Revealed Incomplete Preferences*

Kirby Nielsen[†] Luca Rigotti[‡]

July 16, 2024

Abstract

We introduce a new methodology for eliciting incompleteness distinct from indifference. Subjects rank gambles, and we use these rankings to estimate preferences; payments are based on estimated preferences. About 40–50% of subjects express incompleteness, and revealed incompleteness is consistent with theoretical predictions. Incompleteness is similar for individuals with precise and imprecise beliefs, and in an environment with objective uncertainty, which is consistent with individuals having imprecise tastes. When we force subjects to choose, we observe more inconsistencies and preference reversals. Evidence suggests there is incompleteness that is indirectly revealed—in up to 98% of subjects—in addition to what we directly measure.

*We thank Marina Agranov, David Ahn, Gábor Békés, Federico Echenique, Simone Galperti, Guillaume Fréchette, Yoram Halevy, Paul J. Healy, Muriel Niederle, Lee Pang, Pietro Ortoleva, Charles D. Sprenger, Colin Sullivan, Emanuel Vespa, Ryan Webb, and Lanny Zrill for helpful comments and suggestions. This study was reviewed and granted exemption by the Institutional Review Board at California Institute of Technology.

[†] kirby@caltech.edu; Division of the Humanities and Social Sciences, California Institute of Technology

[‡] luca@pitt.edu, Department of Economics, University of Pittsburgh

> It is conceivable and may even in a way be more realistic to allow for cases where the individual is neither able to state which of two alternatives he prefers nor that they are equally desirable.
>
> von Neumann and Morgenstern, *Theory of Games and Economic Behavior*

> Is it rational to force decisions in such cases?
>
> Aumann, *Utility Theory Without the Completeness Axiom*

Completeness is one of the most fundamental axioms on preferences: Put simply, completeness states that an individual can rank any two alternatives presented to them. In contrast, when preferences are incomplete, there exist alternatives that the individual is unable to rank.[1] Completeness is so central to standard economic theory that "rationality" is often defined as completeness plus transitivity (as, for example, in the very first definition of Mas-Colell et al., 1995), suggesting an implicit normative component to the axiom. Despite the widespread reliance on completeness in models of choice, there have been few attempts to measure the extent to which it is a valid assumption on preferences.[2] We introduce a new method for identifying incompleteness, and find that about 40–50% percent of our subjects do not have complete preferences in a simple stochastic environment.

Incompleteness is inherently difficult to measure because decision-making experiments elicit *choices* and typically force individuals to choose between alternatives. One could attempt to identify incompleteness by looking for indicators that might correlate with it, or for choice patterns that could be manifestations of individuals' attempts to complete their incomplete preferences. The drawback with this approach is that it often requires specifying a model of how individuals complete their preference. Other approaches allow individuals to state incompleteness directly, but incentivizing this statement requires the experimenter to impose a mapping from reported incompleteness to outcomes, and this mapping could muddy the interpretation of any

[1] More formally, given a preference order $\succeq$, completeness states that between any two alternatives p and q, either $p \succeq q$, or $q \succeq p$, or both; preferences are incomplete if there exist p and q such that neither $p \succeq q$ nor $q \succeq p$.

[2] Notable early exceptions include Cohen et al. (1985), Cohen et al. (1987), and more recently Cubitt et al. (2015), Cettolin and Riedl (2019), Bayrak and Hey (2020), and Costa-Gomes et al. (2021). We discuss these papers, among others, in Section V.

reported incompleteness. For example, if the experimenter randomly picks an alternative when subjects report that they do not know which they prefer, then one could be capturing indifference or a preference for randomization rather than incompleteness.[3]

We build on existing methodologies to provide a relatively simple new framework for identifying indifference and incompleteness. Our objective is to elicit the *preference relation*, even when incomplete. To avoid the conceptual issue of how to pay subjects when they report incompleteness, we develop a new procedure that uses subjects' choices but does not pay them for any single choice directly. Instead, we ask subjects to rank lotteries, use their rankings to "estimate" their preferences, and then pay them based on what these estimated preferences predict they would choose in a question they have not seen and cannot influence. Under this procedure, subjects can state indifference directly. Furthermore, subjects can state that they do not know their preferred alternative, and these questions will not enter into our preference estimation. Given these incentives, subjects can answer questions that would accurately inform the preference estimation, but can also directly reveal when their preferences are incomplete. We discuss concerns of incentive compatibility with this procedure, but our data suggest that subjects report truthfully and use the response options as we interpret them.

Our experiment reflects an interpretation of incompleteness that we believe is important for applications: We let individuals communicate that they do not believe some of their choices should be included in the welfare relevant domain (Bernheim and Rangel, 2009). As analysts, we often have a goal of observing an individual's choices and then using these choices to infer underlying preferences so that we can predict future choices, analyze counterfactual environments, and make welfare comparisons. Our methodology directly asks individuals to reveal when they do not want us to use a particular choice to make inference about their underlying preferences. Observing incompleteness in this type of decision implies that individuals know the decisions in which they are unsure about their preference, and suggests that the inability to reveal incompleteness could lead the analyst to inferences and predictions that are welfare-reducing.

We first validate our methodology by comparing it to an established procedure put

[3]This is especially an issue with subjective uncertainty since subjects can use randomization to eliminate ambiguity (Baillon et al., 2022).

forth by Cettolin and Riedl (2019). In this experiment, we ask individuals a sequence of choices comparing bets on an ambiguous urn to bets on various risky urns. In one block, individuals are incentivized with the Cettolin and Riedl (2019) methodology: They can choose either the ambiguous or risky urn, or they can indicate that they are "indifferent" between the two, where this option results in the experimenter randomly selecting one of the two urns. In the other block, individuals are incentivized with our methodology: They can choose either the ambiguous or risky urn, they can indicate (exact) indifference, or they can state that they "do not know" which urn they prefer. As described above, we use the choices in which individuals express strict preference or indifference to estimate their preferences (here, a belief about the distribution of balls in the urn), and we pay them for a different bet using this implied belief.

We conduct the two treatments within-subject, so the comparison of choices across treatments allows us to test whether our methodology produces results in line with the more standard Cettolin and Riedl (2019) elicitation and interpretation (CR inventives henceforth). We find strong correlations: When individuals express a strict preference for either the ambiguous or risky urn under the CR incentives, they report the same strict preference under our incentives upwards of 85% of the time. About 2/3 of the individuals report "indifference" at least once using CR incentives, and our methodology can separate indifference from incompleteness for these individuals. We identify the exact indifference for 76% of them, and we can identify incompleteness separate from indifference for 63%.

With this validation in hand, we use our methodology in a richer choice environment to understand incompleteness and how indifference and incompleteness relate. We conduct a simple online experiment with over 1,300 total subjects across multiple treatments, and we ask these subjects to compare monetary lotteries over a random binary event. We find that 39% of our main sample express incompleteness in at least one comparison in this environment. Thus, a non-trivial minority of individuals have incomplete preferences. However, in any given binary choice, incompleteness is quite rare—individuals express incompleteness in only about 3% of comparisons. Even subjects who do reveal some incompleteness do so in only about 7% of comparisons. So while many individuals have incomplete preferences, the extent of incompleteness is small in our environment. Nevertheless, this incompleteness is systematic and largely conforms to simple predictions from theoretical formulations of incomplete preferences. Furthermore, we show that individuals report incompleteness for lotteries that are more complex according to recently-posited measures of complexity (Enke and Shubatt, 2024), and that response times are slowest when individuals report incompleteness. We also allow individuals to report explicit indifference, and we show how incompleteness and indifference are distinct.

While our experiment can be regarded as a commentary on incompleteness in general rather than a test of a particular theory, we designed it to reflect the growing theoretical literature on incomplete preferences in stochastic environments.[4] This literature typically distinguishes between two possible sources of incompleteness: imprecise beliefs and imprecise tastes. To fix ideas, take an individual who does not know which insurance policy they prefer. They might be unable to compare two policies because they are unsure how likely they are to fall ill in the coming year (imprecise beliefs), or because they are unsure of their risk tolerance (imprecise tastes). Our experiment mirrors this theoretical distinction by asking subjects to make choices between lotteries in a domain of subjective uncertainty. We design binary gambles specifying payoffs that subjects would receive if the Merriam-Webster Dictionary word-of-the-day on a future date would be a verb or not a verb. Subjects can form a subjective belief about the likelihood that the word-of-the-day will be a verb, but they might not be certain about this probability. This allows for incompleteness due to imprecise beliefs, and we identify belief imprecision by eliciting individuals' subjective belief as well as their certainty about this belief in the form of an unincentivized range of beliefs, following Giustinelli et al. (2019).

47% of subjects report uncertainty about their beliefs. Of these subjects, 43% directly reveal incompleteness. However, among the remaining 53% of subjects with certain beliefs, 36% also directly reveal incompleteness. Thus, uncertainty in beliefs does not seem to be the primary source of incompleteness which suggests that imprecise *tastes*—in our case, imprecise risk preferences—-are the main source of incompleteness in our data. We confirm this by running our exact same experiment with lotteries defined over a comparable objective event. We find that the same percentage of subjects report incompleteness in this treatment, and there is only a very small reduction of incompleteness at a comparison-level. Thus, we conclude that a main driver of incompleteness seems to be imprecise tastes rather than imprecise

[4]See, for example, Aumann, 1962; Bewley, 2002; Dubra et al., 2004; Eliaz and Ok, 2006; Gilboa et al., 2010; Ok et al., 2012; Galaabaatar and Karni, 2013; Karni, 2021.

beliefs.

An inherent feature of our methodology is that we do not force the subjects to make a choice. This is in contrasts with the standard "Forced Choice" paradigm of most decision-theoretic inspired experimental work where the researcher asks subjects to choose between two lotteries without allowing them to express indifference or incompleteness. We compare the standard Forced Choice environment to our "Non-Forced" Choice environment using a within-subject elicitation that gives us the ability to detect the extent to which incompleteness could affect the inferences one makes about behavior in forced choice environments (Costa-Gomes et al., 2021). In the Forced Choice treatment, each subject faces the same comparisons as in the Non-Forced treatment, but is asked to make choices without the option of expressing incompleteness or indifference. If preferences are incomplete but we do not give subjects the opportunity to reveal their incompleteness, then we should not be surprised when forced choices exhibit preference reversals or violations of basic properties like transitivity. We find that transitivity violations are far more common in the Forced treatment, and that incompleteness and indifference can explain about a third of these violations.

Preference reversals could indicate that some individuals have underlying incompleteness that they are unaware of or that they do not reveal, and indeed some papers interpret preference reversals and randomization as incompleteness (Eliaz and Ok, 2006; Bayrak and Hey, 2017). We identify this by looking for cases of "clear" preference reversals—comparisons in which subjects report a strict preference in the Non-Forced treatment and report the opposite preference in the Forced treatment. We refer to this as incompleteness that is "indirectly revealed." 95% of our subjects indirectly reveal incompleteness, and only 2% of subjects neither directly nor indirectly reveal incompleteness.

Interestingly, we find that the rates of directly and indirectly revealed incompleteness in any given question are highly correlated across the population. That is, the questions where subjects are most likely to exhibit preference reversals are the same questions in which other subjects are most likely to report incompleteness directly. This lends credibility to the interpretation of preference reversals reflecting underlying incompleteness, and lends further credibility that our methodology accurately elicits the comparisons for which subjects are not sure of their preference.

We see a few important implications of our results. First, incompleteness is important to understand for the reliability of two main goals of economics: assessing welfare and predicting behavior. It is important to understand when individuals are unsure of their choice and would prefer it not to be used as indication of their preferences, which is exactly the motivation behind our elicitation mechanism. Furthermore, when individuals are unsure of their choice, it is likely the case that this choice and its implications on preference are not accurate predictors of future decisions. Identifying incompleteness in turn can yield better predictions.

Second, understanding the source of incompleteness can assist in targeting interventions to help individuals make decisions. Thinking back to our example of an individual who is unable to decide between two insurance plans, this could be because they are unsure about their beliefs, or it could be because they are unsure about their risk aversion. These two sources of incompleteness imply different policy interventions. In particular, belief uncertainty might call for targeted information provision. However, if incompleteness stems from preference uncertainty instead—as we find in our data—then these information interventions would be for naught.

Finally, our objective lotteries are some of the simplest decisions we can ask individuals to make. We see incompleteness even in this environment, and it is natural to conjecture that incompleteness would only increase in more complicated settings. Indeed, we find that 76% of subjects directly reveal incompleteness in a treatment where we make the choice objects more complex and ambiguous (a large increase from the 39% of subjects in our main data). This suggests that the rates of incompleteness that we see in our main data represent a lower bound on the extent of incompleteness in choice. In addition, we find that forced choice leads to less-coherent preferences even in this simple environment, and we conjecture that forced choice would lead to even more inconsistencies in more complex environments. We leave further study of this to future work, but it would be interesting to extend analysis into more naturalistic choice environments with more complex comparisons. We discuss this, and other related open questions, in Section VI.

I. Conceptual Overview of Methodology

Our goal is to identify indifference and incompleteness, but this poses a methodological challenge. In theory, we could simply ask individuals to tell us when they are indifferent or that do not know their preference. However, without incentives to answer carefully and honestly, we might doubt that individuals' responses reflect their underlying preferences. We attempt to keep this simple and straightforward elicitation procedure—directly asking individuals to report indifference and incompleteness—but introduce some incentives to answer carefully and honestly by mapping these reports into payment.

To do this, we adapt the basic idea of the elicitation used in Krajbich et al. (2017) and Kessler et al. (2019), where individuals are asked to make hypothetical choices, and adjust it to a setting that allows for reports of indifference and incompleteness. In particular, we present participants with binary choices over gambles, and they can respond by indicating that they strictly prefer one of the gambles, that they are indifferent between the two, or that they do not know which gamble they prefer. We tell participants that their responses will teach an "algorithm" the kinds of gambles that they prefer, and we pay them based on what the algorithm predicts they would choose in a question they never face. So, we estimate subjects' preferences based on their answers, and use the estimated preferences to make an actual choice between two gambles the subjects have not faced before. The main idea behind the incentive is that reporting strict preference or indifference will help the algorithm to better understand the subject's preferences over gambles, but if they report incompleteness, then this question does not enter into the algorithm, so they are not forced to respond when they are unsure.

We explain this to participants at a high level, but generally do not explain the details of the algorithm. For example, below we reproduce part of the instructions for Experiment 1:

> If you are selected to receive a bonus payment, and if you are selected to be paid for this part of the study, then we will use your answers to guess what you would prefer in another question and we will pay you based on what we think you would prefer in that question. We have a tested algorithm that uses your actual answers to these questions to understand your preferences between the two urns and guesses which urn you'd rather bet on. Importantly, your answers cannot affect which question you are paid for, but can only affect which Urn we draw a ball from in the unknown question.

After this, we tell participants how the algorithm uses strict preference or indifference to understand their relative preference between the two "urns" in the decision problems, and that we will not use reports of indifference to train the algorithm.

We will discuss the more precise details of the algorithms in Sections II and III, and the precise details of the algorithm determine the assumptions under which it is theoretically incentive compatible. For example, in Experiment 1 we use choices to estimate a most pessimistic belief under max-min preferences, so the algorithm would be strictly incentive compatible for subjects whose choices are consistent with max-min decision-making (Gilboa and Schmeidler, 1989); as another example, one of our algorithms in Experiment 2 uses MLE to estimate a CRRA utility function, so the elicitation would be incentive compatible under those assumptions. Since we do not give participants much details on the algorithm—and since they would likely not understand the details anyway—we cannot rule out that subjects form beliefs about these algorithms such that they believe manipulation is profitable. In other words, this methodology might not be "theoretically" incentive compatible for all participants, as was the case in Krajbich et al. (2017) and Kessler et al. (2019). Because of this, we designed our experiment to be able to detect manipulation and untruthful reporting in a few ways. We discuss this evidence throughout and we organize all of the results that we use to validate the elicitation in Section IV.F. The results overwhelmingly suggest that subjects report truthfully.

Danz et al. (2022) show that withholding details on the exact payment incentives—while still telling individuals that truthful reporting is in their best interest— encourages more truthful reporting than giving all of the details of the binarized scoring rule in a belief elicitation task. Thus, there seems to be evidence from the literature that individuals do report truthfully when there is no obvious incentive not to. Thus, our results are consistent with our mechanism falling under this umbrella of "*behaviorally* incentive compatible" mechanisms (Danz et al., 2022), even if not theoretically incentive compatible for all possible beliefs and preferences. We liken this to other methodologies such as dynamically optimized sequential experimentation (DOSE, Wang et al., 2010) and the other papers in the literature that have used preference estimation without entirely fixing subjects' beliefs about the incentives (Krajbich et al., 2017; Kessler et al., 2019).

II. Experiment 1: Validation of Methodology

As an initial demonstration of our methodology, we compare our approach to the established approach of Cettolin and Riedl (2019). We use this exercise to demonstrate that individuals understand our methodology, that the incompleteness they reveal is similar to patterns that have been interpreted as reflecting incompleteness through the Cettolin and Riedl (2019) methodology, and that our methodology allows us to collect more data on preferences by disentangling indifference from incompleteness.

II.A. Experimental Design

We conducted two treatments within-subject in randomized order for each participant.[5] We collected data from 199 participants recruited through Prolific. Participants were paid $2 for completing the experiment (they took fewer than 10 minutes on average). In addition, each participant had a 10% chance of being randomly selected to receive a bonus payment based on their decisions in the experiment. We explain how these bonus payments would be determined in detail below.

One of the treatments nearly-exactly replicates Cettolin and Riedl (2019)—hereafter referred to as the "CR block"—while the other treatment uses the same experimental setup but uses our method of eliciting preferences—hereafter referred to as the "NR block." In both blocks, individuals are introduced to two urns that each contain 100 balls. The ambiguous urn contains an unknown proportion of red and black balls, while the risky urn varies in known color composition. Participants first choose whether to bet on a red or black ball, and then fill out a price list that asks them 21 questions. In each question, a participant decides whether to bet on a ball of their chosen color being drawn from the risky urn specified in that question or from the ambiguous urn. The risky urn composition changes in each row; specifically, the risky urns range from 100 red balls to zero red balls in increments in five.

The two treatments differ in the choice options available to participants and in the way choices are incentivized. In the CR block, in addition to choosing either the risky or the ambiguous urn, individuals can state that they are "indifferent between the two urns." Cettolin and Riedl (2019) argue and demonstrate that multiple

[5]Because we conducted the treatments within-subject, we included a "distractor" task in between the two. In this task, we presented individuals with various different images and asked them to click on the part of the image that first stood out to them.

expressions of indifference reflect incompleteness, since true indifference should be expressed in only one bet. If a participant were selected to receive a bonus payment, and if we randomly determine to pay them for the CR block, then we randomly select one row from the price list and pay them for their selected bet, either from the risky or the ambiguous urn. If a participant indicated indifference, then, following Cettolin and Riedl (2019), we select either the risky or ambiguous urn with equal chance and implement a bet from that urn. If a ball of the participants' chosen color were selected from the chosen urn, then they would receive a bonus payment of $5; otherwise they would receive a bonus payment of $0. Participants were informed of this procedure, and, to contrast with the NR block, we emphasize to participants that they can indicate that they are indifferent as many or as few times as they wished. We followed the instructions from Cettolin and Riedl (2019) very closely.

In the NR block, we use the elicitation procedure outlined above in Section I. Specifically, in addition to choosing either the risky or the ambiguous urn, individuals can state that they are "indifferent between the two urns" or that they "do not know which urn (they) prefer." We tell participants that, if they were to be paid for this block, then we would use their responses to "estimate their preferences" and use these estimated preference to pay them for a separate question. We emphasize that their answers cannot affect the *question* that they are paid for, but can only affect the urn from which we will draw a ball in the unknown question. We tell them that, if they indicate strict preference or indifference, then an algorithm will use this information to understand their relative preferences between the two urns, and if they indicate incompleteness, then we will not use that question to train the algorithm. The exact language of this can be found in Appendix Section C. To contrast with the CR block, we emphasized to participants that they could indicate that they are indifferent a maximum of once, and that they could indicate that they do not know as much or as few times as they wished. We chose to allow only a single stated indifference—in contrast with our subsequent experiments—for theoretical accuracy and to make the difference in incentives between the two blocks more salient to participants.[6]

In practice, our "algorithm" in this experiment is extremely simple. We use a participant's responses to infer their subjective belief of the proportion of balls in the

[6]We ran a pilot session that allowed for multiple statements of indifference in the NR block. The results were very similar.

ambiguous urn. For monotonic individuals who never report incompleteness, this is straightforward. For non-monotonic individuals this is less straightforward, but these cases are rare as we will show below. Most importantly, for individuals who report a range of incompleteness, we ascribe max-min preferences and assume their most "pessimistic" belief (Gilboa and Schmeidler, 1989). For payment, we use this estimated belief to infer a participant's choice in a question very similar to the ones answered throughout the experiment but where the number of balls in the risky urn was different.[7]

II.B. Results

We focus on the within-subject comparison of the two blocks for validation of our methodology. Figure I shows within-subject and within-question comparisons of choice responses between blocks, conditional on response in the CR block. For example, conditional on choosing the risky urn in a given question in the CR block, Figure I shows that individuals also choose the risky urn in that question in the NR block 91.5% of the time. Similarly, when choosing the ambiguous urn in the CR block, individuals choose the ambiguous urn in the NR block 85.3% of the time. Thus, when expressing strict preference, the two methodologies overwhelmingly agree.

Individuals use the option to express indifference and/or incompleteness often in both blocks. 69% of individuals report "indifference" in the CR block at least once, and 74% report indifference or incompleteness at least once in the NR block.[8] Separating responses in the NR block, 59% of individuals report indifference and 50% of individuals report incompleteness. These measures are also highly correlated across blocks. At a question level, conditional on reporting indifference in the CR block, individuals report either indifference or incompleteness in the NR block about 50% of the time.[9] Recall that individuals could only report indifference at most once in

[7]This is the minimal deviation we could imagine while truthfully informing participants that they would be paid for a different question. It allows us to keep the stakes of the bet constant, use the "same" ambiguous urn and therefore the same subjective belief, and we assume that individuals' preferences involving risk only depend on the probabilities and not the number of balls.

[8]77% of subjects select the indifference option at least once in the baseline "Experiment *Risk-Ambi*" of Cettolin and Riedl (2019), so our results are quite consistent with theirs.

[9]On the other hand, it is true that individuals report a strict preference in the NR block about half the time when they reported indifference in the CR block. One might worry that this reflects unwillingness to express indifference or incompleteness under our incentives, but this does not necessarily appear to be the case. Conditioning the other way, when individuals report incompleteness in the NR block, only 55% of the time do they report indifference on that question in the CR block.

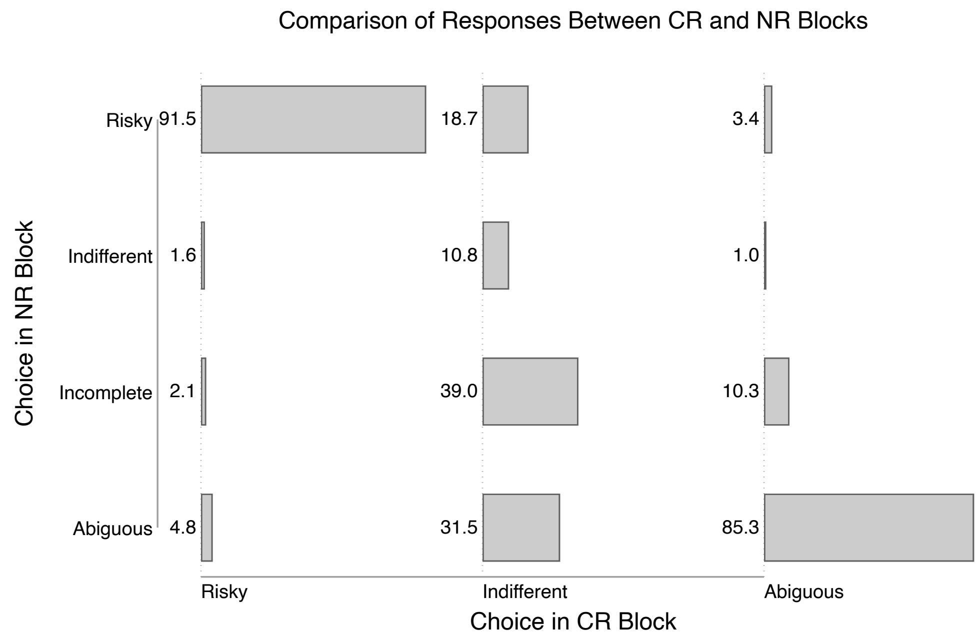

Figure I: Comparison of CR and NR Methodologies
Note: The figure reports the within-subject comparison of responses between the CR and NR blocks. Each choice a participant makes in the CR block is classified as choosing either the risky urn, the ambiguous urn, or stating indifference between the two. Each choice a participant makes in the NR block is classified as choosing either the risky urn, the ambiguous urn, stating indifference, or stating incompleteness. The figure correlates these responses for each subject and each question.

the NR block, so this enables us to identify the single row in which individuals were exactly indifferent, in contrast to the rows in which they expressed incompleteness. Among individuals who ever reported indifference in the CR block, we identify the exact indifference for 76% of individuals and we identify incompleteness separate from indifference for 63%.[10]

We also find similar patterns of expression across the two methodologies. In particular, in the CR block, conditional on reporting any indifference, individuals do so in a single or in consecutive rows 88% of the time. When we consider individuals who express both indifference and incompleteness in the NR block, 84% also express incompleteness and indifference "continuously," meaning that all indifferent and in-

[10]We also find that the row in which individuals were most likely to report indifference was the row in which the risky urn contained 50% winning balls. If individuals were to apply the principle of insufficient reason and act as if the ambiguous urn contained 50% winning balls and 50% losing balls, then this is exactly where we would expect them to be indifferent.

complete choices lie in consecutive rows of the price list.[11] This is a similar pattern to the CR block, further validating our methodology.

III. Experiment 2: Understanding Incompleteness and Indifference

After validating our methodology in Experiment 1, we turn to Experiment 2 to deepen our understanding of incomplete preferences. We designed this experiment with four broad goals: 1. capture the extent of incompleteness and indifference in individuals' preferences, 2. understand how indifference and incompleteness manifest, 3. analyze how forced choice affects inference on preferences relative to non-forced choice, and 4. identify the extent to which incompleteness results from imprecise beliefs relative to imprecise tastes. To do so, we designed an experiment in which participants make binary choices over monetary lotteries. In the following subsections, we first describe the event structure, lottery choices, and treatments, and then describe the algorithms we use.

Throughout the paper, we label incompleteness described by multiple probabilities as "imprecise beliefs" theory and incompleteness described by multiple utilities as "imprecise taste" theory. We formalize this in Appendix Section A. We design Experiment 2 partially to investigate these two sources of incompleteness. Our main environment is one of subjective uncertainty, allowing for incompleteness to result from imprecise beliefs and/or imprecise tastes. We attempt to disentangle these two channels in two ways. First, we classify subjects as having precise or imprecise beliefs based on a self-reported measure of belief precision. We say that, for subjects with precise beliefs, incompleteness can only result from imprecise tastes while for subjects with imprecise beliefs, incompleteness can result from imprecise beliefs, imprecise tastes, or both. Second, we exogenously eliminate belief imprecision in a treatment with objective uncertainty. This allows us to rule out incompleteness due to imprecise beliefs, leaving only incompleteness that results from imprecise tastes; we discuss this in detail below.

To overview the experimental environment, in our main results, we analyze data from a total of 639 participants recruited through Prolific, an online participant re-

[11] Within these continuous individuals, 81% express indifference at the "edge" of their incompleteness region while the remaining 19% express indifference in the interior of their incompleteness region.

cruitment platform.[12] We paid subjects \$5 for completing the experiment. Subjects took 35 minutes to complete the experiment on average, so the ~\$9/hr wage is commensurate with other studies on Prolific where researchers at the time were required to pay at least \$6.50/hr. In addition, each subject had a 10% chance of being randomly selected to receive a bonus payment based on their decisions in the experiment. The average bonus payment among those who received a bonus was \$12.10. We will describe in detail below how individuals' decisions affected their potential bonus payment. The bonus payment incentives were intended to encourage individuals to respond carefully and truthfully.

III.A. The Event

To understand incompleteness in a setting where uncertainty might feel more natural than balls and urns, we chose a novel event structure over which to design our lotteries. The payoff of each lottery was determined by the part of speech of the Merriam-Webster Dictionary word-of-the-day on a pre-specified future date.[13] For example, one lottery would pay \$2 if the word-of-the-day in 3 days is a verb, and would pay \$14 if the word-of-the-day in 3 days is not a verb. We provided subjects with a list of previous words of the day and their parts of speech for a past month to give them a sense of the frequency of each part-of-speech. We do not provide subjects with the empirical frequency of these parts of speech for either the Merriam-Webster word-of-the-day or for the English language in general, and this was not easily found on the internet to the best of our knowledge.[14]

There are a few reasons why we used a subjective event in our main experiment. First, one of our goals in the paper is to understand the role of imprecise beliefs

[12]We restricted to participants of US nationality, with at least 98% approval rate, who had participated in at least 50, but no more than 500, previous studies on Prolific. These 639 subjects were collected in three waves. In May 2021, we recruited 119 subjects. We self-replicated our exact experiment, recruiting 382 additional subjects in November 2021. Results from the two waves are statistically indistinguishable, so we pool the data. We recruited 138 additional subjects in August 2022 using a different elicitation algorithm, which we discuss in Appendix Section C. As shown in the Appendix, these results are also statistically indistinguishable, so we combine all three waves for our main analysis.

[13]The Merriam-Webster Dictionary posts a "word-of-the-day" every day, intended to teach people new words.

[14]In the list that we provided to subjects, 10 out of 30 words were verbs. For this reason, and because we expected individuals to think of nouns, verbs, and adjectives as the three main parts of speech, we calibrated our lotteries to target a region of incompleteness around one-third. As shown in our results, a large mode of subjects do report subjective beliefs near one-third.

versus imprecise tastes in generating incompleteness in preferences. Therefore, we chose an event that was likely to lead to some belief imprecision. Second, most experiments that try to measure incompleteness use objective uncertainty (we review these papers in Section V). Thus, a contribution of our paper is to study incompleteness over simple monetary gambles, but allowing for individuals to form their own—potentially uncertain—subjective beliefs. This also allows us to compare incompleteness in objective versus subjective environments using the same experimental setup as we discuss below.

We chose this specific event for a few reasons. Especially because we ran the experiment online, we wanted an event that was not controlled by the experimenter. We felt this might help subjects trust the ambiguous process of determining the state, without worrying that the experiment was "rigged" in some way. We also felt this would avoid issues of "comparative ignorance" or other concerns about the experimenter knowing the state while subjects did not (Fox and Tversky, 1995).

At the end of the experiment, we elicit participants' subjective belief that the word-of-the-day on the pre-specified future date would be a verb.[15] Following this, we ask subjects whether they are *certain* of this belief or not, and they respond with a simple yes or no answer. If they answer "no," then we allow them to specify a range of beliefs in addition to their point estimate. We followed the elicitation procedure from Giustinelli et al. (2019), who elicit precise and imprecise beliefs about individuals' likelihood of developing late-onset dementia. This elicitation was unincentivized.[16]

As we explain in more detail below in Section IV, we use subjects' reported certainty in their belief to identify whether any observed incompleteness results from imprecise tastes or imprecise beliefs. For participants who are certain of their belief, we assume incompleteness must result from imprecise tastes. For participants who are uncertain of their belief, we assume incompleteness can result either from imprecise tastes, imprecise beliefs, or both.

[15]We discuss payment from the experiment below, but if a participant were randomly selected to have their belief report determine their bonus payment, then we incentivized this report using the standard incentive-compatible mechanism studied by Karni (2009).

[16]Karni (2018) and Karni and Vierø (2020) provide interesting elicitation methods for eliciting sets of subjective probabilities. To the best of our knowledge, these have not been validated behaviorally yet, and it would be interesting to compare these mechanisms to unincentivized reports. Given that our experiment was already quite complex, we use the simpler procedure from Giustinelli et al. (2019) and rely on exogenously inducing certain beliefs to confirm our results.

III.B. Lotteries

Subjects make binary choices over lotteries that specify payoffs to be received if the word-of-the-day is a verb or not a verb. We denote a lottery by (nv, v) where $\$nv$ is the payoff the subject would receive if the word-of-the-day is not a verb and $\$v$ is the payoff they would receive if the word-of-the-day is a verb. There are two within-subject treatments (described below). In each treatment, subjects faced two separate blocks of 25 questions each. In a block, subjects compare 25 lotteries to a "reference lottery" for a total of 25 comparisons per block. Figure II shows the two reference lotteries—(9, 11) and (14, 2)—and the 23 other lotteries used in the experiment; we also list all lotteries in Table VI in the Appendix. We chose one reference lottery to be fairly symmetric and the other asymmetric across states. Note, the reference lotteries were themselves comparison lotteries, so we asked subjects to compare each reference lottery to itself, and asked subjects to compare the two reference lotteries to each other.

We chose these specific lotteries, rather than randomly-generated lotteries or other methods of selection, to target a region of likely incompleteness. In particular, we considered a range of beliefs around $pr(verb) = \frac{1}{3}$ with linear and log utility functions. For reference, this is visualized in the Appendix in Figures VII and VI. In addition to targeting incompleteness, we included some lotteries related by dominance, and some that would be comparable by most preferences.

To subjects in the experiment, we referred to the lotteries as "gambles." We did not make a distinction between reference and comparison gambles. Instead, we said that one gamble would stay the same across a block while the other varied.

III.C. Treatments

We had two within-subject treatments. The "Non-Forced" treatment allowed subjects to report strict preference, indifference, and incompleteness. Specifically, subjects could respond with one of four options, reported verbatim below:

1. I rank Gamble 1 above Gamble 2
2. I rank Gamble 2 above Gamble 1
3. I rank Gambles 1 and 2 exactly the same
4. I don't know how I rank Gambles 1 and 2

We interpret the first two options as strict preference, the third as indifference, and

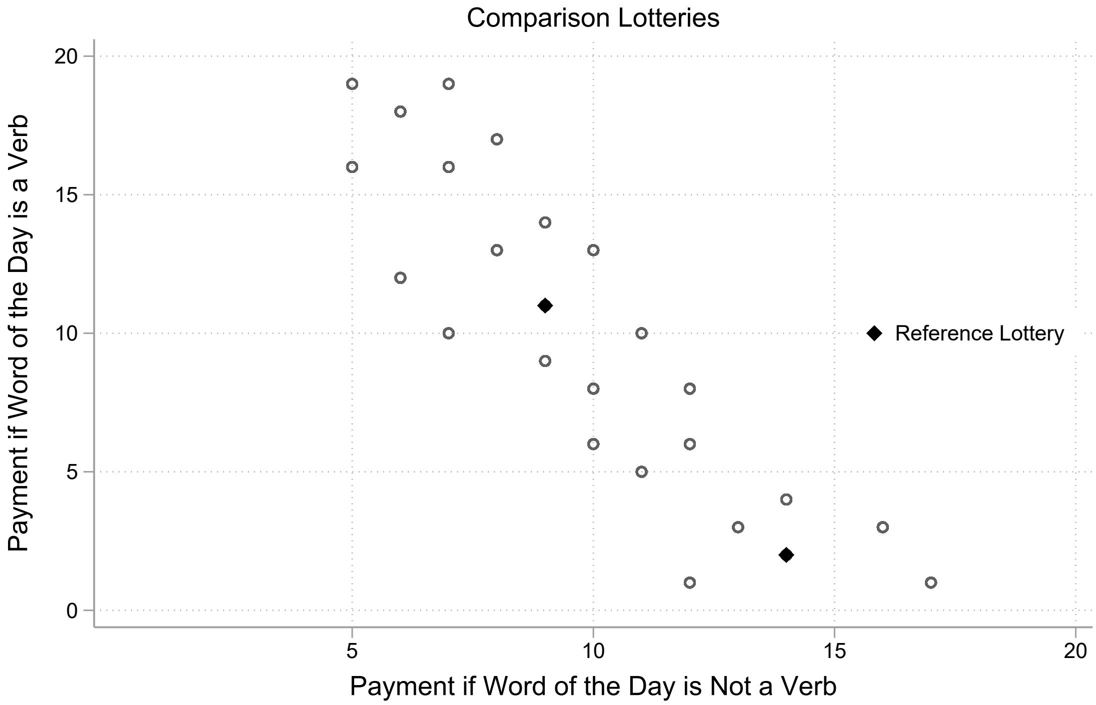

Figure II: Comparison Lotteries
Note: Points show the payoffs associated with the 25 lotteries used in the experiment. The black diamonds show the reference lotteries, $(9,11)$ and $(14,2)$. Both reference lotteries were themselves comparison lotteries, so we had subjects compare each reference lottery to itself, to the other reference lottery, and to the 23 lotteries represented by the open circles.

the fourth as incompleteness. The order of the four options on subjects' screens was randomized independently across each question. We include a screen shot in Appendix Figure VIII.

The "Forced" treatment allowed subjects to report only one of the first two answer options above. While we could have retained the indifference option in the Forced treatment, we intended this treatment to mirror the vast majority of experimental elicitations which do not include the option to report indifference. Again, the order of the options was randomized independently in each question. We told subjects that if they did not know how they ranked the gambles, or if they ranked them exactly the same, then they should "choose one of the two possibilities that (they) think fits best."

Subjects saw the Forced and Non-Forced treatments in random order. Within each treatment, we randomized the order of the two reference lottery blocks. Within each block, we randomized the order of the 25 lotteries. Thus, in total, subjects made 100 binary choices between gambles (25 lotteries × 2 reference lotteries × 2 treatments).

III.D. Payment

As mentioned above, each subject had a 10% chance of being randomly selected to receive a bonus payment based on their decisions in the experiment. If selected, then the subject was paid based on either the Non-Forced treatment, the Forced treatment, or their reported belief that the word-of-the-day is a verb.

If they were paid for the Non-Forced treatment, then we implement our estimation procedure that we describe in detail below. If they were paid for the Forced treatment, then we randomly selected one of the 50 lottery choices and paid them the lottery they chose in this decision. This is an entirely standard incentivization for binary choices, so subjects should choose their preferred lottery in each binary decision. If they were paid for their reported belief, then we paid them according to the Karni (2009) procedure, for a bet worth $5. This is also entirely standard and incentive-compatible under minimal assumptions.

Since payments were based on the word-of-the-day in the future, subjects did not receive event-based payments immediately. They received their $5 show-up fee on the day of the experiment, but, if randomly selected to receive a bonus payment, then they received their bonus payment on the pre-specified date that was 3 days in the future.[17]

III.D.1. The Payment Algorithms for the Non-Forced Treatment

If a participant were paid for their choices in the Non-Forced Treatment, then again we follow the general procedure outlined in Section I. Given that this environment differs from that of Experiment 1, the actual "algorithms" that we use differ as well. We use two specific algorithms between-subject, and we include their detailed descriptions in Appendix Section C. For both algorithms, following Krajbich et al. (2017), Kessler et al. (2019) and Danz et al. (2022), we give subjects minimal details about how the algorithm works.[18]

[17] While this introduces a role for time preferences, there is no natural way for this to interact with the elicitation of incompleteness.

[18] Subjects can click a button to learn more detailed information about the algorithm; 27% of subjects click this button, and these subjects are ~7 percentage points more likely to report incomplete-

One of our algorithms is a standard maximum likelihood estimation procedure. We fix a single payment question which is distinct from the questions that the subjects face in the experiment. We take all of the questions in which a subject reports strict preference or indifference and use these to estimate the risk aversion parameter in a CRRA utility function. The questions in which a subject reports incompleteness do not enter into this estimation. Then, we use this estimated utility function to predict which lottery the subject would prefer in our fixed payment question, and pay them based on this prediction, very similar to Krajbich et al. (2017).

Under this algorithm, reporting a strict preference or indifference can help the maximum likelihood estimation procedure form a more accurate estimate of the subject's risk parameter. When subjects are not sure of their preference in a given comparison, reporting incompleteness prevents the question from potentially biasing the estimation. However, this type of procedure forces complete preferences in the payment questions. Additionally, it relies on assuming a specific CRRA form of utility, and assumes the mapping from choices in the experiment to a choice in the payment question in light of this class of utility functions.

Our second algorithm is based on a non-parametric construction of better-than and worse-than sets compared to the reference lottery. These sets start with randomly-generated lotteries which are replaced via dominance whenever subjects report a strict preference or indifference. For example, if a subject reports preferring a lottery p over the reference lottery, then we replace one of the lotteries in the better-than set by p' which dominates p. We pay subjects based on a randomly-selected lottery from the better-than or worse-than sets.

This procedure does not force complete preferences in an ex-ante specified payment question, since the question we ultimately pay subjects is one where we are "sure," under dominance, that they have complete preferences. It also does not rely on any parametric assumptions on preferences. However, one major drawback of this type of procedure is that the questions subjects answer influence the possible lotteries they could be paid, since the lotteries that are replaced into the better-than and worse-than sets are a function (via dominance) of the questions in which a subject reports strict preference or indifference.

In the end, both these procedures have pros and cons. We ran both as a measure of robustness—in addition to the first mechanism used in Experiment 1—and we find

ness.

that subjects' responses do not depend on the elicitation algorithm. Because of this, our analysis pools both algorithms together; see Appendix Section C for details on the comparison between the two. Given this insensitivity to the algorithms despite their different incentives, we do not believe the details of the estimation affect how subjects respond. That said, our contribution is not to advocate for a specific preference estimation algorithm to be used to elicit incompleteness, but instead to highlight the general idea of using choices to estimate preferences, while acknowledging the limitations of the exercise.

IV. Results

Before outlining our main results, we demonstrate the data that we collect using a few representative subjects. Figure III shows the data from the Non-Forced treatment for three subjects; for each subject, we show the two reference lotteries separately. The circles represent lotteries that are preferred to the reference lottery, while the reference lottery is preferred to the comparison lotteries that are marked with squares. Diamonds mark indifference, and triangles mark incompleteness.

Subject 100—shown in the top two panels—is an example of a perfectly risk-neutral decision-maker who has complete and consistent preferences. They reported a certain belief of $pr(verb) = 0.50$, and we plot the risk-neutral indifference curve implied by this belief on each graph. Subject 100 reports indifference for the lotteries that lie on these indifference curves, and reports strict preference for all other lotteries. Lotteries that lie above the indifference curve are preferred to the reference lottery, while the reference lottery is preferred to lotteries lying below the indifference curve.

In contrast, Subject 359—shown in the middle two panels—has consistent preferences, but they are incomplete. The lotteries that are strictly preferred to the reference lottery (circles) all lie to the northeast of the lotteries that are strictly dispreferred (squares), emphasizing the consistency of the strict portion of their preferences. However, many lotteries are incomparable to the reference lottery. Subject 359 is not sure of their belief and reported a belief range of $pr(verb) \in [0.4, 0.8]$. Note that their incompleteness generally falls within this belief range, suggesting that their incompleteness could result from imprecise beliefs, or imprecise tastes, or both.

Subject 92—shown in the bottom two panels—is one of our most incomplete sub-

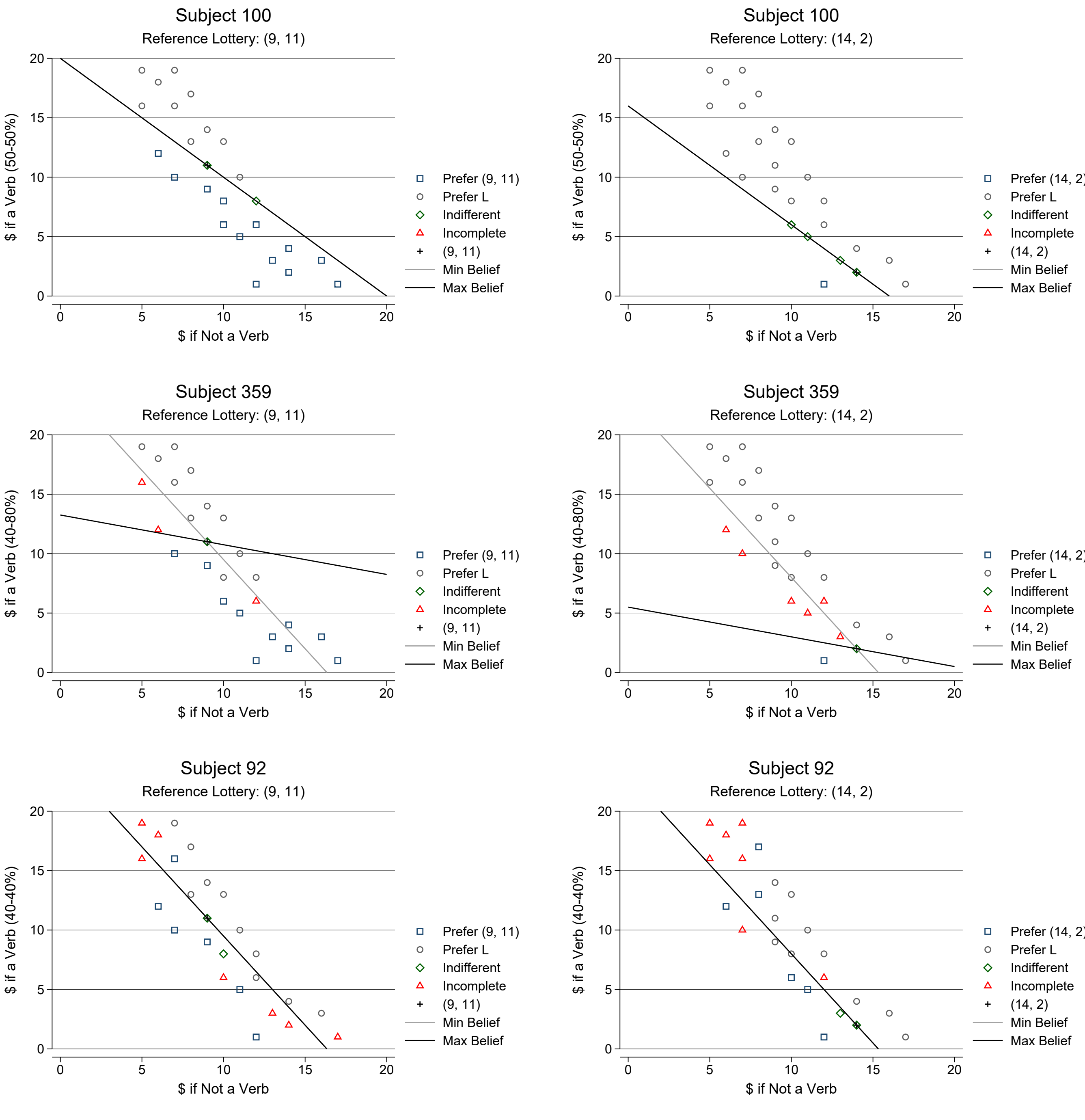

Figure III: Three Example Subjects in the Non-Forced Treatment

Note: Graphs show the full choice data for three example subjects in the Non-Forced treatment, separated by reference lottery. When subjects reported a range of beliefs, we plot the risk neutral indifference curve implied by the minimum of the range as "min belief" and the maximum of the range as "max belief." When subjects reported degenerate beliefs, this belief is reflected in the plot.

jects, with 14 total incomparabilities. However, Subject 92 reported a sure belief of $pr(verb) = 0.40$, so their incompleteness must be due to imprecise tastes under our interpretation.

IV.A. The Prevalence of Incompleteness

Table I reports the aggregate choice data for our two reference lotteries in the Non-Forced treatment. $(9,11)$ is strictly preferred to 55% of our comparison lotteries, while $(14,2)$ is strictly preferred to only 15%. This ensures that our choices sufficiently cover the space of preferences. For both of our reference lotteries, subjects report being indifferent in about 5% of comparisons, while incompleteness is the least common at 2–3%. Recall that we had subjects compare the reference lottery to itself; we exclude those comparisons in Table I. As we report below, almost all subjects report indifference in these cases.

Reference Lottery	Prefer Reference	Prefer Comparison	Indifferent	Incomplete
(9, 11)	55.3%	38.1%	4.5%	2.1%
(14, 2)	15.4%	76.8%	4.5%	3.3%

Table I: Aggregate Choice Data
Note: Subjects made 25 comparisons for each reference lottery. The table presents the percentage of subjects who preferred the reference lottery, preferred the comparison lottery, were indifferent between the two, and were unable to compare the two, aggregated across subjects. Excluded here are the comparisons in which we asked subjects to compare the reference lottery to itself.

These averages mask substantial heterogeneity, as evidenced by Figure III. 39% of subjects (N=252) report some amount of incompleteness, while the remaining 61% have fully complete preferences in the comparisons that we presented. Figure IV shows a histogram of the number of directly revealed-incomplete comparisons by subject. As the histogram shows, most subjects who report incompleteness do so in relatively few comparisons. Among those with any incompleteness, the average number of incomplete comparisons is 3.3 out of 50 total questions, and the participant with the most incomplete preferences reported 17 out of 50 incomparabilities. We note, however, that the prevalence of incompleteness is a function of the exact questions we asked, so the absolute levels are difficult to interpret in isolation.

Consistency A natural question is whether individuals report incompleteness in a way that is consistent with theoretical conceptualizations of incomparability. To as-

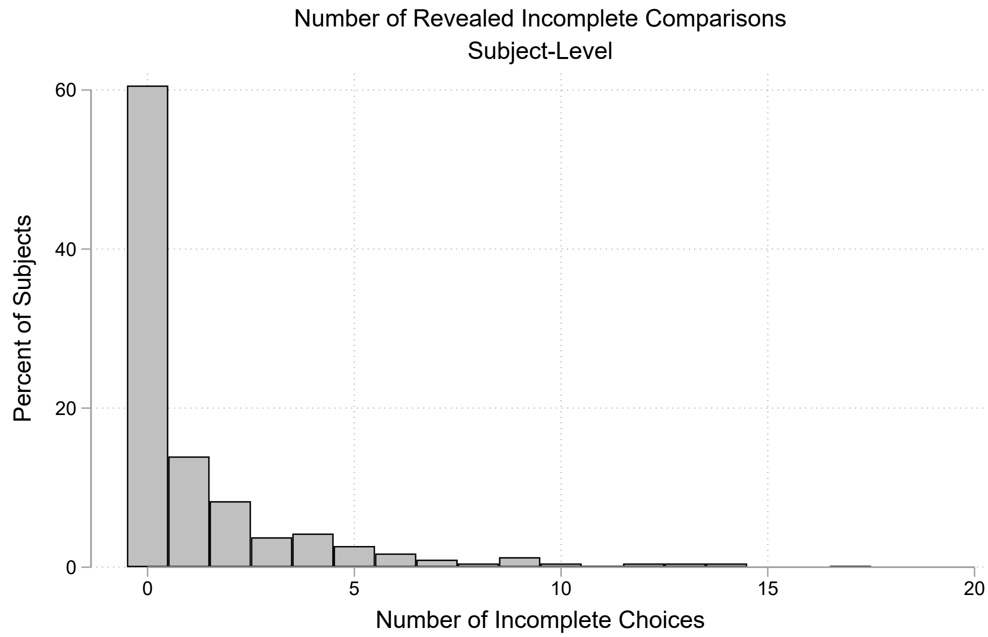

Figure IV: Subject-Level Incompleteness
Note: The figure reports the distribution of the number of incomplete comparisons each subject indicated across the two reference lotteries (out of 50 total choices). 61% of subjects had complete preferences with zero incomplete comparisons.

sess this, we consider a prediction that encompasses both models of imprecise tastes and models of imprecise beliefs. These models predict that if a lottery, p, is incomparable to the reference lottery, r, then r cannot be strictly preferred to any lottery that dominates p in the sense of paying more in every state. This is because, if r is strictly preferred to the dominating lottery, then it should also be strictly preferred—rather than incomparable—to p.[19] Similarly, any lottery dominated by p cannot be strictly preferred to r. Again, this is because, if the dominated lottery were strictly preferred to r, then p should also be strictly preferred rather than incomparable to r. Essentially, this prediction says that the set of lotteries strictly preferred to the reference lottery must lie above the set of lotteries that are strictly worse than the reference lottery, and incompleteness must lie in between these two sets. See Figure IX in the Appendix for an illustration. Furthermore, the same prediction should hold for indifference.

[19] This is easy to see formally by observing that if p and r are not comparable $E_\pi[u(p)] \geq E_\pi[u(r)]$ for some $u \in \mathscr{U}$ and $\pi \in \Pi$, and any lottery q that pays more in each state that p does will have the property that $E_\pi[u(q)] > E_\pi[u(p)]$ and therefore $E_\pi[u(q)] > E_\pi[u(r)]$.

We test this prediction by taking every reported incomparability, identifying the lotteries that strictly dominate or are dominated by this lottery, and calculating the percentage of strict preferences for these lotteries that are in the direction predicted, as described above. We find a very high degree of adherence to theory as summarized in Table II: Upwards of ~80% of strict preferences are consistent with theoretical formulations of incompleteness. We find very similar rates of adherence to the theory for our indifferent comparisons, and we discuss the relationship between incompleteness and indifference more below.

	Incomplete	Indifferent
Dominating lotteries preferred to r	87%	89%
r preferred to dominated lotteries	78%	79%

Table II: Consistency with Implications of Theoretical Models
Note: We calculate this across all incomplete or indifferent comparisons. Percentages report the percentages of strict preferences among dominating/dominated lotteries that are in the direction predicted by theory. The percentages reported consider weak dominance, and the results are even stronger when we consider strict dominance: 87% and 82% for incompleteness, and 90% and 81% for indifference.

Survey Evidence Finally, to aid our interpretation of subjects' understanding of incompleteness, at the end of our study, we asked subjects who ever reported incompleteness why they did so. We gave subjects a few answer options and asked them to select all that applied.[20] 75% indicated that they "did not know which gamble (they) preferred," and 17% indicated that they "did not know how to compare the gambles." We also asked subjects who never reported incompleteness to state why they did not use this answer option.[21] 91% said that they "always preferred one gamble over the other." Less than 1% of subjects stated that they "did not trust the algorithm if (they) said (they) didn't know," and 1% said that they "didn't know what would happen if (they) said (they) didn't know." Thus, subjects' self-reports generally accord with our interpretation of (in)completness.

[20] The full set of options consisted of: "I knew which one I liked better, but I didn't like either gamble," "I didn't know which gamble I preferred," "I did not know how to compare the gambles," "I didn't want to take time to figure out which one I preferred", "I knew which one I liked better, but I thought saying 'I don't know' would give me a better gamble to be paid at the end," and an option to report "Other" and fill out a text box.

[21] The full set of options consisted of: "I always preferred one over the other," "I didn't trust the algorithm," "I didn't know what would happen if I said I didn't know," "I thought it would give me a better gamble to be paid at the end," and an option to report "Other" and fill out a text box.

Result 1. *Over one third (≈ 40%) of subjects directly reveal incompleteness in their preferences. Reported incompleteness behaves systematically and appears in the regions broadly predicted by theoretical formulations of incomplete preferences.*

IV.B. Indifference vs. Incompleteness

As evident in Table I, subjects are more likely to report indifference than incompleteness. Theoretically, indifference is "knife-edge" and should be less prevalent than incompleteness. This brings to question how subjects perceived indifference and incompleteness and how to interpret these choice responses. We present three pieces of evidence suggesting that subjects do differentiate between indifference and incompleteness in a way that aligns with our interpretations.

Mechanical Indifference and Incompleteness As a "sanity check," we included each reference lottery as a comparison lottery with itself. That is, subjects were asked to compare $(9,11)$ with itself, and were asked to compare $(14,2)$ with itself. 93% of subjects reported indifference in each of these comparisons, and less than 1% report incompleteness.[22] This gives us reassurance that subjects report indifference when they "should" report indifference, and that subjects properly distinguish between indifference and incompleteness.

It is harder to test whether subjects report incompleteness when they should since we cannot as easily induce incompleteness like we can indifference. That said, we attempt to demonstrate this at least directionally with an additional treatment. We recruit 200 new participants through Prolific. For these subjects, we withhold payoff-relevant information in some comparisons, and argue that subjects should be more likely to report incompleteness when they do not have full information.

In particular, we asked subjects about the following comparisons:[23]

1. $(14,x)$ vs. $(14,x)$
2. $(14,x)$ vs. $(8,x)$
3. $(14,x)$ vs. $(14,y)$
4. $(14,5-3x+y+(1\times-2))$ vs. $(7+(1-x)+2\times y-(6+4),19)$

[22]As noted above, these comparisons are excluded from Table I, so it does not explain the higher prevalence of indifference than incompleteness.

[23]In implementation, we set $x=2$ and $y=5$, since (14, 2) and (5, 19) were lotteries we included in our original treatments.

5. $(5, 6+5x-2(y+1))$ vs. $(5, 8-y+3x-(2\times 3))$

We tell subjects that x and y represent some possible payment amount between \$0 and \$20, but we do not tell them the exact amount.[24] If subjects understand and trust our algorithm, then they would be indifferent between $(14,x)$ and $(14,x)$, despite uncertainty about x. Similarly, they would have a strict preference between $(14,x)$ and $(8,x)$. However, they might not be able to compare $(14,x)$ and $(14,y)$, given the uncertainty about x and y. It is possible that subjects still form beliefs about x and y, enabling them to form a strict preference, or that they are indifferent between them. Nevertheless, we predict that subjects would be more likely to report incompleteness for this comparison than for our comparisons with full information. Finally, we attempt to exaggerate incompleteness by including the last two options that are deliberately complex. We predict that we would see the most incompleteness for these comparisons.

	Strict Preference	Indifferent	Incomplete
$(\$14,\$x)$ vs. $(\$14,\$x)$	5%	93%	3%
$(\$14,\$x)$ vs. $(\$8,\$x)$	94%	4%	3%
$(\$14,\$x)$ vs. $(\$14,\$y)$	13%	51%	36%
Complex_1	46%	6%	48%
Complex_2	39%	11%	51%

Table III: Aggregate Choice Data

Notes: Complex_1 refers to $(14, 5-3x+y+(1\times -2))$ vs. $(7+(1-x)+2\times y-(6+4), 19)$ and Complex_2 refers to $(5, 6+5x-2(y+1))$ vs. $(5, 8-y+3x-(2\times 3))$.

Results, shown in Table III, confirm our hypotheses. 93% of individuals are indifferent between $(14,x)$ and $(14,x)$ and 94% report strict preference between $(14,x)$ and $(8,x)$. We take this as reassurance that the act of withholding information itself does not lead to incompleteness, and as further validation of reported indifference. However, individuals are much more likely to report incompleteness in comparisons where the lack of information makes it difficult to form a preference: 36% report incompleteness between $(14,x)$ and $(14,y)$, and half of subjects report incompleteness in our complex comparisons. While we still cannot say whether the remaining half of

[24] In implementing the experiment, we used % and # symbols rather than x and y. We tell subjects that, when the same symbol appears in both options, it represents the same amount of money. When two different symbols appear, they represent different amounts of money.

subjects "should" report incompleteness here, results demonstrate that, at least directionally, individuals are willing to report incompleteness when they do not know how to compare two alternatives. Indeed, 76% of individuals report incompleteness at least once in this treatment, compared to 39% in our original data (Fisher's exact, $p < 0.001$).

Complexity Enke and Shubatt (2024) recently developed a rich index that captures the features of a menu that make lottery choice more complex, both in an objective sense—namely failure to choose the lottery with the higher expected value—and in a subjective sense—namely self-reported cognitive uncertainty (Enke and Graeber, 2020). They refer to these measures as "objective problem complexity" (OPC) and "subjective problem complexity" (SPC), respectively. One might expect that individuals would be more likely to report incompleteness for lottery choices that are more complex, while it is not clear how complexity and indifference are related.

To test this hypothesis, we calculate OPC and SPC for all of the lottery choices that each participant faces, using that participant's self-reported belief.[25] Then, for a given OPC or SPC level, we find the average likelihood of reporting incompleteness and indifference. Both OPC and SPC are positively and significantly correlated with the percentage of reported incompleteness (OPC: 0.0601 $p < 0.001$, SPC: 0.0559 $p < 0.001$), but these measures are actually negatively correlated with the percentage of reported indifference (OPC: -0.0225 $p < 0.0384$, SPC: -0.0207 $p = 0.0565$).[26] Thus, complexity is a strong predictor of incompleteness, but is not a predictor of indifference. This further establishes that indifference and incompleteness are distinct to participants, and our methodology allows us to isolate these in a clear way.

Response Times Finally, we analyze response times conditional on choice. In each comparison, we record the time it takes a subject to submit their decision. Subjects take 8.84 seconds to submit a strict preference, on average. This is directionally though not significantly faster than indifference (9.73 seconds, $p = 0.147$) and incom-

[25] This is an adaptation of the Enke and Shubatt (2024) exercise, as their environment contained only objective risky lotteries. We conduct this exercise for our objective lottery treatment, discussed below, and find similar results.

[26] This correlation is even stronger in our treatment with objective lotteries that we discuss below. This is not surprising since we have more power in this treatment given that all subjects faced the same probabilities, so we essentially have more observations per unique lottery.

pleteness (10.49 seconds, $p = 0.076$).[27] Thus, it appears that incompleteness is the slowest response type in our data, and is directionally slower than indifference.

Result 2. *We can separate indifference from incompleteness using our methodology, and the two concepts appear distinct to individuals. Incompleteness is related to existing notions of complexity while indifference is not, and incompleteness and is associated with directionally longer response times than indifference.*

IV.C. The Source of Incompleteness

Given that we observe incompleteness in preferences over uncertainty, a natural question relates to the source of incompleteness. One theory is that incompleteness reflects imprecision in beliefs, while an alternative hypothesis is that incompleteness reflects imprecision in tastes. It is also possible that both beliefs and tastes are imprecise and both contribute to incompleteness. We look to see whether subjects with incomplete preferences are more likely to have imprecise beliefs, lending support for the first hypothesis, or instead whether incompleteness is as prevalent in individuals with precise beliefs, indicating a source of imprecise tastes.

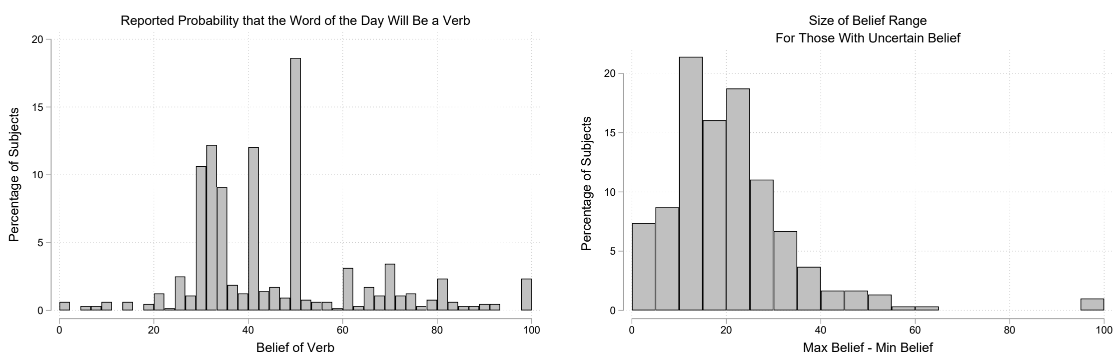

Figure V: Distribution of Reported Beliefs of the Likelihood that the word-of-the-day will be a Verb

Note: The left panel presents the distribution of reported beliefs that the word-of-the-day would be a verb. The right panel presents the distribution of the size of belief ranges for the 47% of subjects who reported having an uncertain belief.

[27] We exclude exact comparisons in this calculation, since the response time literature typically does not ask subjects to choose between two identical objects. $p-$values are reported from a linear regression with standard errors clustered at the subject level.

The left panel of Figure V shows the distribution of reported beliefs that the word-of-the-day will be a verb. Most subjects are clustered around a belief of one-third, with another group clustered around one-half. 47% of subjects indicate uncertainty about their reported belief.[28] Subjects who report uncertainty about their belief have the opportunity to report a range of beliefs, and on average, they report ranges that span 19 percentage points. The full distribution of the size of the belief ranges for those with uncertain beliefs can be found in the right panel of Figure V.

If incompleteness stems mainly from uncertainty in beliefs, then we would expect the subjects who report incompleteness to be the subjects who indicate uncertainty about their beliefs. Table IV shows the joint distribution between incompleteness and belief uncertainty. Of those with incomplete preferences, only half have uncertain beliefs. Among subjects with belief uncertainty, 43% report incomplete preferences, directionally but only marginally significantly higher than the 36% reporting incompleteness among those with certain beliefs (Fisher's exact, $p = 0.075$). Furthermore, among the subsample of subjects with imprecise beliefs, we find no significant correlation between tendency to report incompleteness and the size of the belief range (correlation: 0.0447, $p = 0.441$).

Thus, despite evidence that many subjects have imprecise beliefs—and it might be the case that this belief imprecision contributes to incompleteness to some extent—it does not appear to be the case that belief imprecision is the main contributor to incompleteness in preferences in our experiment. Note that, because we ask a finite number of questions in a particular portion of the lottery space, it's possible, and indeed expected, that individuals with uncertain beliefs would have revealed incompleteness in other questions even if they did not report incompleteness in our study. We can only detect incompleteness as a function of our specific questions, and we would expect belief imprecision to manifest as incompleteness at least somewhere in lottery space.

Objective Lotteries To confirm the previous finding, we run two additional treatments, each with about 150 new subjects recruited through Prolific. These treatments are exactly the same as our original experiment, except that we provide subjects with an objective probability. Specifically, in one treatment, we tell subjects that

[28]This is on par with the 49% of subjects who report uncertainty about developing late-onset dementia—a dramatically different event—in Giustinelli et al. (2019).

	Preferences	
	Complete	Incomplete
Certain Belief (53% of subjects)	34.0%	19.2%
Uncertain Belief (47% of subjects)	26.6%	20.2%

Table IV: Relationship Between Incompleteness and Imprecise Beliefs
Note: The table presents the aggregate percentage of subjects broken down by whether they have certain or uncertain beliefs and whether they have complete or incomplete preferences. Reported percentages are unconditional.

there is a $\frac{1}{3}$ chance of the event realizing—analogous to a $\frac{1}{3}$ chance that the word-of-the-day would be a verb—and in the other, we tell subjects that there is a $\frac{1}{2}$ chance of the event realizing.[29] These objective probabilities reflect the modal beliefs from our subjective version of the experiment, allowing us to make comparisons about the likelihood of incompleteness across objective and subjective uncertainty.[30]

At an individual level, we find no significant reduction in the tendency to report incompleteness when lotteries are defined over events with objective probabilities. Recall that, in our original data, 39% of subjects reported incompleteness at least once. We find a similar result in our new objective treatments: 37% of subjects report incompleteness at least once when probabilities are objectively given (Fisher's exact, $p = 0.615$).

To look at incompleteness on the level of an individual question, we compare individuals with beliefs near one-third or one-half in the subjective treatment who are *uncertain* about this belief to individuals with the same beliefs in the objective treatment who are *certain* about their belief. The number of subjects who fit these criteria are referenced in Table V. For this analysis, we call reported beliefs in $[30, 35]$ "near one-third" and beliefs in $[48, 52]$ "near one-half" in order to allow for small deviations in reporting.

[29]Since we want to truthfully tell subjects the objective probabilities while keeping everything as similar as possible to our subjective treatments, we show subjects pictures of cards that have "verb" or "not a verb" written on them. For example, in the treatment where we induce an objective prior of $\frac{1}{3}$, we show subjects three cards, one of which says "verb" and the other two say "not a verb." We tell subjects that we will randomly select one of these cards and the label of the card will determine their payment from a given lottery. We tell subjects that they can watch us live-streaming the card draw on Twitch in three days to further instill trust in the randomization process. We did live-stream the card draws, but did not have any viewers.

[30]We still elicit subjects' beliefs at the end of the experiment and we ask them whether they are certain about this belief. Aggregating across all subjects in both objective treatments, only 11% indicated that they were unsure about their reported belief. This is significantly lower than the 47% who indicated belief uncertainty in the subjective treatment (Fisher's exact, $p < 0.001$).

		Strict Preference	Indifferent	Incomplete
Near One-Third	Uncertain (N=92)	93.3%	3.8%	2.9%
	Certain (N=117)	93.5%	3.4%	3.1%
Near One-Half	Uncertain (N=50)	91.9%	5.1%	3.0%
	Certain (N=126)	94.5%	3.5%	2.0%

Table V: Aggregate Choice Data

Note: The table presents the percentage of subjects who had a strict preference, were indifferent, and were unable to compare the gambles, aggregated across subjects and reference lotteries. Excluded here are the comparisons in which we asked subjects to compare the reference lottery to itself. Uncertain $\frac{1}{3}$ are subjects who reported beliefs in $[30,35]$ in the subjective treatment and reported that they were not sure of this belief; Certain $\frac{1}{3}$ are subjects who reported beliefs in $[30,35]$ in the objective treatment and reported that they were sure of this belief. The $\frac{1}{2}$ are defined analogously in the range $[48,52]$. Sample sizes, N, are reported as number of subjects who satisfy the given belief restrictions.

Table V presents the aggregate choice data. For beliefs near one-third, we find extremely similar rates of incompleteness. For beliefs near one-half, we find slightly more incompleteness reported by individuals with uncertain beliefs in our subjective treatment compared to individuals with certain beliefs in our objective treatment (rank-sum $p = 0.485$ for beliefs near one-third, $p = 0.0066$ for beliefs near one-half). Nevertheless, most of the incompleteness remains even with objective probabilities that individuals are certain about. Thus, we conclude that belief uncertainty is not the main contributor to observed incompleteness in our study.

Result 3. *Imprecise beliefs cannot fully explain incompleteness. Of those with incomplete preferences, only half report being uncertain of their subjective belief. Furthermore, individuals are equally likely to express incompleteness in an objective environment where they are certain of their beliefs.*

IV.D. The Impact of Forced Choice

Given that we can identify incompleteness and indifference and that these appear to be non-trivial components of preference, we turn finally to understand the consequences of not allowing individuals to express incompleteness and indifference. To do so, we compare choices in the Non-Forced treatment to those in the Forced treatment. One interpretation of the distinction between these two treatments follows Mandler (2005) and Nishimura and Ok (2016). They model two preference relations: *Choices* are complete by construction, but might be intransitive, while *tastes* can be incomplete, but the complete portion of the relation is transitive. Under this interpretation, our Non-Forced treatment measures tastes and the Forced treatment measures choices. We analyze the extent to which intransitive choices can be explained by imprecise tastes.

Our experimental design allows us to test transitivity as follows. In both the Forced and Non-Forced blocks, individuals compare lotteries, p, to our two reference lotteries, r_1 and r_2. As one of these comparisons in each block, individuals also compare r_1 and r_2 directly. An individual's ranking between r_1 and r_2 allows us to "link" all of the choices together by transitivity. For example, an individual who prefers p to r_1, and who prefers r_1 to r_2, should also prefer p to r_2. As such, a violation of transitivity is observed whenever $p \succeq r_i \succeq r_j \succeq p$.[31]

In our Forced choice treatment, individuals are only given the option to report $p \succeq r_i$ or $r_i \succeq p$, so the expression above presents the only possible transitivity violations. In our Non-Forced treatment, individuals can report indifference explicitly, which presents additional transitivity violations. For example, $p \sim r_i \succ r_j \succ p$ represents a Non-Forced transitivity violation, as well. We analyze Non-Forced violations first only in strict preference and then include indifference.

In forced choice, we find 4.0% of choices constitute a transitivity violation. This is significantly higher than the 1.7% of strict preferences in the Non-Forced treatment that constitute transitivity violations ($p < 0.0001$). This gap shrinks but is still significant when we included Non-Forced intransitivities that involve indifference (3.3%, $p = 0.0184$). This is stark, since individuals have four answer options in the Non-Forced treatment so trembles would be more likely to lead to intransitivities. Of the transitivity violations observed in Forced choice, 29% involve a comparison identified as indifferent or incomplete in the Non-Forced treatment. This is significantly higher than the percentage of transitive comparisons that include an indifferent or incomplete comparison (22% vs. 29%, Fisher's exact $p < 0.001$). Thus, a significant proportion of intransitivities in Forced choice can be explained by incompleteness or indifference, though the majority of intransitivities cannot be explained in this way. This leaves room for alternative explanations for intransitivities that are consistent with the data, such as random utility (Block and Marschak, 1959; He and Natenzon,

[31]We had subjects compare the two reference lotteries to each other twice, once in each block. It is possible that a subject reports a different preference across these two choices. We consider it a violation of transitivity if either report constitutes a violation.

2024).

Result 4. *Forced choice leads to more inconsistencies in preferences compared to Non-Forced choice. Incompleteness can explain a non-trivial portion of intransitivities in forced choice.*

IV.E. Indirectly Revealed Incompleteness

We make use of the Forced treatment for one additional comparison. Despite the high degree of consistency in strict preference between the Forced and Non-Forced treatments, we find that subjects exhibit preference reversals in 14% of choices in which they revealed a strict preference. Although this could be evidence of stochastic preferences or "learning" one's preferences throughout the experiment, another interpretation is that these preference reversals reflect underlying "indirectly revealed" incompleteness that subjects themselves are not aware of (Bayrak and Hey, 2017). For each of our comparison lotteries, we calculate the percentage of subjects who report incompleteness when comparing this lottery against one of the reference lotteries in the Non-Forced treatment, and we correlate this with the percentage of subjects who report a strict preference in the Non-Forced treatment but then choose the other lottery in the Forced treatment.

We find a strong positive correlation between these two percentages: The comparisons in which subjects are more likely to report incompleteness are the same comparisons in which other subjects are likely to exhibit preference reversals (correlation of 0.83 for reference lottery (9,11) and 0.80 for reference lottery (14,2)); see Figure X in the Appendix for visualization. This provides suggestive evidence in favor of the interpretation that preference reversals can be understood as indirectly revealed incompleteness—when we include both directly and indirectly revealed incompleteness, 98% of subjects have incomplete preferences. Furthermore, this provides evidence that there may be comparison-specific features of decisions that make it more difficult to form preferences, in line with the notion of complexity from Enke and Shubatt (2024).

Finally, we note that almost all subjects (94%) who directly reveal incompleteness in the Non-Forced treatment also indirectly reveal incompleteness by exhibiting a preference reversal. Thus, we do not interpret these preference reversals to mean that subjects were distrustful of our experiment and were unwilling to reveal incompleteness directly, since these preference reversals occur at similar rates in subjects who we know were willing to reveal incompleteness directly. Instead, this suggests that even individuals who are sometimes aware of their incompleteness have underlying incompleteness that they are not aware of.

Result 5. *The percentage of subjects who report incompleteness in a given question is highly correlated with the percentage of subjects who exhibit a preference reversal in this question.*

IV.F. Did Subjects Report Truthfully?

Naturally, one might worry about how subjects perceived our elicitation and whether they reported truthfully. Our design allows for many ways in which to evaluate subjects' responses and detect evidence of manipulation. We collect the evidence here, and believe that it paints a picture of truthful reporting.

First, we look for evidence to evaluate subjects' reports of strict preferences. There are two features of the data that suggest subjects reported strict preferences truthfully. First, we find that subjects who report a strict preference when incentivized by our mechanism report the same choice 86% of the time under standard incentives in the Forced treatment, which is a high degree of consistency (and similar to the ≥85% consistency in strict preferences observed in Experiment 1). As a second piece of evidence, we included a few lottery comparisons that were related by strict dominance (i.e., one lottery paid strictly more in both states). In these questions, subjects report to our mechanism a strict preference for the dominant lottery in 89% of instances. Thus, it appears that subjects were truthfully reporting their strict preferences when incentivized by our mechanism.

Second, we look for evidence to evaluate subjects' reports of indifference. Our clearest validation of these reports is when we ask subjects to compare a lottery to itself. Here, over 90% of subjects report indifference. We also find high rates of reported indifference (93% of subjects) even when we withhold the exact payment values, as reported in Section IV.B. Furthermore, in Experiment 1, we find that the row in which individuals were most likely to report indifference was the row in which the risky urn contained 50% winning balls. If individuals were to apply the principle of insufficient reason and act as if the ambiguous urn contained 50% winning balls and 50% losing balls, then this is exactly where we would expect them to be indifferent. Thus, we find compelling evidence that subjects do not attempt to manipulate our mechanism—by reporting a strict preference or incompleteness—when they are actually indifferent.

Given that we have evidence that subjects report strict preferences and indifferences truthfully, we find it natural to conclude that subjects also report incompleteness truthfully. This is the hardest to validate in the data, precisely for the fact that incompleteness is difficult to identify. Nevertheless, there are four features of the data to support this claim. First, as reported in Section IV.A, we find that ~80% of reported incomparabilities are consistent with theoretical conceptualizations of incompleteness. Second, we find that over twice as many subjects report incompleteness (from 39% of subjects to 76% of subjects) in our treatment that deliberately increases complexity and removes information, which was designed to increase rates of incomparability. Third, at a question level, we see a strong positive correlation between choice inconsistencies (i.e., preference reversals) and incompleteness that is revealed directly through our mechanism. If one believes preference reversals and stochastic choice to be an indicator of underlying incompleteness, then our mechanism measures behavior that is highly correlated with this. Finally, we note that survey responses confirm our interpretation, as well: 91% of subjects who do not report incompleteness say that it is because they always preferred one over the other, and 84% of subjects who report incompleteness say that it's because they did not know which gamble they preferred or did not know how to make the comparison between the two gambles.

Taken together, we believe there is compelling evidence that our experimental participants reported their preferences thoughtfully and truthfully. This suggests that estimating preferences and paying based on estimated preferences is behaviorally incentive compatible (Danz et al., 2022). We chose to start with simple implementations as a proof-of-concept, but, as discussed in Appendix C, there are many ways in which one could change the precise details of the estimation to address specific potential sources of manipulation.

V. Related Literature

Our paper is most closely related to other papers that have attempted to identify incompleteness in preferences (see Bayrak and Hey, 2020 for a survey of the literature and methodologies related to preference imprecision). Cohen et al. (1985) and Cohen et al. (1987) are the first experimental papers we are aware of in which subjects are presented with an "indifference" and an "I do not know" option. In their case, subjects compare a certainty equivalent to a binary risky option with given objective probabilities using a price list (they also have an unknown probabilities option). They find that about 10% of subjects display indecision. In their case, reporting "I do not know" implied the experimenter chose for the subject.

Cubitt et al. (2015) have subjects fill out a standard multiple price list by reporting a switch point, but also allow subjects to report that they are "not sure about (their) preference" in any rows. However, in order to incentivize these decisions, subjects were also required to complete their incomplete preferences by indicating a single switch point even if they were unsure. Across all of their questions, they find that 87% of subjects report some preference imprecision by using this "unsure" option.

This is in line with Agranov and Ortoleva (2020) who ask subjects to fill out a multiple price list but allow for the option to explicitly randomize in each row. For subjects who choose to randomize across multiple consecutive rows, one interpretation is that preferences are imprecise in this range. They find that between 50–75% of subjects choose to randomize in a way that is consistent with preference imprecision. This is also consistent with their earlier work in which 70% of subjects randomize across multiple repetitions of the same decision (Agranov and Ortoleva, 2017).

While one interpretation of randomization is that subjects have imprecise preferences, individuals might randomize instead if they are indifferent, have convex preferences, utility from gambling, misspecified beliefs about the uncertainty generating process, or other potential explanations (Agranov et al., 2021). Thus, the literature has attempted to identify evidence of incompleteness in other ways. In a paper related to ours, Costa-Gomes et al. (2021) use costly deferral as indication of incompleteness. Subjects are presented with choices over consumption goods (in their case, headsets). In one treatment, subjects were forced to make a choice of headset, while in the other they could pay a small cost to defer choice to the end of the experiment. They find that 35% of subjects use the deferral option. They also find that choices are more coherent in the deferral treatment than in the forced choice treatment, similar to our results comparing the Forced and Non-Forced treatments.

These experiments used very different elicitation methodologies to ours, but our 40–50% of subjects expressing incompleteness is in line with other estimates in the literature. Randomization seems to be more prevalent than deferral or explicit statements of incompleteness. This could be because randomization captures additional preferences as discussed above, or it could be that deferral and explicit incompleteness require more "sophistication" and awareness of incompleteness. This is consistent with the conclusion from Cettolin and Riedl (2019), which is that some amount of randomization can be attributed to incompleteness while for other subjects it is a manifestation of their inherent preference for randomization.

Our paper builds on this previous literature and contributes to the understanding of incomplete preferences in a few ways. First, our elicitation procedure reveals when individuals "do not know" how to make a choice or are unsure of their preference. This methodology is designed to capture welfare-relevant incompleteness, where individuals do not want the analyst to make inference based on these decisions. We view this as complementary to other elicitation methodologies in the literature and think it would be interesting to compare what is revealed across these different methodologies. Second, we design our experiment to conform to theoretical studies of incompleteness, and directly elicit from subjects a distinction between indifference and incompleteness. Third, we separate imprecise beliefs from imprecise tastes, and find that incompleteness results from imprecise tastes more than imprecise beliefs. Fourth, and related to the previous point, we study incompleteness in a domain of subjective uncertainty while many previous papers have focused on objective uncertainty. We use a novel but natural event structure that allows for imprecise subjective beliefs, but are still able to compare this to an equivalent environment with objective uncertainty. Finally, we confirm what Costa-Gomes et al. (2021) find—that forced choice results in less consistent decisions—and our experiment further suggests that preference reversals can indicate underlying incompleteness.

Finally, from a methodological perspective, using a preference estimation algorithm to elicit otherwise-hard-to-elicit information is very much in the spirit of Krajbich et al. (2017) and Kessler et al. (2019). Like our paper, neither of these papers discuss the actual "algorithm" used, nor do they tell subjects what it is.[32] We extend their methodologies to allow for the elicitation of indifference and incompleteness, and, in addition, we complement these papers by testing for untruthful reporting

[32]For example, Kessler et al. (2019) tell employers that they use "a newly developed machine-learning algorithm to identify candidates who would be a particularly good fit for your job based on your evaluations."

through a variety of measures summarized above, finding no evidence of it.

VI. Discussion

Using a method that allows us to elicit incompleteness directly, we find that about 40–50% of subjects have incomplete preferences over simple money lotteries. Reported incompleteness generally conforms to theoretical formalizations of preference, and contributes to standard behavioral anomalies such as intransitivities and preference reversals in forced choice. Our results suggest incompleteness stems mainly from imprecise tastes rather than imprecise beliefs. Furthermore, our methodology allows us to separate indifference and incompleteness and the two concepts appear distinct.

How should one interpret the incompleteness that subjects reveal directly, or the potential underlying incompleteness that they do not reveal? There are a few alternate interpretations of our data that one could take. First, there are some theories of choice that assume intransitivities must reflect incompleteness; since we do not find that all intransitivities can be explained through incompleteness, under these theories, there must be residual incompleteness that we cannot detect (Mandler, 2005; Nishimura and Ok, 2016). This suggests a higher prevalence of incompleteness in the population—up to 94% of our subjects. However, these preference reversals can also result from stochastic choice, which we cannot differentiate in our data.

Furthermore, it is possible that subjects have incomplete preferences in this space but develop heuristics or procedures that enable them to complete their preferences (Arrieta and Nielsen, 2024). We cannot detect this so we cannot rule out this interpretation, and indeed we believe this to be a reasonable hypothesis. Under this interpretation, the reported incompleteness we observe should be interpreted as situations where subjects are unable to find a way to complete their preferences given the heuristics and procedures they have developed. The relationship between heuristics and incompleteness is an interesting open question.

In addition, and related to our discussions above, the incompleteness we measure in this design is in some sense a *sophisticated* incompleteness, where subjects must know that they don't know their preference. For example, if we assume intransitivities and preference reversals reflect incompleteness, then we would conclude that subjects are only aware of a fraction of their underlying incompleteness. Other elicitation mechanisms in the literature potentially require less sophistication at the expense of clean interpretation as incompleteness (e.g., randomization). The extent to which individuals are sophisticated about their incompleteness is another interesting open question.

Finally, the question remains as to how one should interpret the levels of incompleteness we see, especially as it pertains to our unusual elicitation methodology. One can think about this question from two different perspectives: at the level of a given individual or at the level of a given comparison. At the individual-level, we find a very stable proportion of individuals—about 40–50%—report incompleteness across Experiments 1 and 2, and in both our subjective and objective treatments. In our deliberately-complex treatment, this increases to 76%. This suggests a lower bound of at least three-quarters of subjects trust our elicitation mechanism and are willing to report incompleteness, but many of these subjects do not have incomplete preferences in our standard simple binary choices.

At a comparison-level, it is difficult to say how the level of incompleteness would change in different environments. As noted, we find more incompleteness in complex questions. This suggests that the "magnitude" of incompleteness in individuals' preferences is underestimated by the simple questions we use. For example, one could imagine that choices involving compound lotteries, lotteries with more outcomes, or multi-dimensional objects would reveal more incompleteness. From this perspective, it is perhaps surprising that incompleteness remains when we strip away all uncertainty about probabilities and outcomes. On the other hand, there is evidence that individuals are more likely to follow and/or develop decision rules in complex environments (Nielsen and Rehbeck, 2022; Arrieta and Nielsen, 2024), so the interaction between complexity and incompleteness is an interesting open question.

Our results leave open a number of interesting questions. As we discuss in reviewing related papers in the literature, there are other methods of identifying incompleteness, and different measures can lead to different conclusions on preferences. It would be interesting to understand the compare these methodologies—and related concepts such as cognitive uncertainty (Enke and Graeber, 2020)—to understand the extent to which they measure the same uncertainty in preferences.

Finally, it would be interesting to understand better how individuals complete their incomplete preference. In our Forced treatment, individuals are asked to make a choice even if their preferences are incomplete. Understanding this completion process better could help interpret standard choice data and potentially could allow for identification of incompleteness even when individuals are unable to report incompleteness directly. Along these lines, it would be very interesting to identify any neurological or biological indicators of completeness.

References

Agranov, M., P. J. Healy, and K. Nielsen (2021): "Stable Randomization," *Working Paper*.

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

——— (2020): "Ranges of Randomization," *Working Paper*.

Arrieta, G. and K. Nielsen (2024): "Procedural Decision Making in the Face of Complexity," *Working Paper*.

Aumann, R. (1962): "Utility Theory Without the Completeness Axiom," *Econometrica*, 30, 445–462.

Baillon, A., Y. Halevy, and C. Li (2022): "Randomize at your own Risk: on the Observability of Ambiguity Aversion," *Econometrica*.

Bayrak, O. K. and J. D. Hey (2017): "Expected Utility Theory with Imprecise Probability Perception: Explaining Preference Reversals," *Applied Economics Letters*, 906–910.

——— (2020): "Understanding Preference Imprecision," *Journal of Economic Surveys*, 154–174.

Bernheim, B. D. and A. Rangel (2009): "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics," *Quarterly Journal of Economics*, 124, 51–104.

Bewley, T. F. (2002): "Knightian Decision Theory. Part I," *Decisions in Economics and Finance*, 25, 79–110.

Block, H. D. and J. Marschak (1959): "Random orderings and stochastic theories of response," *Cowles Foundation Discussion Paper*.

Cettolin, E. and A. Riedl (2019): "Revealed Preferences Under Uncertainty: Incomplete Preferences and Preferences for Randomization," *Journal of Economic Theory*, 181, 547–585.

COHEN, M., J.-Y. JAFFRAY, AND T. SAID (1985): "Individual Behavior under Risk and under Uncertainty: An Experimental Study," *Theory and Decision*, 18, 203–228.

——— (1987): "Experimental Comparison of Individual Behavior under Risk and under Uncertainty for Gains and for Losses," *Organizational Behavior and Human Decision Processes*, 39, 1–22.

COSTA-GOMES, M., C. CUEVA, G. GERASIMOU, AND M. TEJIŠČÁK (2021): "Choice, Deferral and Consistency," *Quantitative Economics*, forthcoming.

CUBITT, R. P., D. NAVARRO-MARTINEZ, AND C. STARMER (2015): "On Preference Imprecision," *Journal of Risk and Uncertainty*, 50, 1–34.

DANZ, D., L. VESTERLUND, AND A. J. WILSON (2022): "Belief Elicitation and Behavioral Incentive Compatibility," *American Economic Review, Forthcoming*, 112, 28511–83.

DUBRA, J., F. MACCHERONI, AND E. A. OK (2004): "Expected Utility Theory without the Completeness Axiom," *Journal of Economic Theory*, 115, 118–33.

ELIAZ, K. AND E. A. OK (2006): "Indifference or Indecisiveness? Choice-Theoretic Foundations of Incomplete Preferences," *Games and Economic Behavior*, 56, 61–86.

ENKE, B. AND T. GRAEBER (2020): "Cognitive Uncertainty," *Working Paper*.

ENKE, B. AND C. SHUBATT (2024): "Quantifying Lottery Choice Complexity," *Working Paper*.

FOX, C. R. AND A. TVERSKY (1995): "Ambiguity Aversion and Comparative Ignorance," *The Quarterly Journal of Economics*, 110, 585–603.

GALAABAATAR, T. AND E. KARNI (2013): "Subjective Expected Utility with Incomplete Preferences," *Econometrica*, 81, 255–284.

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. AND D. SCHMEIDLER (1989): "Maxmin expected utility with non-unique prior," *Journal of Mathematical Economics*, 18, 141—-153.

GIUSTINELLI, P., C. F. MANSKI, AND F. MOLINARI (2019): "Precise of Imprecise Probabilities? Evidence from Survey Response on Late-onset Dementia," *NBER Working Paper*.

HE, J. AND P. NATENZON (2024): "Moderate Utility," *American Economic Review: Insights*, 6, 176–195.

KARNI, E. (2009): "A MEchanism for Eliciting Probabilities," *Econometrica*, 77.

——— (2018): "A Mechanism for Eliciting Second-Order Beliefs and the Inclination to Choose," *American Economic Journal: Microeconomics*.

——— (2021): "Incomplete Preferences and Random Choice Behavior: Axiomatic Characterizations," *Working Paper*.

KARNI, E. AND M.-L. VIERØ (2020): "Comparative Incompleteness: Measurement, Behavioral Manifestations and Elicitation," *Working Paper*.

KESSLER, J. B., C. LOW, AND C. D. SULLIVAN (2019): "Incentivized Resume Rating: Eliciting Employer Preferences without Deception," *American Economic Review*, 109, 3713–3144.

KRAJBICH, I., C. CAMERER, AND A. RANGEL (2017): "Exploring the Scope of Neurometrically Informed Mechanism Design," *Games and Economic Behavior*, 101, 49–62.

MANDLER, M. (2005): "Incomplete Preferences and Rational Intransitivity of Choice," *Games and Economic Behavior*, 255–277.

MAS-COLELL, A., M. D. WHINSTON, AND J. R. GREEN (1995): *Microeconomic Theory*, New York: Oxford University Press.

NIELSEN, K. AND J. REHBECK (2022): "When Choices Are Mistakes," *American Economic Review*, 112, 2237–68.

NISHIMURA, H. AND E. A. OK (2016): "Utility Representation of an Incomplete and Nontransitive Preference Relation," *Journal of Economic Theory*, 166, 164–185.

OK, E. A. (2002): "Utility Representation of an Incomplete Preference Relation," *Journal of Economic Theory*, 104, 429–449.

OK, E. A., P. ORTOLEVA, AND G. RIELLA (2012): "Incomplete Preferences Under Uncertainty: Indecisiveness in Beliefs versus Tastes," *Econometrica*, 80, 1791–1808.

RIGOTTI, L. AND C. SHANNON (2005): "Uncertainty and Risk in Financial Markets," *Econometrica*, 73, 203–243.

WANG, S., M. FILIBA, AND C. CAMERER (2010): "Dynamically Optimized Sequential Experimentation (DOSE) for Estimating Economic Preference Parameters," *mimeo*.

Online Appendix

A. Theories of Incompleteness

Here we overview the theoretical framework underlying our experiment. A model of decision making under subjective uncertainty is presented in Bewley (2002). It shows that a strict preference relation that is not necessarily complete, but satisfies all other axioms of the standard Anscombe-Aumann framework, can be represented by a unique utility index and a *set* of probability distributions. In this model, lack of completeness is thus reflected in multiplicity of beliefs: The unique subjective probability distribution of the standard expected utility framework is replaced by a set of probability distributions.

Since we restrict attention to monetary outcomes that depend on a binary event, we can denote the state as $\{nv, v\}$ (corresponding to "not verb" and "verb") and describe the set of all probability distributions over this state space using the interval $[0,1]$ with generic element π. Let $\succsim$ denote an individual's preference relation over elements of $\boldsymbol{R}^2$ (these are pairs of monetary outcomes in each state). Bewley's Knightian Decision Theory can be summarized by saying that for any $p, q \in \boldsymbol{R}^2$

$$p \succsim q \qquad \text{if and only if} \qquad E_\pi[u(p)] \geq E_\pi[u(q)] \text{ for all } \pi \in \Pi \tag{1}$$

where we define $E_\pi[u(p)] \equiv \pi u(p_{nv}) + (1-\pi)u(p_v)$, and let $u(x)$ denote the utility yielded by the monetary amount x, and let $\Pi \subseteq [0,1]$. That is, p is preferred to q if it has higher expected utility for all probability distributions corresponding to the interval Π. If the inequality in (1) changes direction for different elements of Π, the two alternatives are not comparable. In this model, alternatives are evaluated one probability distribution at a time, and they can be ranked only when all those evaluations agree. In this theory, the incompleteness is reflected by multiple subjective probabilities, and if Π reduces to a singleton the preferences are complete and this is standard subjective expected utility.[33]

Bewley's Knightian Decision Theory is extended in Galaabaatar and Karni (2013) by admitting not only multiple probabilities but also multiple utility functions. This

[33] See Rigotti and Shannon (2005) for a precise statement of this result.

model can be described by the following representation: for any $p, q \in \boldsymbol{R}^{\mathbf{2}}$

$$p \succsim q \qquad \text{if and only if} \qquad E_\pi[u(p)] \geq E_\pi[u(q)] \text{ for all } u \in \mathscr{U} \text{ and } \pi \in \Pi \tag{2}$$

where $\mathscr{U}$ is a set of utility functions for monetary amounts. That is, p is preferred to q if it has higher expected utility for all probability distributions and all utility functions. Alternatives are evaluated one probability and utility function pair at a time, and they can be ranked only when all those evaluations agree. Notice that even when the individual's subjective probability is unique incompleteness is reflected by many utility functions.

In a world where probabilities are objective Dubra et al. (2004) presents a model of incompleteness that weakens the original von-Neumann & Morgenstern axioms by dropping completeness. Preferences are described by a single *objective* probability distribution and many utility functions: for any $p, q \in \boldsymbol{R}^{\mathbf{2}}$

$$p \succsim q \qquad \text{if and only if} \qquad E_\pi[u(p)] \geq E_\pi[u(q)] \text{ for all } u \in \mathscr{U} \tag{3}$$

where, again, $\mathscr{U}$ is a set of utility functions for monetary amounts.[34] Here, p is preferred to q if it has higher expected utility for all utility functions. If the inequality in (3) changes direction for different elements of $\mathscr{U}$, the two alternatives are not comparable. In this model, alternatives are evaluated one utility function at a time, and they can be ranked only when all those evaluations agree. Whenever $\mathscr{U}$ contains a single utility function, this model reduces to von-Neumann & Morgenstern expected utility with objective probabilities.

[34]For other models inspired by Aumann (1962) see Ok, 2002; Eliaz and Ok, 2006; Ok et al., 2012.

B. ADDITIONAL TABLES AND FIGURES

Payment if Not Verb ($)	Payment if Verb ($)
5	19
5	16
6	18
7	10
7	16
8	17
7	19
8	13
6	12
9	14
9	9
9	11
10	8
10	13
10	6
13	3
16	3
11	5
12	1
11	10
12	6
14	2
14	4
17	1
12	8

Table VI: List of All Lotteries
Note: This presents the state-dependent payment values of all gambles as depicted in Figure II. (9, 11) and (14, 2) were the two reference lotteries.

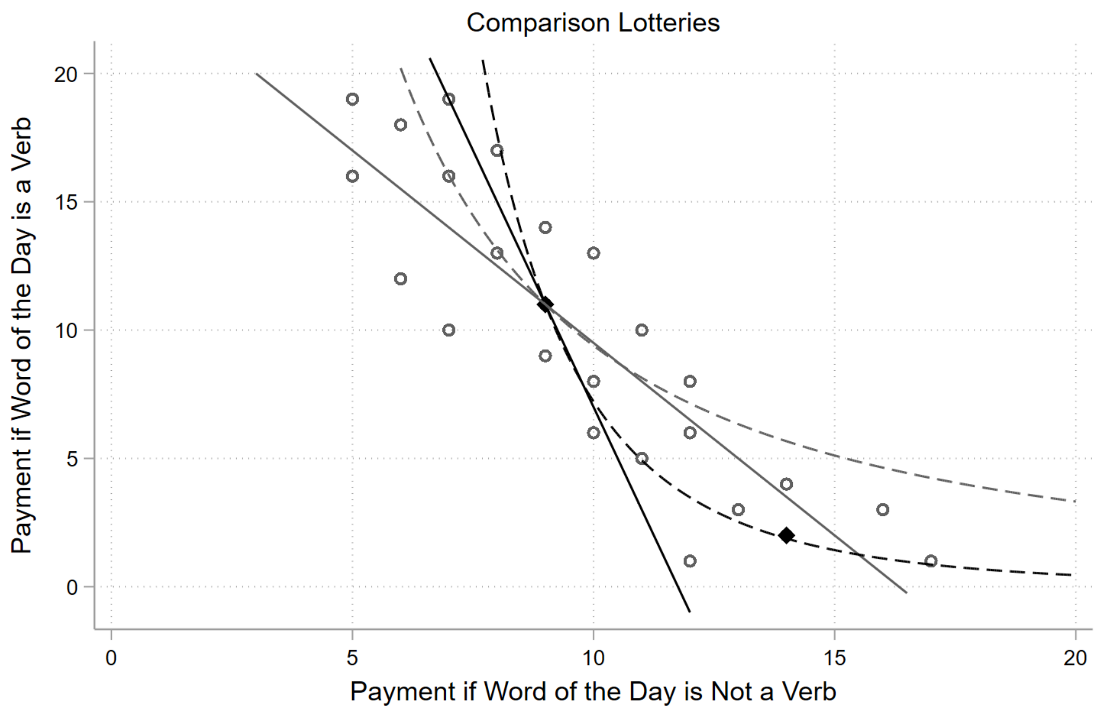

Figure VI: Log Utility Functions Through Reference Lottery (9,11)
Note: Solid lines show linear indifference curves while dashed lines show log-utility. The black lines show $pr(verb) = 0.2$ while grey lines show $pr(verb) = 0.4$.

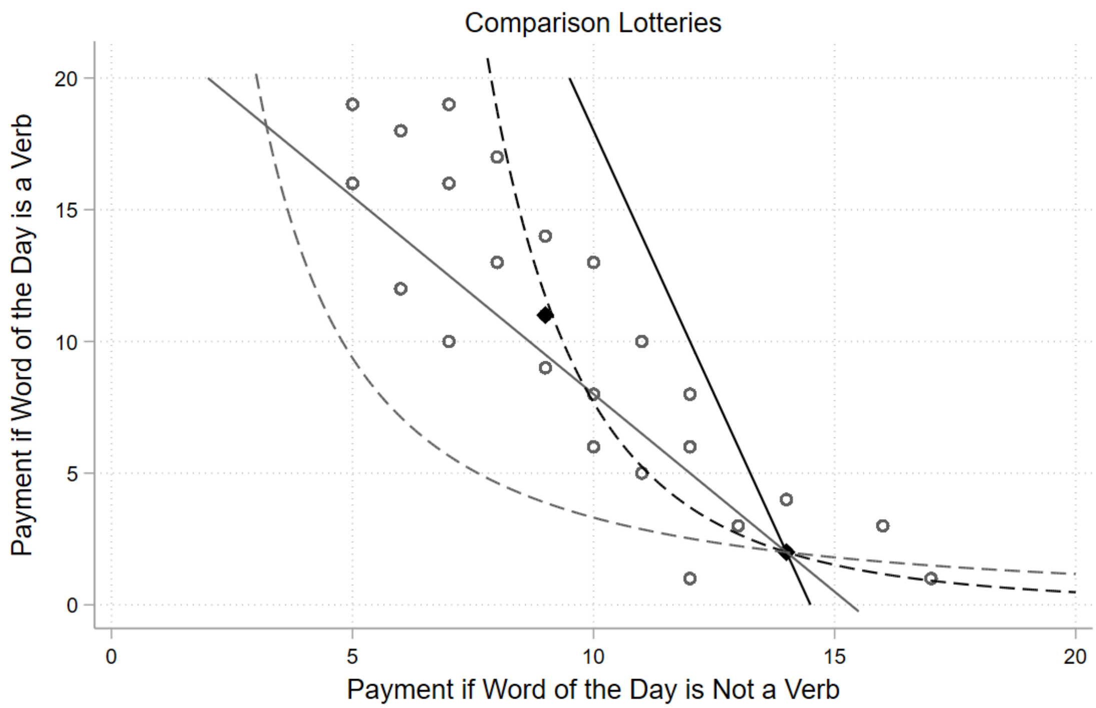

Figure VII: Log Utility Functions Through Reference Lottery (14, 2)
Note: Solid lines show linear indifference curves while dashed lines show log-utility. The black lines show $pr(verb) = 0.2$ while grey lines show $pr(verb) = 0.4$.

Decision 1 of 25:

Gamble 1	Gamble 2
$9 if the word-of-the-day is not a verb	$16 if the word-of-the-day is not a verb
$11 if the word-of-the-day is a verb	$3 if the word-of-the-day is a verb

I rank **Gamble 1** above Gamble 2

I rank Gambles 1 and 2 **exactly the same**

I **don't know** how I rank Gambles 1 and 2

I rank **Gamble 2** above Gamble 1

Figure VIII: Screenshot of Decision Screen in Non-Forced Treatment
Note: Decisions in the Forced treatment looked exactly the same, without the indifferent and incomplete answer options.

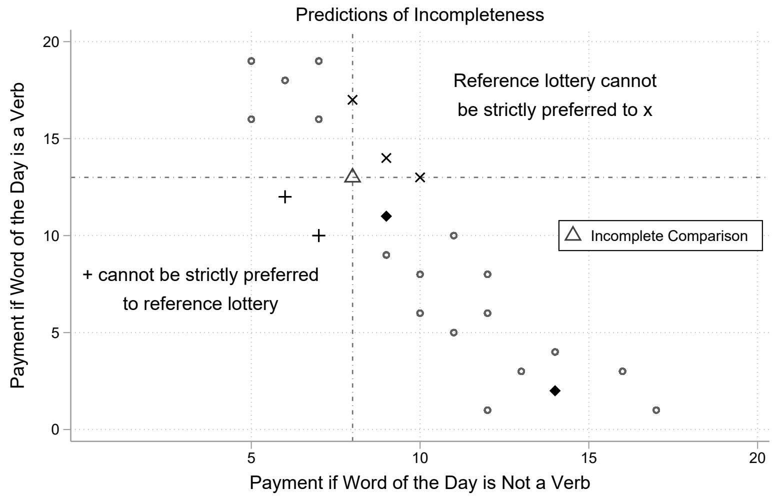

Figure IX: Predictions on Preferences Given Observed Incompleteness

Note: Points show the 25 lotteries used in the experiment. Given a reported incomparability—here the example shown is (8,13)—models of incompleteness constrain the reports for dominant and dominated lotteries. The reference lottery should not be strictly preferred to lotteries that dominate (8,13)—lotteries marked with an (x)—and lotteries that are dominated by (8,13)—marked with a (+)—should not be strictly preferred to the reference lottery.

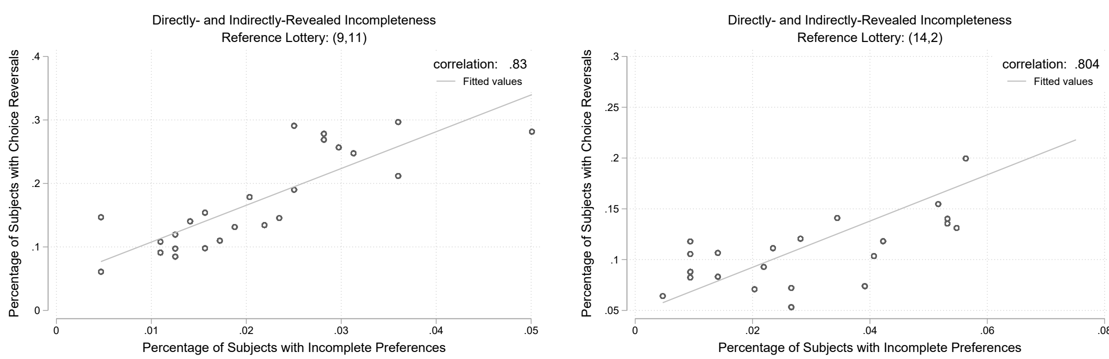

Figure X: Preference Reversals as a Potential Indicator of Underlying Incompleteness

Note: Each point on a graph represents one of our 24 comparison lotteries; we exclude questions in which subjects compare the reference lottery to itself. The two panels separate the two reference lotteries.

C. Instructions and Details on Algorithms

Here we present the exact language used for the instructions relating to our elicitation methodology. We first present the instructions for Experiments 1 and 2, respectively. Then we discuss the details of the algorithms in Experiment 2.

C.A. Experiment 1 Instructions

> *If you are selected to receive a bonus payment, and if you are selected to be paid for this part of the study, then we will use your answers to guess what you would prefer in another question and we will pay you based on what we think you would prefer in that question. We have a tested algorithm that uses your actual answers to these questions to understand your preferences between the two urns and guesses which urn you'd rather bet on. Importantly, your answers cannot affect which question you are paid for, but can only affect which Urn we draw a ball from in the unknown question. Therefore, it's in your best interest to answer what you really think.*
>
> - *If you choose to bet on one of the urns, then our algorithm will use this question to understand your relative preferences between the two urns.*
> - *If you say that you are indifferent between the two urns, then our algorithm will know when you think the urns are exactly the same. Because of this, you can only say that you are indifferent in at most one question, since this is the question that the algorithm will use to figure out when you are exactly indifferent.*
> - *If you say that you don't know which urn you prefer, then our algorithm will not use this question to understand your preferences. You can say that you don't know as many or as few times as you wish.*

We also included understanding questions to ensure that individuals understood what would happen in each of those three scenarios.

C.B. Experiment 2 Instructions

> *We will not actually pay you directly for the gambles in these question groups. Instead, we will pay you based on what your responses imply about what gambles you prefer in some other decision that you will not*

face. We will use your choices in this question group to understand what types of gambles you like or dislike.

At the end of the experiment, we will pick two gambles. We will pay you the gamble that we think you prefer out of those two randomly selected gambles. We will use your earlier choices to decide which gamble we think you prefer in this decision.

How do we use your choices to understand which gamble you prefer?

We have an algorithm, based on previous research, that learns how you rank gambles.

For example, if you tell us that you like \$5 more than \$4, the algorithm learns that you like more money to less money. If instead you had said that you like \$4 more than \$5, the algorithm would learn that you like less money!

Your choices teach the algorithm what type of gambles you like.

- *If you know which gamble you like better:*
  If you say that you rank one gamble over the other, then we will use this information to help our algorithm understand which gambles you would rather have.
- *If you like the two gambles equally:*
  If you say that you rank the two gambles equally, then we will use this information to help our algorithm understand when having either of two gambles is the same to you.
- *If you are not sure which gamble you prefer:*
  If you say that you do not know how to rank the two gambles, then we will not use that question in our algorithm.

Therefore, you have an incentive to tell us which gamble you rank higher, or that you rank them equally. If you do that consistently, our algorithm will use your answers to get a better idea of your preferred gamble at the end of the experiment.

You also have an incentive to tell us when you do not know so that we do not enter your choice into our algorithm. If you told us that you rank one above the other when in reality you were unsure, our algorithm might get the "wrong idea" about the types of gambles you prefer.

We also included understanding questions to ensure that individuals understood what would happen in each of those three scenarios.

C.C. Experiment 2 Algorithm Details

Our main data contains responses from 639 participants recruited using two different elicitation algorithms. These two algorithms potentially offer different reporting incentives and have different theoretical properties, but we show in Table VII that subjects' responses were the same regardless of the exact incentive mechanism.

We designed our first algorithm, that we call the "Set Construction Algorithm," to map out the theoretical objects of better-than and worse-than sets. We start with a "better-than set" and a "worse-than set" that each contain 10 randomly-selected lotteries from the space of all possible lotteries (x, y), with $x, y \in [0, 20]$. When a subject reports that (x, y) is strictly preferred to a reference lottery, we replace one of the lotteries in the better-than set with $(x+i, y+i), i \in [1,5]$.[35] When a subject reports that the reference lottery is strictly preferred to (x, y), we replace one of the lotteries in the worse-than set with $(x-i, y-i)$. When a subject reports indifference between a reference lottery and (x, y), we replace one of the lotteries in the better-than set *and* one in the worse-than set as described above. When a subject reports incompleteness, we do not change the better than and worse than sets.

If a subject were paid from this procedure, then we randomly selected to pay them from the better-than set or worse-than set with equal chance. If we randomly selected the worse-than set, then the subject would receive the reference lottery as payment (since they preferred the reference lottery to any lottery in the worse-than set).[36] If we randomly selected the better-than set, then we would randomly select one of the lotteries in the better-than set, and the subject would receive this lottery as payment.

In theory, this procedure is desirable for a few reasons. First, given that subjects are never paid for any of the exact choices p vs. q that they make throughout the experiment, we are never forced to complete their preference between p and q com-

[35] In most of our data, $i = 1$. However, technically there are two comparison lottery pairs for which $(x_1 + 1, y_1 + 1) = (x_2, y_2)$, so one of the lotteries that could be put into the better-than or worse-than set was itself a comparison lottery. Because of this, in later treatments, we randomly select i and put a restriction to prevent this. We do not believe this makes a difference in practice.

[36] Technically, this means that subjects do not have a strict preference to report when a lottery is worse than the reference lottery, since they will be paid the reference lottery anyway. We could have introduced strict incentive by replacing a lottery from the worse-than set and *also* changing the better-than set, for example, by removing any lotteries in the better-than set that were worse than the current lottery. We avoided this for the additional complication, and so that there was no risk of the better-than set becoming empty. Instead, we included lotteries that were dominated by the reference lottery so that we could identify whether subjects report when the reference lottery is strictly better. As we discuss in Section IV.F, 90% of subjects do.

parisons in which they report incompleteness. Second, the procedure described above allows for estimated preferences to be incomplete. We are simply constructing better-than and worse-than sets, and the union of these sets need not (and in fact does not) span the entire space of lotteries. Third, in this simple set construction procedure, we make no functional form or other parametric assumptions, so choices only inform preferences through dominance. Finally, the question we ultimately pay subjects is one where we are "sure," under dominance, that they have complete preferences.

The one major drawback of this type of procedure is that the questions subjects answer influence the possible lottery choices they could be paid, exactly because of this last feature where we only pay subjects for a choice where they have complete preferences. In particular, the lotteries that are replaced into the better-than and worse-than sets are a function (via dominance) of the questions that a subject answers with strict preference or indifference. As a result, this algorithm is not incentive compatible for all possible beliefs and preferences. For example, if a subject thought that the better than set contained "very good" lotteries to start with, then they would not want to indicate that any comparison lotteries were preferred to the reference lottery, since this would replace a "very good" lottery with an inferior one. There is no reason for subjects to hold this belief and it is not accurate—the better-than and worse-than sets start out with randomly-generated lotteries, and we state this in the additional information about the algorithm—but we cannot entirely rule out such beliefs.[37]

Given these pros and cons, we re-ran our experiment using a different—and potentially more standard—incentive mechanism, that we call the "MLE Algorithm." Here, we fixed a single payment question for all participants, $(14,2)$ vs. $(9,5)$, which was not one of the questions that subjects faced in the experiment. Using a standard maximum likelihood approach, we took all of the questions in which a subject reported strict preference or indifference to estimate a CRRA utility function for each subject who received a bonus payment. Incomparabilities simply did not enter into this estimation. We then calculated whether the subject would prefer $(14,2)$ or $(9,5)$ given their estimated utility function, and we paid them based on this prediction.

Here, the incentives are, loosely, that reporting strict preferences or indifference

[37]We needed the sets to start with randomly-selected lotteries to give subjects an incentive to answer questions where they had a strict preference. Otherwise, a subject could, for example, answer one question and be guaranteed that lottery for payment.

can help the estimation procedure form a more precise estimate of the utility function. Since subjects do not know what question is being paid, if their response could change the estimated utility function, then there are questions where this change in estimate could lead to different payment lotteries. When subjects are not sure of their preference in a given comparison, reporting incompleteness prevents this comparison from potentially biasing the estimation.

The drawbacks of this type of procedure are, in some sense, the opposite of the Set Construction Algorithm. Here, we fix a payment question ex-ante and we force a complete preference in this comparison. It could be the case that subjects actually have incomplete preferences in this question, but this procedure does not allow for that. Furthermore, we also have to use a specific functional form for the estimation. In theory, it could be that subjects have complete but non-CRRA preferences, which could lead to biases in reporting.

Reference Lottery	Prefer Reference	Prefer Comparison	Indifferent	Incomplete
Set Construction	34.2%	55.0%	8.2%	2.6%
MLE	33.4%	56.2%	7.7%	2.7%

Table VII: Aggregate Choice Data
Note: Subjects made 25 comparisons for each reference lottery. The table presents the percentage of subjects who preferred the reference lottery, preferred the comparison lottery, were indifferent between the two, and were unable to compare the two, aggregated across subjects.

Table VII shows that subjects' responses are statistically indistinguishable across these algorithms (Fisher's exact $p = 0.249$). The same percentage of subjects clicked to learn more about the algorithm ($p = 0.829$), and the choices among those who read the algorithm details are also statistically indistinguishable ($p = 0.098$)). Given that these two algorithms are very different from one another, and that the theoretical pros and cons are quite opposite, we find this compelling evidence that the details of the underlying algorithm do not affect subjects' responses.

In running our MLE Algorithm treatment, we also recruited 166 additional subjects and incentivized them under the same algorithm but exogenously provided them with the information about the algorithm. That is, rather than choosing to learn the information or not by clicking a button, we showed this information to all participants. Specifically, we tell subjects the following:

> *We will use maximum-likelihood estimation to estimate a constant relative risk aversion utility function. Maximum likelihood estimation will find*

the constant relative risk aversion parameter such that your choices are most probable under this model. Then, we will use that model to predict what you would choose in another question, and will pay you based on this prediction. The choices that you make will help this estimation procedure to choose the parameter that best fits your preferences. If you say that you "do not know" which gamble you prefer in a question, we will not use this question in our maximum likelihood estimation. We will use only the questions where you know which gamble you prefer.

We link Wikipedia pages for maximum likelihood estimation, constant relative risk aversion, and utility functions.

We find a slight reduction in the percentage of subjects who report incompleteness in this treatment (31% vs. 41%, $p = 0.071$). We find it implausible that this reflects subjects reading and optimally responding to the MLE and CRRA detailed incentives. Instead, this seems in line with Danz et al. (2022) who find that providing detailed information on incentives can lead to subjects distorting their responses. In support of this, we find that, among subjects who ever report incompleteness, they report the same percentage of incomplete comparisons at a question level ($p = 0.414$) regardless of whether information was exogenously provided. Thus, it appears that some subjects are “scared off” or confused by the elicitation details, so too much information can cause a reduction in incompleteness on the extensive margin but not on the intensive margin.