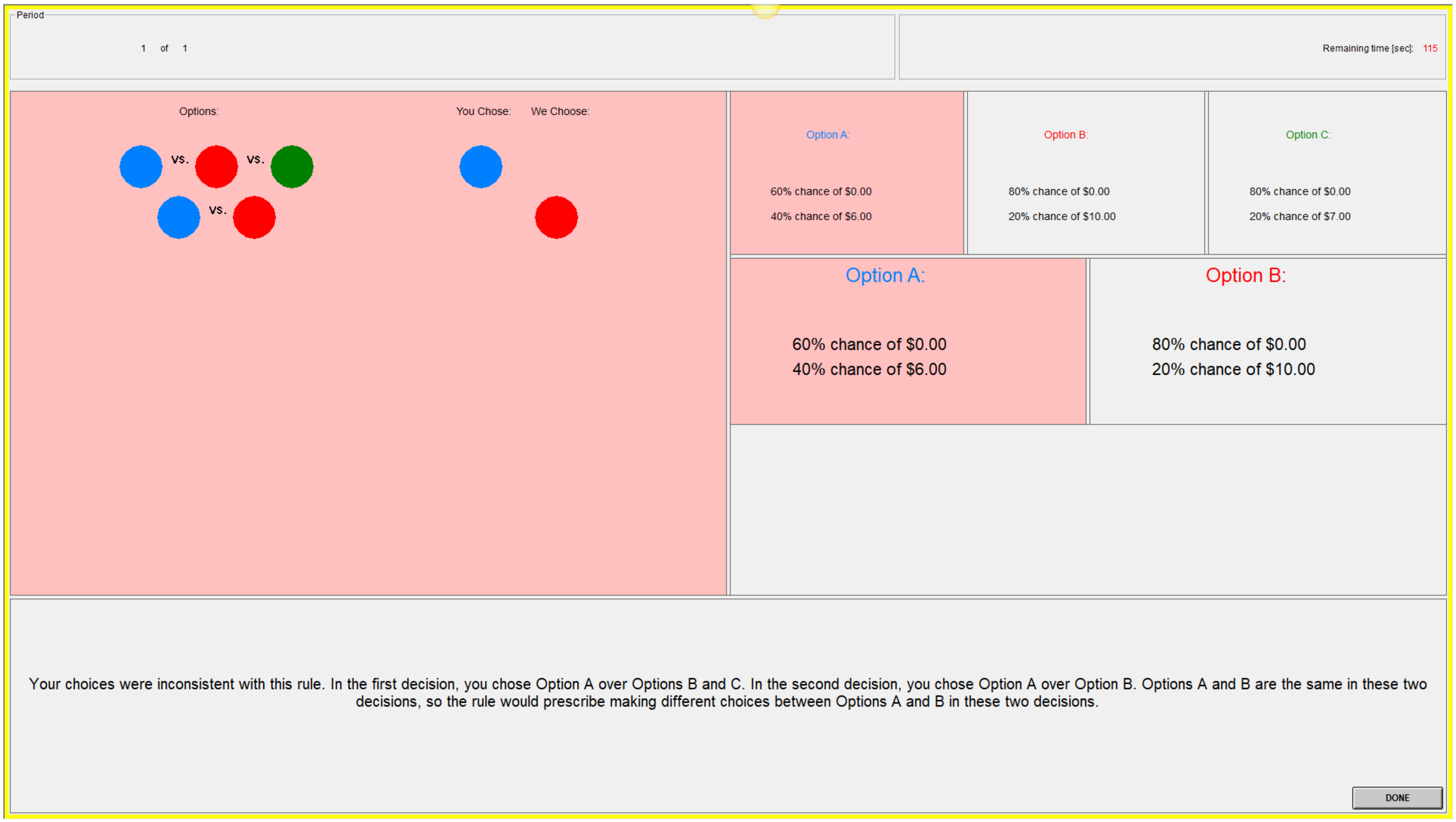
THURSTONE, L. L. (1927): "A law of comparative judgment," *Psychological review*, 34, 273.

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

WEDELL, D. H. (1991): "Distinguishing among models of contextually induced preference reversals." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 767.

A. Screenshots

Period
1 of 1
Remaining time [sec]: 118
Options:
vs.
vs.
vs.
You Chose:
We Choose:
Option A:
60% chance of $0.00
40% chance of $6.00
Option B:
80% chance of $0.00
20% chance of $10.00
Option C:
80% chance of $0.00
20% chance of $7.00
Option A:
60% chance of $0.00
40% chance of $6.00
Option B:
80% chance of $0.00
20% chance of $10.00
Your choices were inconsistent with this rule. In the first decision, you chose Option A over Options B and C. In the second decision, you chose Option B over Option A. Options A and B are the same in these two decisions, so the rule would prescribe making the same choice between Options A and B in these two decisions.
DONE


Period
1 of 1
Remaining time [sec]: 115
Options:
vs.
vs.
vs.
You Chose:
We Choose:
Option A:
60% chance of $0.00
40% chance of $6.00
Option B:
80% chance of $0.00
20% chance of $10.00
Option C:
80% chance of $0.00
20% chance of $7.00
Option A:
60% chance of $0.00
40% chance of $6.00
Option B:
80% chance of $0.00
20% chance of $10.00
Your choices were inconsistent with this rule. In the first decision, you chose Option A over Options B and C. In the second decision, you chose Option A over Option B. Options A and B are the same in these two decisions, so the rule would prescribe making different choices between Options A and B in these two decisions.
DONE

B. Additional Results

	Number of Control Axioms Selected							
#Axioms Selected	0	1	2	3	4	5	6	Total
0	2.7%	0.0%	0.0%	0.0%	0.0%	0.0%	0.9%	3.6%
1	0.9%	0.0%	0.9%	0.0%	0.0%	0.0%	0.0%	1.8%
2	2.7%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	2.7%
3	3.6%	0.0%	0.0%	0.0%	1.8%	0.9%	0.0%	6.4%
4	3.6%	2.7%	1.8%	0.0%	0.0%	0.0%	0.0%	8.2%
5	11.8%	2.7%	1.8%	0.0%	0.9%	0.0%	0.0%	17.3%
6	39.1%	17.3%	0.9%	1.8%	0.0%	0.0%	0.9%	60.0%
Total	64.5%	22.7%	5.5%	1.8%	2.7%	0.9%	1.8%	100.0%

Table IV: Number of Axioms and control axioms Selected

	Select IND	Select c-IND
CRT Score	0.142	0.111
	(0.0890)	(0.0903)
Understanding Score	0.165	-0.162
	(0.0675)	(0.0721)
CRT × Understanding Score	-0.0178	-0.0192
	(0.0137)	(0.0146)
Constant	-0.486	0.377
	(0.410)	(0.428)

Table V: Relationship Between CRT and Understanding Score on IND and c-IND Axiom Selection

Notes: This reports results from a probit regression. Standard errors are clustered at the subject-level.

	Keep Inconsistent	Change Lotteries	Change and Still Inconsistent
CRT Score	-0.00395	0.0498	0.0319
	(0.0717)	(0.0718)	(0.100)
Understanding Score	0.111	-0.0461	0.0280
	(0.113)	(0.109)	(0.173)
Constant	0.0230	0.406	-1.571
	(0.645)	(0.633)	(1.126)

Table VI: Relationship Between CRT and Understanding Score on c-IND Reconciliation Decision

Notes: This reports results from a multinomial logistic regression. The omitted category is those who keep their lottery choices and unselect the axiom. Standard errors are clustered at the subject-level.

C. Lotteries

A description of references and behavioral effects for each question can be found in Table VII, along with the question numbering that we use to present the results. The payment amounts span from \$0 to \$30, and lotteries range in expected value from \$1.40 to \$26. Within each question (with either two or three lotteries), the difference in expected value between the lotteries ranges from \$0 to \$6. On average, the expected value difference is just shy of \$1.75. Furthermore, these lottery questions are not the same as the lottery questions that incentivized Block 1 rule choices. This was to ensure that subjects do not update on the type of lotteries where the rules would apply. We briefly describe the source of the questions. We note that none of the choices directly replicate questions from the source paper. We did this since the payments were not comparable across experiments.

Question	Reference
IIA1–IIA3	Huber et al. (1982) (Attraction Affect)
IIA4	Huber et al. (1982) (Compromise Effect)
FOSD1–FOSD4	Birnbaum and Martin (2003)
TRANS1–TRANS2	Loomes et al. (1991) (Regret Theory)
TRANS3	Brown and Healy (2018) (Multiple Switch Points)
IND1	Birnbaum and Chavez (1997)
IND2	Kahneman and Tversky (1979) (Certainty Effect)
IND3	Jain and Nielsen (2019) (Reverse Certainty Effect)
BRANCH1	Birnbaum and Chavez (1997)
CONS1–CONS2	Brown and Healy (2018) (Near Indifference)

Table VII: Description of Questions

INDEPENDENCE OF IRRELEVANT ALTERNATIVES — We used four IIA questions similar to Huber et al. (1982). Three of the four questions targeted violations of IIA by adding a dominated lottery to a binary choice problem to "attract" the subject to the dominating alternative. We refer to these as IIA1, IIA2, and IIA3. The fourth question, IIA4, targeted a violation of IIA by adding a lottery to a binary choice problem that makes one of the initial two lotteries a "compromise" option.

TRANSITIVITY — We used two transitivity questions (TRANS1 and TRANS2) similar to Loomes et al. (1991) that were used to examine regret theory. In addition, we included six binary questions which together comprised a separated price list, as demonstrated in Brown and Healy (2018).[35] These six questions involved binary comparisons between a risky lottery and a sure payment. The risky lottery was the same in all six questions while the sure payment varied. Multiple switch points on the price list constitute a violation of transitivity, which we refer to as TRANS3.

FIRST-ORDER STOCHASTIC DOMINANCE — We asked four binary questions to target violations of FOSD (FOSD1–FOSD4). All four questions followed the structure in Birnbaum and Martin (2003).

INDEPENDENCE — We included three questions that targeted violations of independence, including one question from Birnbaum and Chavez (1997) (IND1), one question from (Kahneman and Tversky, 1979) demonstrating the certainty effect (IND2), and one question from Jain and Nielsen (2019) demonstrating the reverse certainty effect (IND3).

BRANCH INDEPENDENCE — We included one question targeting a violation of Branch Independence from Birnbaum and Chavez (1997) (BRANCH1).

CONSISTENCY — We included two questions to target violations of consistency in which we asked two binary decision problems that were each repeated twice (CONS1 and CONS2).[36] We chose the binary decision problems based on the questions that were nearest to indifference in Brown and Healy (2018).

Table VIII: IIA Questions

Question	Lottery A	Lottery B	Lottery C
IIA1	60% chance of \$0 40% chance of \$6	80% chance of \$0 20% chance of \$10	80% chance of \$0 20% chance of \$7
IIA2	60% chance of \$0 40% chance of \$6	80% chance of \$0 20% chance of \$10	85% chance of \$0 15% chance of \$10
IIA3	60% chance of \$0 40% chance of \$6	80% chance of \$0 20% chance of \$10	85% chance of \$0 15% chance of \$7
IIA4	60% chance of \$0 40% chance of \$6	80% chance of \$0 20% chance of \$10	70% chance of \$0 30% chance of \$8

[35]For example, one question was a binary choice between lottery p and \$14.00. Another was the choice between p and \$14.50, another between p and \$15.00, etc. Presenting questions of this form is a common procedure used to elicit a certainty equivalent for lottery p.

[36]The questions are not repeated back to back.

Table IX: FOSD Questions (*B* FOSD *A*)

Question	Lottery *A*	Lottery *B*
FOSD1	10% chance of $1.25 5% chance of $9 85% chance of $9.75	5% chance of $1.25 5% chance of $1.50 90% chance of $9.75
FOSD2	10% chance of $2 5% chance of $16 85% chance of $19	5% chance of $2 5% chance of $3 90% chance of $19
FOSD3	21% chance of $1 18% chance of $10.25 61% chance of $11	1% chance of $1 19% chance of $2 80% chance of $11
FOSD4	21% chance of $0.50 18% chance of $13 61% chance of $16	1% chance of $0.50 19% chance of $4 80% chance of $16

Table X: Transitivity Questions

Question	Lottery *A*	Lottery *B*	Lottery *C*
TRANS1	30% chance of $6 30% chance of $6 40% chance of $20	30% chance of $0.50 30% chance of $11 40% chance of $11	30% chance of $8 30% chance of $8 40% chance of $8
TRANS2	45% chance of $7.50 25% chance of $7.50 30% chance of $19	45% chance of $1.25 25% chance of $10.50 30% chance of $10.50	45% chance of $9 25% chance of $9 30% chance of $9

Table XI: Price List Transitivity Questions

Question	*A*	*B*	*C*	*D*	*E*	*F*	*G*
TRANS3	25% chance of $5 75% chance of $25	100% chance of $14	100% chance of $14.50	100% chance of $15	100% chance of $15.50	100% chance of $16	100% chance of $16.50

Table XII: Independence Questions

Question	Lottery A	Lottery B	Lottery C	Lottery D
BRANCH1	80% chance of \$0 10% chance of \$2 10% chance of \$12	80% chance of \$0 10% chance of \$8 10% chance of \$9	10% chance of \$2 10% chance of \$12 80% chance of \$15	10% chance of \$8 10% chance of \$9 80% chance of \$15
IND1	80% chance of \$0 10% chance of \$6 10% chance of \$11	80% chance of \$0 10% chance of \$8 10% chance of \$9	40% chance of \$0 30% chance of \$6 30% chance of \$11	40% chance of \$0 30% chance of \$8 30% chance of \$9
IND2	100% chance of \$10	20% chance of \$0 80% chance of \$13.50	75% chance of \$0 25% chance of \$10	80% chance of \$0 20% chance of \$13.50
IND3	20% chance of \$10 80% chance of \$30	100% chance of \$20	24% chance of \$10 8% chance of \$20 68% chance of \$30	8% chance of \$10 84% chance of \$20 8% chance of \$30

Table XIII: Consistency Questions

Question	Lottery A	Lottery B
CONS1	50% chance of \$3 50% chance of \$15	25% chance of \$5 75% chance of \$12
CONS2	50% chance of \$5 50% chance of \$10	30% chance of \$0 70% chance of \$15

Question	# of Violators	# Keep Inconsistent	# Unselect Axiom	# Change Lotteries	# Change Inconsistently
IIA1	18	3	0	15	0
IIA2	17	2	1	13	1
IIA3	17	4	0	13	0
IIA4	11	3	0	8	0
FOSD1	59	27	11	20	1
FOSD2	55	25	12	18	0
FOSD3	48	26	11	10	1
FOSD4	32	17	7	8	0
TRANS1	6	1	2	3	0
TRANS2	3	1	0	2	0
TRANS3	32	5	0	22	5
IND1	22	7	4	11	0
IND2	37	21	4	11	1
IND3	37	17	7	11	2
BRANCH1	22	9	0	12	1
CONS1	19	1	0	14	4
CONS2	33	6	0	27	0

Table XIV: Number of Violations and Revisions Per Question, Conditional on Selecting Axiom

D. Strict Cost of Deciding

As we described in Section III, we ran an additional treatment where subjects had to pay a strictly positive cost, \$1, to make decisions on their own rather than following a rule. Figure VI shows the difference in rule selection rates across treatments. For each of the axioms, individuals are more likely to select both the axiom and the control axiom in the \$1 cost treatment.

51% of inconsistencies are revised, which is significantly lower than the 63% in our main treatment (Fisher's Exact, $p < 0.001$). However, among choices reconciled to be consistent, it is still the case that individuals reconcile in favor of the axiom. 91% of reconciliations change lottery choices, which is significantly higher than the 79% in the main treatment (Fisher exact, $p < 0.001$). We find a similar story for the control axiom revisions. 39% of inconsistencies are revised, which is significantly lower than the 67% in the main treatment (Fisher exact, $p < 0.001$). Among the revised choices, 46% are revised in favor of the $c-$axiom, which is directionally but not significantly higher than the 36% in the main treatment (Fisher exact, $p = 0.243$).

Overall, we find the same qualitative results in both treatments: Individuals find the axioms to be normatively appealing and revise inconsistencies in favor of the axioms. The treatment differences do give some interesting insights into how subjects perceive these axioms. Individuals are more willing to follow rules when it saves

Question	# of Violators	# Keep Inconsistent	# Unselect Axiom	# Change Lotteries	# Change Inconsistently
$c-$IIA1	10	3	4	2	1
$c-$IIA2	12	4	5	2	1
$c-$IIA3	11	4	6	1	0
$c-$IIA4	9	5	3	1	0
$c-$FOSD1	5	1	0	4	0
$c-$FOSD2	4	2	1	1	0
$c-$FOSD3	2	1	1	0	0
$c-$FOSD4	5	2	1	2	0
$c-$TRANS1	11	3	5	3	0
$c-$TRANS2	11	2	6	0	3
$c-$IND1	11	5	3	1	2
$c-$IND2	8	1	3	4	0
$c-$IND3	10	5	2	2	1
$c-$BRANCH1	8	3	3	2	0
$c-$CONS1	4	0	1	1	2
$c-$CONS2	3	0	0	2	1

Table XV: Number of Violations and Revisions Per Question, Conditional on Selecting control axiom

Note: There were six total instances where a subject changed their lottery choices in such a way that they were still inconsistent with the $c-$axiom, but did not unselect the $c-$axiom. These are included in the last column.

Question	# of Violators	# Keep Inconsistent	# Unselect Axiom	# Change Lotteries	# Change Inconsistently
IIA1	14	3	0	11	0
IIA2	21	7	0	14	0
IIA3	18	6	0	12	0
IIA4	12	7	0	4	1
FOSD1	60	38	5	17	0
FOSD2	55	28	5	21	1
FOSD3	50	39	2	9	0
FOSD4	31	17	2	11	1
TRANS1	2	1	0	1	0
TRANS2	3	1	0	2	0
TRANS3	35	7	0	27	1
IND1	21	7	1	13	0
IND2	42	28	1	10	3
IND3	37	24	3	10	0
BRANCH1	17	6	1	10	0
CONS1	28	0	0	25	3
CONS2	22	10	0	12	0

Table XVI: Number of Violations and Revisions Per Question, Conditional on Selecting and Violating Axiom, in $1 Cost to Decide Treatment

Note: There were three total instances where a subject changed their lottery choices in such a way that they were still inconsistent with the axiom, but did not unselect the axiom. These are included in the last column.

Question	# of Violators	# Keep Inconsistent	# Unselect Axiom	# Change Lotteries	# Change Inconsistently
$c-$IIA1	17	10	4	1	2
$c-$IIA2	16	12	3	1	0
$c-$IIA3	17	11	3	3	0
$c-$IIA4	10	7	3	0	0
$c-$FOSD1	6	2	1	3	0
$c-$FOSD2	8	4	2	2	0
$c-$FOSD3	12	8	2	2	0
$c-$FOSD4	15	8	4	3	0
$c-$TRANS1	23	12	4	0	7
$c-$TRANS2	24	14	5	0	5
$c-$IND1	29	21	1	4	3
$c-$IND2	17	11	1	3	2
$c-$IND3	17	10	3	4	0
$c-$BRANCH1	24	16	3	4	0
$c-$CONS1	14	5	0	3	6
$c-$CONS2	13	7	2	2	2

Table XVII: Number of Violations and Revisions Per Question, Conditional on Selecting and Violating control axiom, in $1 Cost to Decide Treatment

Note: There were fifteen total instances where a subject changed their lottery choices in such a way that they were still inconsistent with the axiom, but did not unselect the axiom. These are included in the last column.

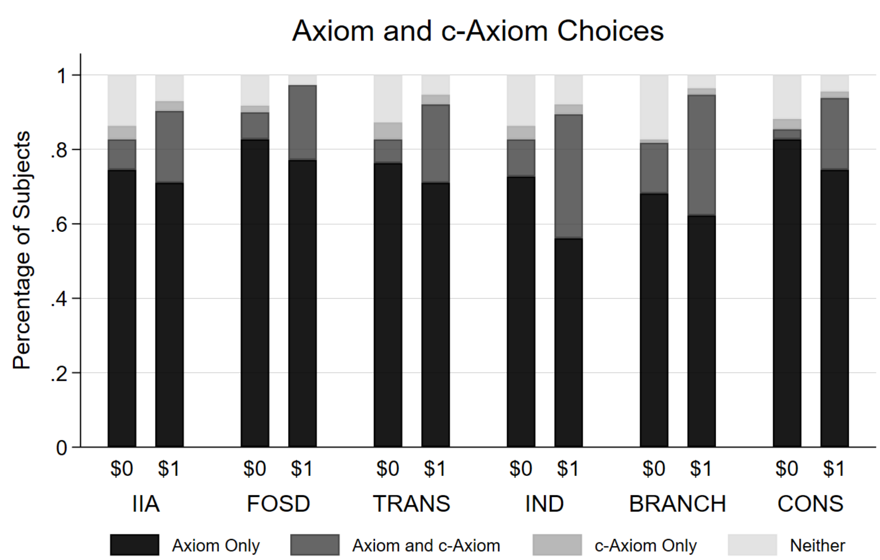


Figure VI: Axiom Selection Rates Across Treatments them \$1, which is intuitive. They are also less likely to revise inconsistent choices. This suggests that individuals in the \$0 cost treatment view selecting the axiom as indicating that it should be "always" true, while individuals in the \$1 cost treatment view it as something that should be "mostly" true.

In Tables XVI and XVII, we report the percentage of violations and direction of reconciliation for each question.

E. Axiom Ranking

To elicit willingness to reconcile inconsistent choices, we had subjects rank any of the six main axioms they selected against a \$1 outside option. For example, a subject who selected all six axioms would see seven boxes on their screen—one for each of the axioms, and one with an option that says "I would rather have \$1 than reconcile choices associated with any of the remaining rules." Subjects first select the axiom they would most want to reconcile should their choices violate it, or select the outside option if they would rather have \$1 than reconcile their choices. Then, subjects select the axiom they would next-most want to reconcile among the remaining axioms, and so on.

If the ranking were chosen for payment, we would randomly select two of the available axioms or the outside option and pay the subject according to the Block 4 reconciled choices from whichever axiom they ranked higher. For example, take a subject who ranks the axioms in the order FOSD ≻ IIA ≻ TRANS ≻ CONS ≻ Outside Option ≻ BRANCH ≻ IND. We would randomly select two options, say IIA and the Outside Option. The subject ranked IIA higher, so they would be paid based on their reconciled choices in the IIA question, as described below.[37] Regardless of their ranking, a subject still faces all of the reconciliation decisions in Block 4; the ranking only impacts which would be paid.

Technically, the reconciled choices for the last-ranked axiom would not be incentivized in this procedure. The reconciled choices of last-ranked axiom would never be implemented since we implement the reconciled choices of whichever axiom is ranked higher. To ensure incentive compatibility, there's an independent chance that

[37] If they had ranked the outside option higher, they would be paid for their original choices in the IIA question and would receive an extra \$1 bonus. If they did not violate IIA, they would be paid for their original choices and would receive a \$1 bonus, which ensures their ranking is not affected by their perceptions of which axioms they were more likely to have violated.

we would randomly select the reconciled choices to pay, as described below. This means that the reconciled choices are almost twice as likely to be paid as the original choices. Therefore, if anything, subjects should be more concerned with their reconciled choices than their original choices. This also helps encourage subjects to carefully consider the reconciliation opportunity.

The main purpose of the ranking is to see whether subjects strictly prefer to re-evaluate their choices. This gives us only a coarse measure—whether individuals are willing to give up at least $1. We felt this would be easier for subjects to understand than trying to elicit a finer willingness-to-pay for each axiom. We also use the reported rankings to look for any consistent patterns within rankings.[38] This gives us a finer measure of subjects' perceptions of the rules compared to the binary ranking in Block 1.[39]

We find that 48% of subjects rank IND before the $1 outside option, and this is similar at 48% for TRANS, 47% for IIA, 45% for CONS, 44% for FOSD, and 43% for BRANCH. On an individual level, 46% of subjects are willing to give up $1 to re-evaluate choices corresponding to at least one axiom.

Looking at the average ranking of each axiom, again we find very few differences. Figure VII presents the average ranking for each of the alternatives. Transitivity is ranked lowest (most preferred), while BRANCH is the least preferred, on average. There are some minor differences across axioms, but overall it seems that subjects do not have any systematic preferences among the axioms they wish the follow.

[38] With this procedure, we cannot rule out that subjects are indifferent among the axioms or to the outside option. We look for systematic patterns in the rankings, but acknowledge this as a shortcoming of our ranking procedure. We also cannot rule out that rankings are driven by subjects' beliefs about the expected value differences of the associated lotteries.

[39] Eliciting a ranking over axioms is also similar to the work of MacCrimmon and Larsson (1979).

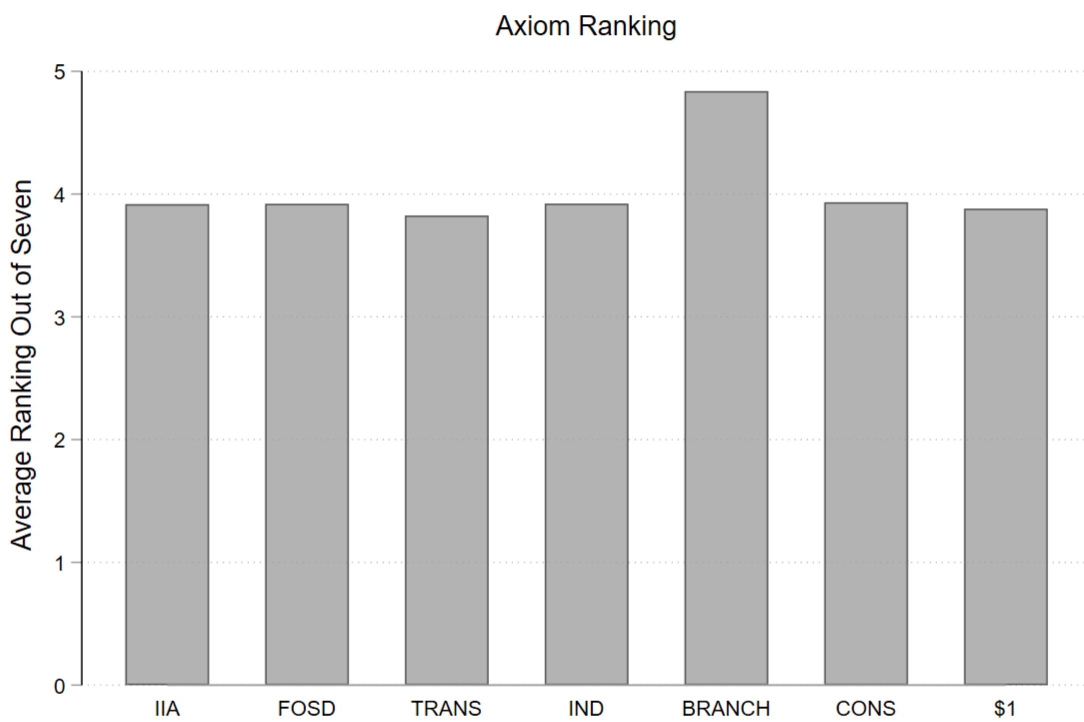


Figure VII: Average Ranking of Each Axiom
Notes: Lower ranking corresponds to the rule being more preferred.

F. SUPPLEMENTAL APPENDIX

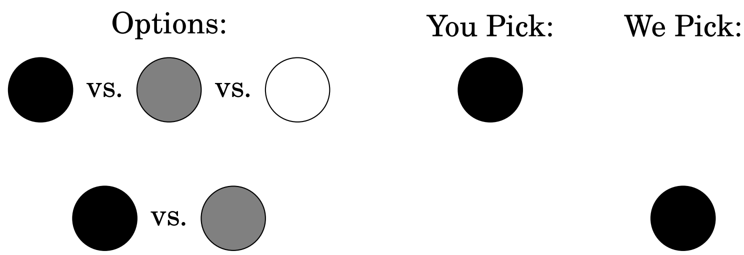


Figure VIII: IIA Rule

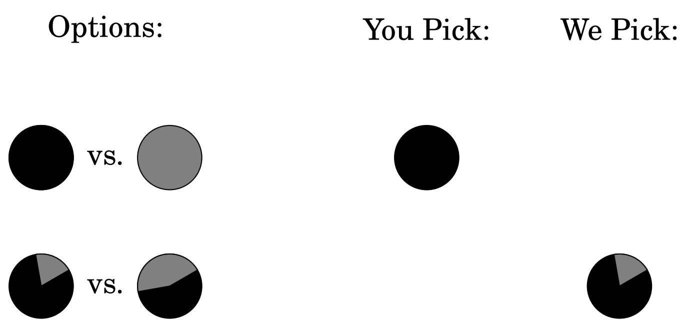


Figure IX: First Order Stochastic Dominance Rule

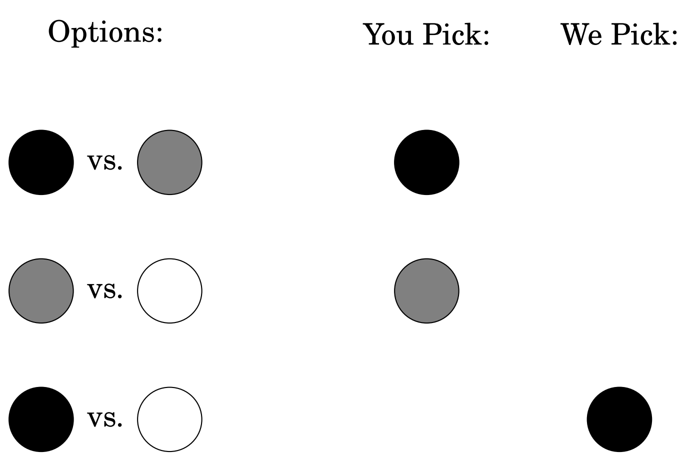


Figure X: Transitivity Rule

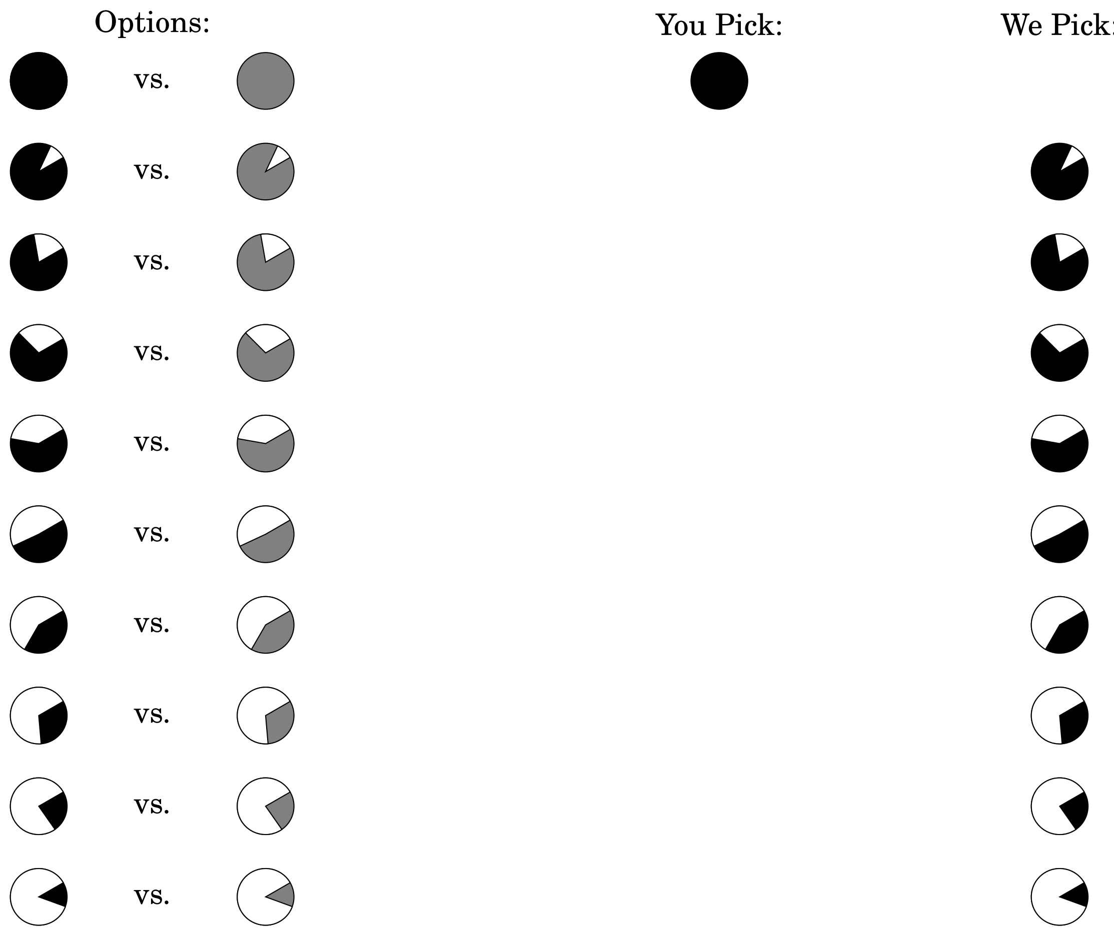


Figure XI: Independence Rule

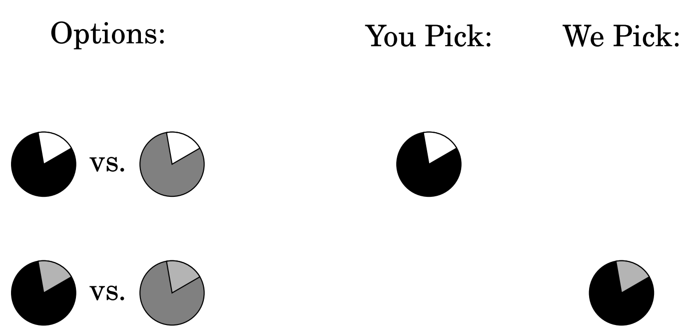


Figure XII: Branch Independence Rule

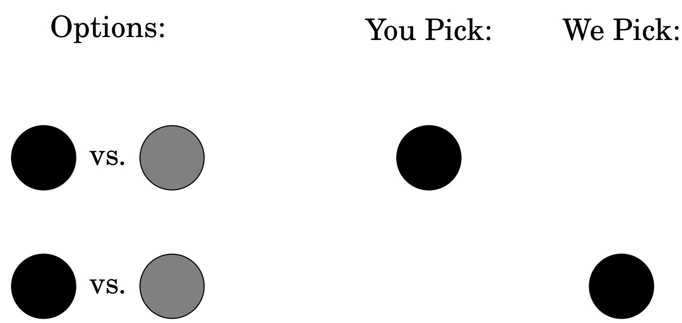


Figure XIII: Consistency Rule

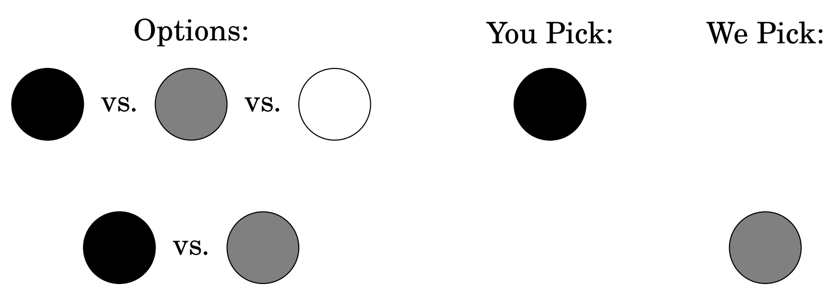


Figure XIV: $c-$IIA Rule

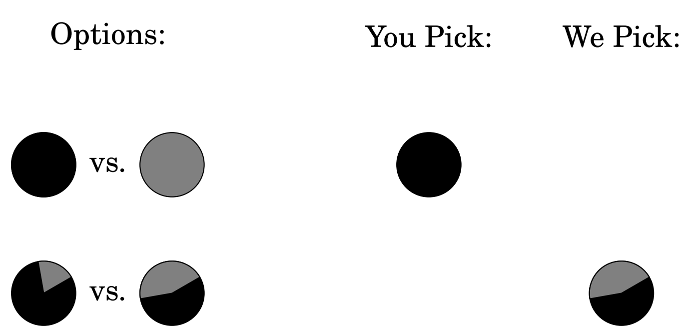


Figure XV: $c-$First Order Stochastic Dominance Rule

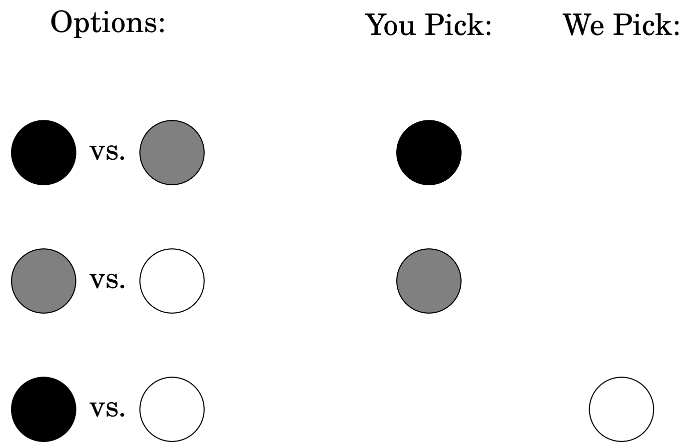


Figure XVI: $c-$Transitivity Rule

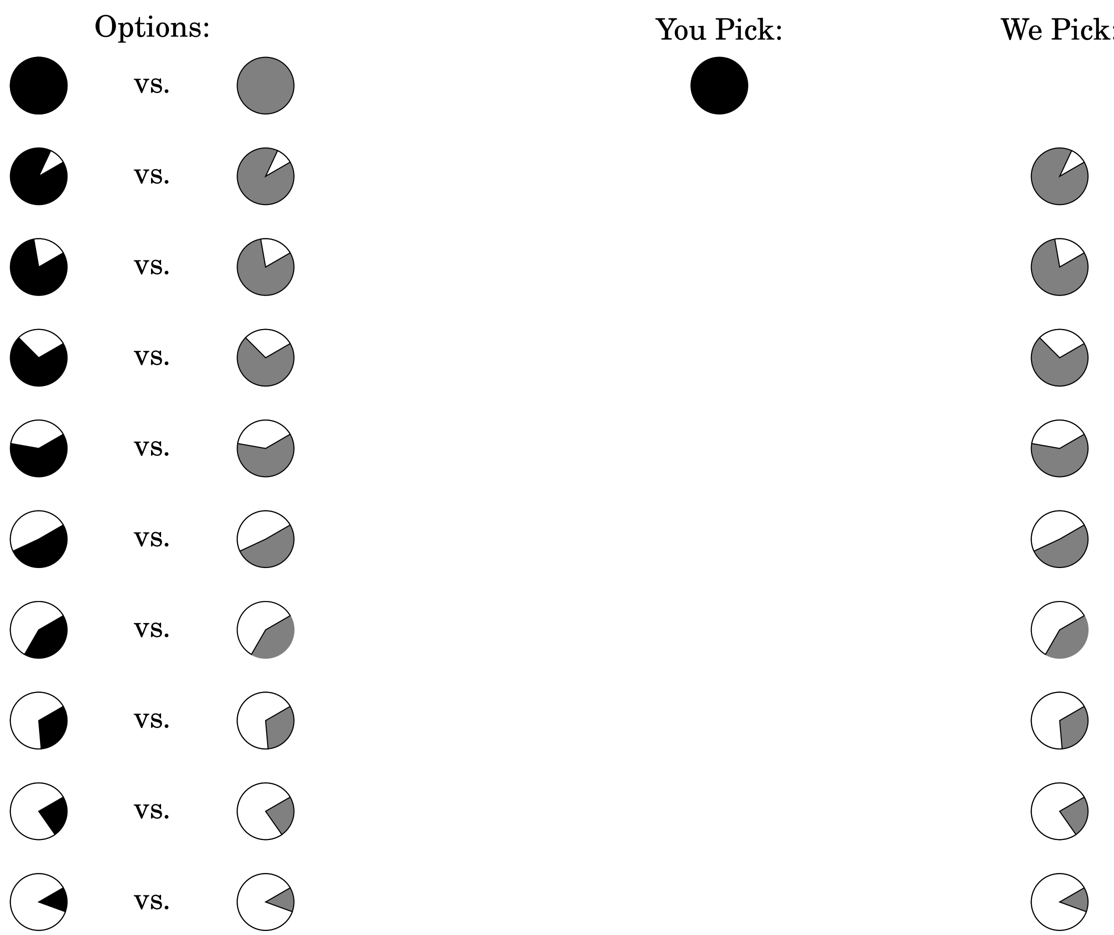


Figure XVII: $c-$Independence Rule

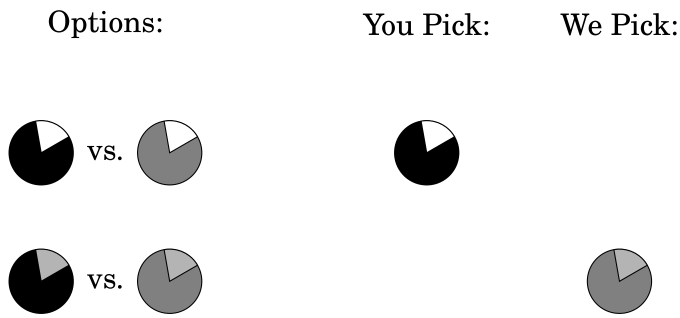


Figure XVIII: $c-$Branch Independence Rule

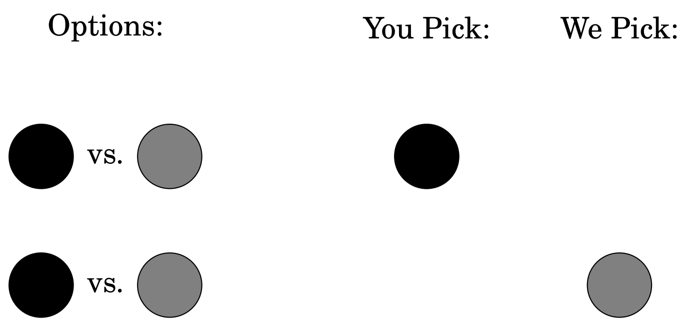


Figure XIX: $c-$Consistency Rule

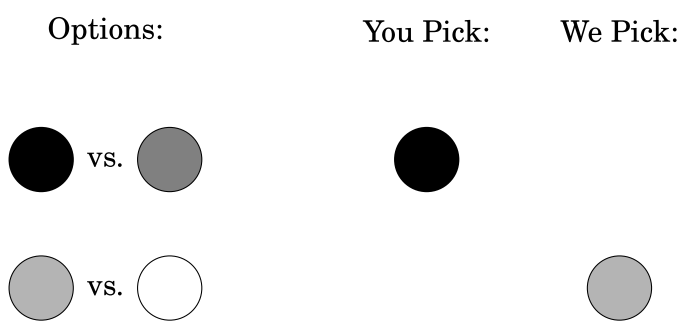


Figure XX: Distractor Rule #1

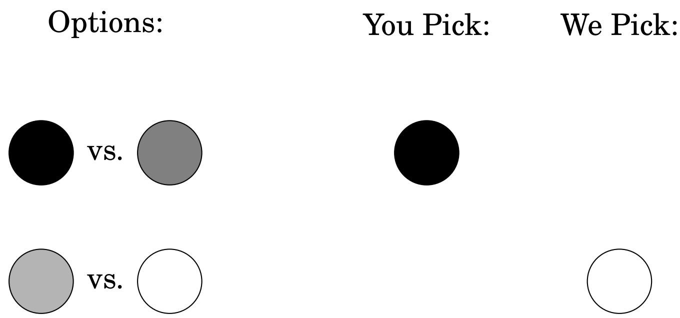


Figure XXI: Distractor Rule #2

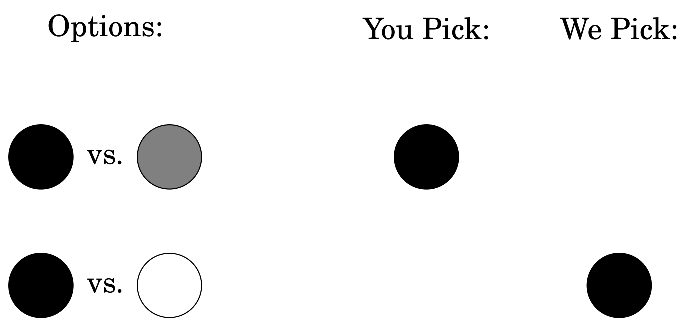


Figure XXII: Distractor Rule #3

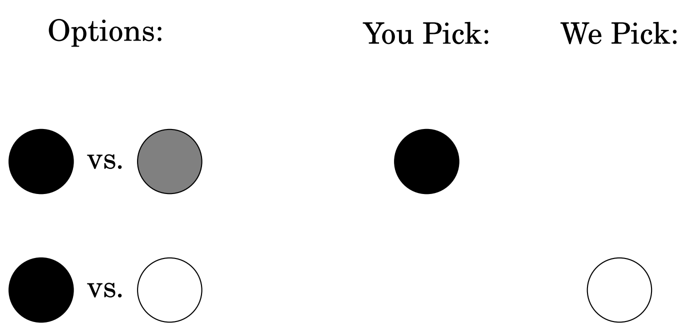


Figure XXIII: Distractor Rule #4

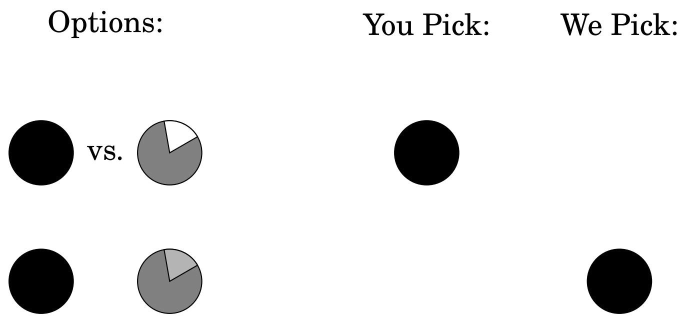


Figure XXIV: Distractor Rule #5

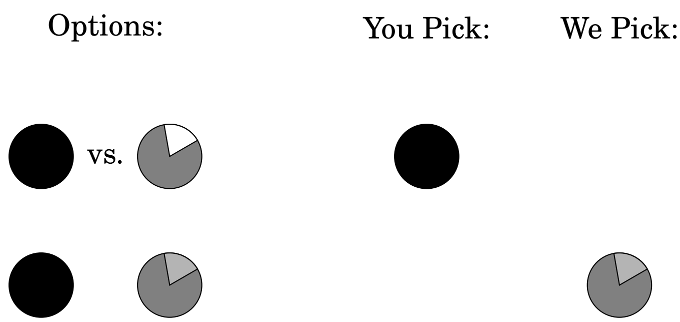


Figure XXV: Distractor Rule #6

INSTRUCTIONS

This is an experiment in the economics of decision making. Stanford University and the Ohio State University have provided the funds for this research. Feel free to ask questions while we go over the instructions. Please do not speak with any other participants during the experiment and please put away your cell phones and anything that you might have brought with you.

This experiment has three different parts, and each part has many decisions. These instructions are for the first part. You will be paid based on your choice in exactly ONE randomly selected decision from the entire experiment. Each of your decisions is equally likely to be paid. In addition, you will receive $7 for completing the experiment.

For each part, we will explain how you would be paid from a decision in that part. The decision chosen to determine your payment will be shown at the end of the experiment, and we will roll dice at the front of the room to determine which decision it will be.

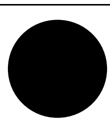
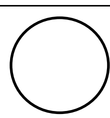
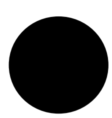
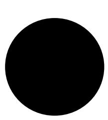
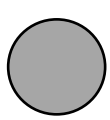
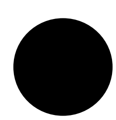
RULES

In this part of the experiment, you will be presented with various decision “rules.” These rules are abstract ways to represent choices, and we are interested in how people perceive them. Your task is to evaluate various rules and decide whether you would want to implement choices using the rule or make choices yourself. If you select your preferred rules carefully, will implement your preferred choices from the rules and you won’t have to make these choices on your own.

The rules will use colors to represent **positive money payments.** These payments could be “lotteries” or sure payments, which we’ll describe below. The possible payments range from $0 to $30. You will choose whether you want the rule to make decisions for you or whether you want to make the choices yourself. Remember, when you evaluate these rules, you should only be considering money payments, and only positive amounts (meaning you can’t lose money).

A “rule” is like an algorithm that observes your initial choices and then figures out what to pick later based on those initial choices. You can choose to use various rules to make your choices, or you can choose to make each later choice on your own.

For example, here’s one possible rule:

Options:	You Choose:	We Choose:
 vs. 		
 vs. 		

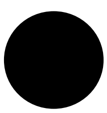
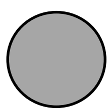
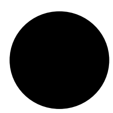
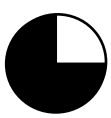
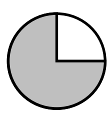
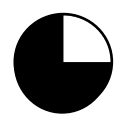
This rule establishes that, when given the choice between "black" and "white," you chose "black." Then, the rule says that if you were given the choice between "black" and "grey," we would choose black for you.

The colors don't have any inherent meaning (e.g. grey is not "in between" black and white). They can represent any possible money amount or lottery. As a result, you probably wouldn't want to select this rule. If you like black over white, whatever they may be, this doesn't necessarily mean you'd like black over grey, since grey can represent any possible payment. Grey might be $30 and black might be $5!

You want to select rules that would be always best for you, regardless of what the colors stand for. It's worthwhile to make the choice on your own if the rule would not always be best for you, like in the example above. This way, if "grey" were $30, you could choose grey over black. If instead grey were $0, you'd be able to choose black.

MIXTURES

Some of the rules will involve "mixtures." Here's an example.

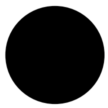
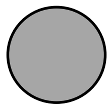
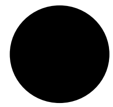
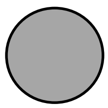
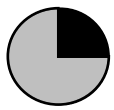
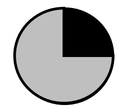
Options:	You Choose:	We Choose:
 vs. 		
 vs. 		

This rule says that if you prefer black over grey, then you would prefer a mix of black and white to the same mix of grey and white.

You can think of the mixtures like a spinner wheel. If the spinner wheel lands in the "white" region, you'd receive white. If it lands in the black (or grey) region, you'd receive black (or grey). Essentially, this rule would say that if you like black over grey, then you'd also like black-mixed-with-white over grey-mixed-with-white, regardless of what black, white, and grey are.

For example, imagine black and grey are both lotteries (remember, they can be any lottery), and imagine white is $10. If the spinner lands in the white region (which happens with probability ¼), you would receive $10. If it lands in the black or grey region (which happens with probability ¾), you would receive the original black or grey lottery.

Here's another example:

Options:	You Choose:	We Choose:
 vs. 		
 vs. 		

Many people would find this to be a desirable rule. If you like black more than grey, you might prefer the mix of black-and-grey to grey. Because there's a chance you can get black, which you prefer to grey.

TYPES OF QUESTIONS

In the experiment, the colored circles will represent various lotteries. These lotteries can have any payments from $0 to $30, and can have any probabilities. For example, maybe one lottery is a 100% chance of $20. Another lottery might be a 40% chance of $5, 30% chance of $11, and 30% chance of $16.

You won't know what the exact lotteries are while you're making your decisions about the rules, but the lotteries are sufficiently varied. Some might be very risky, some are safe, some involve low payments, some high, etc.

Therefore, when choosing your rules, you'll want to select only rules which are always best for you. By selecting rules which are always best for you, you'll get the option which is best for you without having to make the decision on your own. By not selecting rules which aren't always best for you, you'll ensure that you have the opportunity to pick your most preferred option in the cases where the rule wouldn't be good for you.

PAYMENT

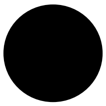
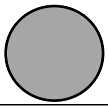
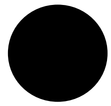
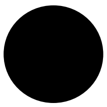
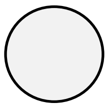
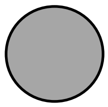
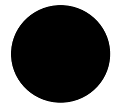
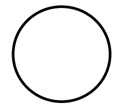
Along with all decisions you make in this experiment, each of your rule decisions is equally likely to be paid, and we will roll dice at the end of the experiment to determine which decision will be paid. If one of the rules is the randomly selected question to determine your payment, here's how we would pay you.

Each rule corresponds to some specific lottery questions. If you selected to make the choice on your own, you will be shown the lottery question(s) and you will make the choice on your own.

If you selected to implement your choices using the rule, we will make the choice for you based on what the rule says.

You will be paid based on the lottery chosen, either by you or by the rule.

For example, imagine you are being paid based on the following rule:

Options:	You Choose:	We Choose:
 vs. 		
 +  vs. 		 + 

There would be two lotteries corresponding to "black" and "grey." Remember, these lotteries could be anything, but let's say the two lotteries were

1. 100% chance of $12
2. 75% chance of $20, 25% chance of $10

Let's say you prefer option 2 (75% chance of $20, 25% chance of $10), so think of this as "black" and the other (100% chance of $12) as "grey." In this case, you preferred black over grey, as the rule assumes. Then, the rule says we will pay you "black & white" instead of "grey." The white lottery also could be anything, say it's 100% chance of $5. Then we would pay you "black & white," or 75% chance of $20, 25% chance of $10, as well as 100% chance of $5, instead of "grey," which would have been 100% chance of $12.

If you hadn't selected this rule, you would make the choice of black + white vs. grey on your own.

HOW TO ANSWER

There are no "right" or "wrong" answers in these questions. We are interested in people's preferences over these rules, so we simply want to know which rules you like and which you don't.

Since we want to know people's true preferences, the payment is set up such that you'll be paid your "favorite" thing if you answer truthfully. Therefore, you have no incentive to answer differently from what you really think. If you think the rule should describe your choices, you should select it (and then you'll save the cost of making the decision on your own). If you think there are situations where the rule would not give you your favorite option, you should not select it.

LOTTERIES

In this part of experiment, you will be making choices between “lotteries.” A lottery specifies the chance of receiving certain payoffs. In this experiment, the possible payoffs will range from $0 to $30. The chance of each payoff can be anything from 0% to 100%.

For example, one lottery could give you an 80% chance of $18, 10% chance of $7, and a 10% chance of $4. Another lottery might give a 100% chance of $13. There are many different possible lotteries.

In each decision, you will see two or three lotteries on your screen. Your screen will display the written values for the payoffs and probabilities in a table.

This is an example of what that could look like:

Option A:	Option B:
40% chance of $5 30% chance of $7 30% chance of $19	50% chance of $3 50% chance of $25

To visualize these lotteries, imagine rolling two 10-sided dice to generate a number from 1 to 100. In Option A, the first 40 numbers pay $5, the next 30 numbers pay $7, and the last 30 numbers pay $19. In Option B the first half of the numbers pay $3 and the other half pay $25. If this is the question to determine your payment, we would roll the dice and pay you according to which number is rolled.

Your task is simply to choose the lottery you prefer. The computer will record your choice and then will bring you to the next decision.

PAYMENT

You will be making 33 decisions. Each decision will be presented on a different screen. Along with all decisions you make in this experiment, each of your lottery decisions is equally likely to be paid, and we will roll dice at the end of the experiment to determine which decision will be paid.

For example, in the randomly selected problem, imagine you chose the lottery which gives

30% chance of $0

50% chance of $11.50

20% chance of $17

If we roll a number 1—30, you’d receive $0. There are 30 out of 100 possible numbers between 1 and 30, so this corresponds to a 30% chance of $0.

If we roll a number 31—80, you’d receive $11.50. There are 50 out of 100 possible numbers between 31 and 80, so this corresponds to a 50% chance of $11.50.

If we roll a number 81—100, you'd receive $17. There are 20 out of 100 possible numbers between 81 and 100, so this corresponds to a 20% chance of $17.

HOW TO ANSWER

Simply choose the lottery you prefer! In any given decision, if it's the decision to determine your payment, we will give you the lottery you selected. So you should select the lottery that you'd rather have determine your payment.

There are no "right" or "wrong" answers in these questions. We are interested in people's preferences over these lotteries, so we simply want to know which lotteries you prefer.

These lotteries are not the same questions as the lotteries which would determine your payment from your "rule" decisions.

RULES AND CHOICES

In the first part of the experiment, you indicated which "rules" you liked and which ones you did not like. In the second part of the experiment, you made choices over lotteries. It's possible that you selected a rule in Part 1 which could have been applied to make one of your choices in Part 2.

In this final part of the experiment, you will have the opportunity to reevaluate your choices. We will show you if you ever selected a rule and then made lottery choices which were not what the rule would have chosen for you. This way, you can analyze the rule and the choices together. You might decide that you want to change your lottery choices to align with what the rule would have chosen for you. Or you might decide that the rule isn't always true of your choices, so you can "un-select" it. Or, you can leave your choices as they were.

Here's how that will work.

RANKING OF THE RULES

First, for some of the rules you selected in Part 1, we will ask you to "rank" them in order of how much you would like to revise your choices if they conflict with the rule. An example is shown on the board. Essentially, we are asking you to rank the rules according to how important you think they are. The rule you think is most important would be ranked first, the next most important would be ranked second, etc.

At the bottom of the screen, there is also an option which says "I would rather have $1 than revise my choices for any of the remaining rules." This is an alternative which can enter into your ranking, too.

For example, let's say you selected six rules. In order of importance, you rank them

Rule 4, Rule 1, Rule 2, Rule 3, Rule 6, Rule 5.

Now suppose that you'd be willing to pay at least $1 to revise choices that conflict with rules 4, 1, 2, and 3, but you'd rather have an extra $1 than revise choices that conflict with rules 6 and 5.

Then, you would report a ranking of

Rule 4, Rule 1, Rule 2, Rule 3, $1, Rule 6, Rule 5.

If this ranking question is randomly chosen for payment at the end of the experiment, we will randomly select two of the things you ranked. We would pay you according to whichever you ranked *higher*. If that is one of the rules, we would pay you for your revised choices in that rule (more on this below). If that is the $1, you would receive a $1 bonus and would be paid for your original choices.

Your decision will be implemented only IF your choices are not in agreement with the rule. If your choices *are* in agreement with the rule, we will just implement your original decisions and you still get a $1 bonus. This means that you don't need to worry about whether you think your choices were in agreement with the rule or not. If they were, your decision in these questions won't matter (since we'll just implement your original choices). So you should report your ranking as if your choices conflict with the rules.

REVISING YOUR CHOICES

Regardless of your ranking, we will show you your rule/lottery decisions and you can choose to revise them or not. These decisions actually could be paid if you are paid for your revised choices, as described above.

To give you the opportunity to revise your choices, we will show you the rule on your screen and we'll show you your choices which are inconsistent with that rule. An example is shown on the board.

The choices you made in Parts 1 and 2 will be shown in red. Below that, you'll see an explanation of why the rule would have selected something different than what you chose. To change your choice of the rule, just click the box containing the rule. It will change from red to grey and vice versa. To change your lottery decisions, just click the box of the option you prefer instead. It will turn red, also. You can change any of your choices as many or as few times as you wish. Your final choices will be those in red.

You can revise your choices in any way you wish. For example, your choices being inconsistent with the rule might indicate that you don't like the rule after all and it is not a rule which is always best for you. In that case, you could choose to "un-select" the rule. On the other hand, you might decide that the rule does give you things which are always best for you, and therefore you want to change your choices to align with the rule. Finally, you might decide that you like the rule but also like your choices, even though they're inconsistent. In this case, you can leave all your choices as they are.

In other words, you can change any of your choices or keep them the same. You are under no obligation to make your choices agree with the rules, or vice versa.

PAYMENT

If one of your revised choices is selected for payment, we would pay you in the same way that we described in Parts 1 and 2. That is, if a revised rule were selected for payment, we would pay you just like how we described for the rules in Part 1. If you selected the rule *in your revised choices*, we would use the rule to make decisions for you. If you did not select the rule, you would make the decision for yourself.

If a revised lottery decision were selected for payment, we would pay you the lottery you selected *in your revised choices*. We would roll dice to generate a number between 1—100 to determine which payoff you receive.

HOW TO ANSWER

There are no "right" or "wrong" answers in these questions. We are interested in people's preferences over these rules and choices, so we want to know which options you prefer.

Since we want to know people's true preferences, the payment is set up such that you'll be paid your "favorite" thing if you answer truthfully. Therefore, you have no incentive to answer different from what you really think. If you think the rule should describe your choices, you should select it. If you think there are situations where it should not describe your choices, you should not. If you like one lottery over the other, you should choose that.

Your revised rule selections or your revised lottery choices could be selected to determine your final payment. If this is the case, we will pay you based on your revised rule or lottery choice, as described in Parts 1 and 2. Therefore, since your revised choices might be selected to determine your payment, you should choose the options you most prefer.

You are also under no obligation to make your lottery choices and rules agree. And, if you do want to change your decisions, you can change either the rule or the lottery choices depending on your preferences.