When Are Decisions Improvable? An Evaluation of Diagnostic Methods *

B. Douglas Bernheim   Aldo Lucia   Kirby Nielsen   Charles D. Sprenger
Stanford   Ohio State   Caltech   Caltech

March 10, 2026

Abstract

We evaluate three methods for identifying improvable choices: documenting specific misconceptions (the Characterization Assessment method), gauging confidence in choices (the Decision Confidence method), and showing that specific behavioral patterns in the domain of interest also emerge in a related domain where they are objectively suboptimal (the Pattern Matching method). In experiments involving risky choice, the three methods imply that different choices are improvable and have conflicting implications regarding legitimate risk preferences. We clarify the assumptions underlying each method and reevaluate the evidence on risk-taking in light of their limitations.

1 Introduction

A central theme of Behavioral Economics is that people sometimes make suboptimal choices. This observation motivates paternalistic interventions that aim to improve well-being by correcting decision-making imperfections (see, e.g., Bernheim and Taubinsky, 2018, for an overview). Examples include sin taxes, bans, mandates, and "nudges." Evaluating such policies involves a fundamental diagnostic challenge—detecting when choices are *improvable*, meaning that a social planner with the same information as the decision-maker, full knowledge of the decision-maker's normative objectives, and perfect optimization technology would make a different selection.[1] Without reliable identification of improvable choices, it is impossible to evaluate the welfare impacts of even the most well-intended behavioral policies.

The standard approach to evaluating the welfare of a fallible decision-maker is to construct a *Welfare-Relevant Domain* (WRD)—the set of choices taken to reveal the individual's normative values—by excluding from the set of all choices those flagged as improvable according to diagnostic criteria. For any given criterion, a positive test result is taken to indicate that the choice is improvable. In contrast, a negative result is *not* taken to indicate that the choice is unimprovable; rather, it implies that the choice does not exhibit the particular imperfections the test seeks to detect. In other words, diagnostic criteria justify *exclusions* from the WRD rather than *inclusions*. Because the analyst can use multiple diagnostic criteria in combination, a test that fails to detect many improvable choices can still be useful provided it does not incorrectly classify many unimprovable choices as improvable. Consequently, in this context, false positives are more problematic than false negatives.[2]

The leading diagnostic tool is *characterization assessment* (CA): The analyst removes choices from the WRD when there is evidence that the decision-maker misunderstood key aspects of the problem (i.e., suffered from *characterization failure*); see, e.g., Bernheim and Rangel (2004), Chetty et al. (2009), Bronnenberg et al. (2015), Allcott and Taubinsky (2015), Bernheim and Taubinsky (2018), Allcott et al. (2019), and Ambuehl et al. (2022).[3] The relia-

---

[1]Throughout, we depart from the existing literature by referring to problematic choices as *improvable* rather than as *mistakes*. If a choice is a mistake, then it is also improvable, but not all improvable choices are mistakes. For example, if a shortcut is optimal when accounting for the decision-maker's cognitive costs, then its use is not a mistake. However, the choice is still improvable in the sense that someone with a superior decision technology, or an opportunity to conserve cognitive effort by making decisions for multiple parties simultaneously, could do better. Given the difficulty of establishing whether decision-making imperfections reflect optimal shortcuts or suboptimal heuristics, it is more accurate to call such choices *improvable.*

[2]When evidence of improvability is used to justify paternalistic interventions that overrule freely made choices, the principles of classical liberalism argue for a high standard of proof (see, e.g., Bernheim, 2025), and consequently also underscore the importance of avoiding false positives.

[3]To be clear, the existence of a misunderstanding does not necessarily mean that the individual selected the wrong option. However, because it introduces this possibility, it renders the choice an unreliable guide for the planner.

bility of the CA method hinges on the analyst's ability to identify essential facts and determine whether the decision-maker understands them. Furthermore, the method is narrow: Diagnosis is limited to the facts the analyst specifies, and the method is not designed to detect other forms of decision-making imperfections. This narrowness motivates consideration of broader diagnostic methods.

In this paper, we first evaluate whether CA is a reliable and comprehensive method for diagnosing improvable choices. We then assess two alternative methods that potentially offer broader diagnostics: statements of self-reported decision confidence (DC) and "pattern matching" (PM)—the comparison of anomalous choice patterns between a domain of interest and objective errors in an induced-value matching domain. We find that CA misses a meaningful share of improvable choices. However, inferences about improvability based on the DC and PM methods are neither reliable nor consistent. Consequently, while the narrowness of CA is an important limitation, the DC and PM methods, as currently formulated, are not well-suited for detecting instances of improvability that the CA method overlooks.

Section 2 outlines a conceptual framework for understanding the defining characteristics of improvable choices. As we discuss, decision-making imperfections can arise from failures of either characterization or optimization. While the empirical detection of characterization failure is relatively straightforward, diagnosing optimization failure is more challenging. Consequently, the CA method has become the most prominent tool for identifying improvable choices. However, recognizing that CA is narrow, the literature has explored other diagnostic criteria, some of which cast broader nets. Alternatives to CA include documenting (1) misunderstandings of the principles that govern the relationships between actions and their consequences (e.g., Ambuehl et al., 2022), (2) conditions that impair understanding, including cognitive limitations (e.g., Bernheim and Rangel, 2004), (3) inconsistencies, including choice reversals (e.g., O'Donoghue and Rabin, 1999) and violations of choice axioms (e.g., Nielsen and Rehbeck, 2022), (4) dominated choices (e.g., Bronnenberg et al., 2015), and (5) seemingly unreasonable choices that violate normative consensus or expert opinion (e.g., Bronnenberg et al., 2015). For a discussion and evaluation of these methods, see Bernheim and Taubinsky (2018).

Interest in using the DC and PM methods to detect improvable choices has emerged more recently. A potential advantage of both is that they may diagnose a broad range of decision-making imperfections rather than a narrow set of characterization failures.

The DC method relies on self-reported confidence in the quality of an individual's judgments (*decision confidence*). This method has a 140-year track record in cognitive psychology dating back to Peirce and Jastrow (1884). This literature establishes that self-reported confidence often provides a good indicator of performance in objective perceptual tasks.[4] More recently, as

---

[4]Psychologists have termed the relationship between confidence and performance in these tasks "metacogni-

Enke et al. (2025) have noted, the behavioral economics literature has proposed using measures of decision confidence in economic contexts as "indicators that a decision was made using an imperfect (i.e., heuristic or noisy error-prone) choice rule" when objective performance benchmarks are unavailable.[5] Indeed, they describe these measures as "signatures of error-prone decision making." Whether low confidence is actually a good proxy for perceived improvability in the context of economic decision-making is an open question. However, the reasoning behind the DC approach is clear: Even without an objective benchmark, when someone thinks they may have made a decision poorly, it may be wise to credit their assessment.[6] Of course, the opposite is not true: Because misconceptions are often confidently held, *high* confidence in a decision is almost certainly a poor indicator of unimprovability. However, as noted above, the purpose of improvability diagnostics, including those based on the DC method, is to justify exclusions from the WRD rather than inclusions. Regardless of what *high* confidence signifies, *low* confidence may indeed capture "decision difficulty" and "imprecision, and the errors that result" (Butler and Loomes, 1988).

The PM method detects improvability through parallel choice patterns. Specifically, suppose one wishes to know whether a behavioral pattern in some primary domain reflects preferences or, alternatively, improvable choices. The strategy of the PM method is to create "similar" induced-value tasks in a "matching domain" where the same pattern would unambiguously imply that choices are improvable.[7] If the pattern arises in the matching domain, then the analyst attributes it to common decision-making imperfections in both domains, and consequently concludes that the behavioral pattern in the primary domain is a manifestation of improvability.[8] Examples of studies that employ the PM method include Cason and Plott (2014),

---

tive sensitivity" (Maniscalco and Lau, 2012; Fleming and Lau, 2014). A rich literature exploring this connection has developed. Rahnev et al. (2020) constructs a database of over 100 studies linking confidence to performance and meta-analyzes the relationship between the two.

[5]Much of the literature in behavioral economics on decision confidence is concerned with relationships between *cognitive uncertainty* and behavioral regularities (see, e.g., Enke and Graeber, 2023). To be clear, cognitive uncertainty and improvability are different concepts.

[6]In this spirit, Enke et al. (2025) conclude that hyperbolic discounting in monetary discounting experiments is a mistake based in part on correlations with low confidence.

[7]Structuring the "similar" induced-value tasks of the matching domain is by no means straightforward. First, the analyst must preserve the sources of improvability they hypothesize within the primary domain. For instance, as Oprea (2024) notes, if the matching domain tasks are too transparent, they may eliminate common sources of improvability, attenuating pattern matching. Second, the analyst must avoid introducing other sources of improvability that are specific to the matching domain, or alternatively provide ways to identify the choices they infect. Critically, the analyst cannot determine whether their efforts are successful based on observed pattern matching—i.e., they cannot conclude that the matching domain tasks are appropriate or inappropriate based on whether patterns do or do not match. To do so would involve circular reasoning. Without evidence-based criteria for evaluating whether specific induced-value tasks are appropriate, the analyst is forced to rely on impressionistic judgments. The difficulty of resolving debates involving impressionistic arguments about the adequacy of specific induced-value tasks potentially limits the value of improvability diagnostics based on the PM method.

[8]Along these lines, Oprea (2024) concludes that pattern matching between prospect-theoretic deviations from

Martínez-Marquina et al. (2019), Oprea (2024), and Vieider (2024). The reasoning behind the PM method is also clear: If objective errors in a matching domain mirror the behavioral patterns in a primary domain, then the primary-domain phenomenon may also involve errors.

We investigate CA, DC, and PM as tools for identifying improvable choices in several experiments with more than 4,000 subjects and 74,000 decisions. Section 3 outlines the design of our main experiment, and we provide details concerning other experimental conditions where relevant. Much of our analysis focuses on choices over lotteries—specifically, on whether small-stakes risk aversion and the Common Ratio Effect (CRE) are manifestations of improvability rather than preferences. We also study Risk-Free decision tasks that resemble the Risky tasks and serve as the "matching" domain for the PM method. Critically, improvability is objectively identifiable in the Risk-Free tasks.

In Section 4, we present our main analysis of the CA method. Our first finding is that CA does not attribute small-stakes risk aversion or the CRE to improvability. However, our second finding is that, despite being highly diagnostic, CA misses a substantial fraction of objectively improvable choices in the Risk-Free domain. Considering that the method is likely more comprehensive in the Risk-Free domain than in the Risky domain, it may also fail to detect many improvable Risky choices, which calls the first finding into question.

Therefore, in Section 5, we turn to the broader DC and PM diagnostics. We find that they have conflicting implications: While the PM method attributes small-stakes risk aversion and the CRE to improvability, the DC method attributes their *absence* to improvability. This conflict motivates a deeper dive into the assumptions and performance of these methods.

Section 6 evaluates the assumptions and reliability of the DC method. We show that, with high frequency, expressions of low confidence do not imply that people think their choices are improvable.[9] For example, people often express low confidence because they are uncertain their choice will turn out favorably. In other cases, low confidence signifies near-indifference or the perception of normative ambiguity. Furthermore, the proper interpretation of these confidence judgments, and thus the degree to which low confidence reflects perceived improvability, varies considerably between the Risky and Risk-Free domains. These findings cast doubt on the

---

risk-neutrality and objective mistakes in Risk-Free induced-value tasks at both the aggregate and individual level implies that "theories of risk preferences designed to explain these anomalies...are unlikely to contain much normative content and therefore should not be accommodated in the inference of welfare or the design of policy."

[9]Prior research in cognitive psychology on metacognitive sensitivity and related work in economics shows that greater confidence does not always correlate with better performance (Rahnev et al., 2020; Jin et al., 2022; Enke et al., 2023). However, this work does not distinguish between our hypothesis, that decision confidence typically mismeasures perceived competence, and the alternative hypothesis that perceived competence is sometimes negatively correlated with performance. Unlike our hypothesis, the alternative hypothesis does not necessarily challenge the use of decision confidence as a criterion for excluding a choice from the WRD. High perceived competence may correlate with poor performance because it is associated with confidently held misconceptions. Assuming the analyst is reasonably successful at detecting those misconceptions through characterization assessments, perceived competence may be *incrementally* diagnostic of improvability.

suitability of low decision confidence as an improvability diagnostic. To be clear, our analysis does not rule out the existence of meaningful correlations between low confidence and other phenomena such as attenuated behavioral responses to changes in a task's parameters (as discussed in Enke and Graeber, 2023; Enke et al., 2024).[10] Rather, we highlight the risks of inferring the perceived improvability of choices from reports of low confidence.

Section 7 evaluates the assumptions and reliability of the PM method. We first outline six hypotheses that could account for parallelism in classical risk-taking patterns (small-stakes risk aversion and the CRE) between Risky and Risk-Free choices. Pattern matching implies that Risky choices are improvable for three of these hypotheses but not for the other three. For example, because this method employs Risk-Free tasks that resemble Risky tasks, people may mistakenly think the Risk-Free tasks involve risk. These *recognition failures* mechanically produce pattern matching. To be clear, a high incidence of recognition failures tells us nothing about the improvability of Risky choices or the general reliability of the PM method. It may simply mean that the analyst has adopted a confusing format for presenting Risky and Risk-Free choices. The critical questions are whether pattern matching persists when subjects are *not* confused and, if so, why.

One can attempt to evaluate pattern matching when subjects are not confused in two ways: Either try to ensure through experimental design that few if any subjects are confused, or focus on the subset of tasks they properly understand. To implement the first strategy, we include "recognition enhancement" treatments that highlight the absence of risk in each Risk-Free task. Given that some confusion is inevitable, we also implement the second strategy. For example, our main experiment assesses each subject's characterization of every decision task they perform, providing a measure of recognition failure in Risk-Free tasks. Regardless of which approach we use, we find relatively little evidence of pattern matching when subjects are not confused: Small-stakes risk-aversion and the CRE are highly robust for Risky tasks but are sharply attenuated for Risk-Free tasks. Moreover, additional analyses attribute the residual pattern matching to the other two hypotheses that do not imply improvability of Risky tasks, rather than to the three that imply the opposite. These findings highlight the delicacy of the PM method and underscore the importance of thoroughly investigating the causes of parallelism.

Thus, while our results highlight the limitations of the CA method, they also imply that the DC and PM approaches, as currently formulated, suffer from serious confounds and are not well-suited for detecting the instances of improvability that the CA method overlooks.

---

[10] Though not central to our analysis, we find that low confidence is correlated with *amplification* of responses to changes in parameters for Risky choices, and with attenuation for Risk-Free choices; see Appendix A.4. Other studies have also found amplification in specific contexts, such as near the boundaries of parameter ranges; see Enke et al. (2024) for examples and some potential explanations. Further investigation of the factors that determine whether attenuation or amplification prevails in a given context would be useful.

Taken together, our findings highlight the need for new methods—or improvements on existing methods—that can reliably and accurately detect improvability. Moreover, our analysis provides tools that are potentially useful for evaluating such methods. We conjecture that these improved methods and evaluative tools might also prove useful in applications beyond the scope of this paper, such as analyzing the complexity of choice problems and choice domains.

2 When are choices improvable?

To make progress toward an operationalizable definition of an improvable choice, it is useful to distinguish between two logically separable components of decision-making: *characterization* and *selection*. Characterization involves acknowledging the existence of a decision problem, identifying available options, and developing an understanding of how those options map to consequences. Selection follows characterization: The individual determines which bundle of consequences is best and makes a choice from the alternatives they identified.

We can also usefully distinguish between two aspects of selection: *normative judgment* (or "judgment" for short), which refers to the criteria a decision-maker uses when deciding whether one bundle of consequences is better than another, and *optimization*, which entails identifying and choosing the best option among a collection of alternatives according to those criteria.

For the sake of greater precision, we adapt the formulation of Bernheim and Taubinsky (2018). Suppose a decision-maker must choose an action, x, from an opportunity set, X. They care intrinsically (as opposed to instrumentally) about a vector of consequences, $y \in Y$. Conditional on choosing x and knowing information I,[11] the CDF for consequence y is $F(y \mid x, I)$. The function u captures the individual's normative objectives; its argument is the CDF, $F(\cdot \mid x, I)$.[12] The best possible choice is then the solution to the following problem:

$$\max_{x \in X} u\left(F(\cdot \mid x, I)\right). \tag{1}$$

To make a choice, the decision-maker must form an understanding of the opportunity set and the mapping from choices to distributions of consequences, formulate objectives, and optimize. In principle, difficulties might crop up in any of these stages: They may think the constraint set is Z rather than X (for example, ignoring options would imply $Z \subset X$) or that the mapping from actions and information to distributions of consequences is G rather than F; they may take their objective function to be v rather than u; or they may fail to identify or choose the maximal element. The first type of phenomenon ($(Z, G) \neq (X, F)$) is characterization

---

[11] For the sake of simplicity, we take the individual's information set to be fixed. However, one can apply a similar framework to settings where the individual makes choices impacting information acquisition.

[12] This formulation subsumes expected utility as a special case, but is also consistent with other aggregators.

failure,[13] the second ($v \neq u$) is judgment failure, and the third is optimization failure. When any of these three imperfections is present, the decision-making process is improvable. To be clear, the choice itself is not necessarily improvable because the decision-maker may fortuitously select the right option for the wrong reason. However, improvable decision processes do not yield choices that reliably reflect the decision-maker's normative objectives.

Even though one can define failures of normative judgment with mathematical precision ($v \neq u$), it is unclear that their existence is amenable to objective evidence-based proof. What evidence could prove that someone who perfectly characterized a decision task and optimized over their options should have pursued a different objective? Claims of judgment failure that appear in the literature often implicitly reference misunderstood or overlooked consequences, which means they instead invoke characterization failure (see, e.g., the discussion of Sunstein, 2020 in Bernheim, 2025). For example, consider the claim that people consume sugary drinks excessively because their objectives place "too little weight" on health. How might one prove this claim? One possibility is to demonstrate that people do not consider health consequences even though they care about health. Another is to show that they have false beliefs concerning the health risks of sugary drinks. A third is to establish that they misunderstand the implications of health risks for longevity and quality of life. While these strategies may legitimately establish that the consumption of sugary drinks is excessive, they do so by documenting various characterization failures rather than judgment failures. If someone with full knowledge of all consequences thinks the benefits of sugary drinks exceed their costs, then we cannot prove that their judgment is wrong by marshaling facts and citing consequences they already acknowledge.

It follows that efforts to diagnose improvability should focus on characterization failure and optimization failure. One can provide evidence of characterization failure either directly, by eliciting the decision-maker's understanding of the available options and/or the mapping from options to consequences, or indirectly, by demonstrating that a decision-maker misunderstands the general principles that govern the relationships between actions and their consequences, or by documenting conditions that likely impair that understanding.

While characterization failures lend themselves to empirical investigation, detecting optimization failures is more challenging. Sometimes, optimization failures are detectable because they go hand-in-hand with characterization failures. To illustrate, imagine someone knowingly employs a cognitive shortcut involving a simplified representation of the relationship between choices and outcomes; they act as if the simplification is accurate even though they know it is only an approximation. Their beliefs about consequences may or may not correspond to those the simplification implies but, either way, they will not hold correct beliefs, except by chance.

---

[13]This formulation of characterization failure assumes F describes objective probabilities. For settings in which all probabilistic assessments are subjective, one would focus instead on the accuracy and completeness of information and objective inferences used in the process of forming beliefs.

Diagnosing other forms of optimization failure is more difficult. One challenging possibility is that a decision-maker who fully understands the set of options and their consequences might simplify by restricting consideration to a subset of their alternatives.[14] Another challenging possibility is that someone who must expend effort to assess or apply their true preference, u, might settle for a rough approximation. For instance, if the application of risk preferences involves more cognitive effort than simple calculations, then a risk-averse decision-maker might intentionally act as an expected-value maximizer in settings involving small-stakes risk. Unlike the optimization failures mentioned in the preceding paragraph, these shortcuts would not necessarily manifest in ways characterization assessments would detect.

One might nevertheless hope to capture elusive optimization failures by using broad diagnostics that detect general symptoms of cognitive difficulties. Methods that rely on documentation of choice reversals (e.g., O'Donoghue and Rabin, 1999), dominated choices (e.g., Bronnenberg et al., 2015), and violations of choice axioms (e.g., Nielsen and Rehbeck, 2022) fall into this category, but also have limitations (see Bernheim and Taubinsky, 2018). The decision confidence (DC) and pattern matching (PM) methods may provide effective alternatives. The DC method can potentially capture any imperfection provided the chooser is aware of it. For example, someone who knowingly restricts consideration to a subset of their options may express less-than-complete confidence that they have made the best possible selection. The PM method casts an even broader net, in that it can potentially detect imperfections even if the decision-maker is unaware of them. Thus, we design and implement a series of experiments to evaluate these methods along with the more targeted Characterization Assessment (CA) method.

3   Overview of Experimental Design

We evaluate methods for diagnosing improvability in the context of risky choice. This domain is particularly suitable for our investigation because the literature on risk-taking identifies a collection of robust patterns that are irreconcilable with standard consumer theory. Whether these patterns reflect improvability or preferences remains an open question. We focus on two such regularities: small-stakes risk aversion and the common ratio effect (CRE). Our experiments allow us to evaluate these regularities using characterization assessments, decision confidence, and pattern matching. Other features of our experimental design provide opportunities for testing the reliability of each method's implications.

In the following subsections, we first outline the structure of the Risky tasks we study, as well as the "Risk-Free" tasks that serve as the matching domain when we deploy the PM method.

---

[14]Recent work attempts to identify individuals' consideration sets and highlights the challenges involved in these exercises (see Abaluck and Adams-Prassl, 2021; Barseghyan et al., 2021; Ellis et al., 2024, among others).
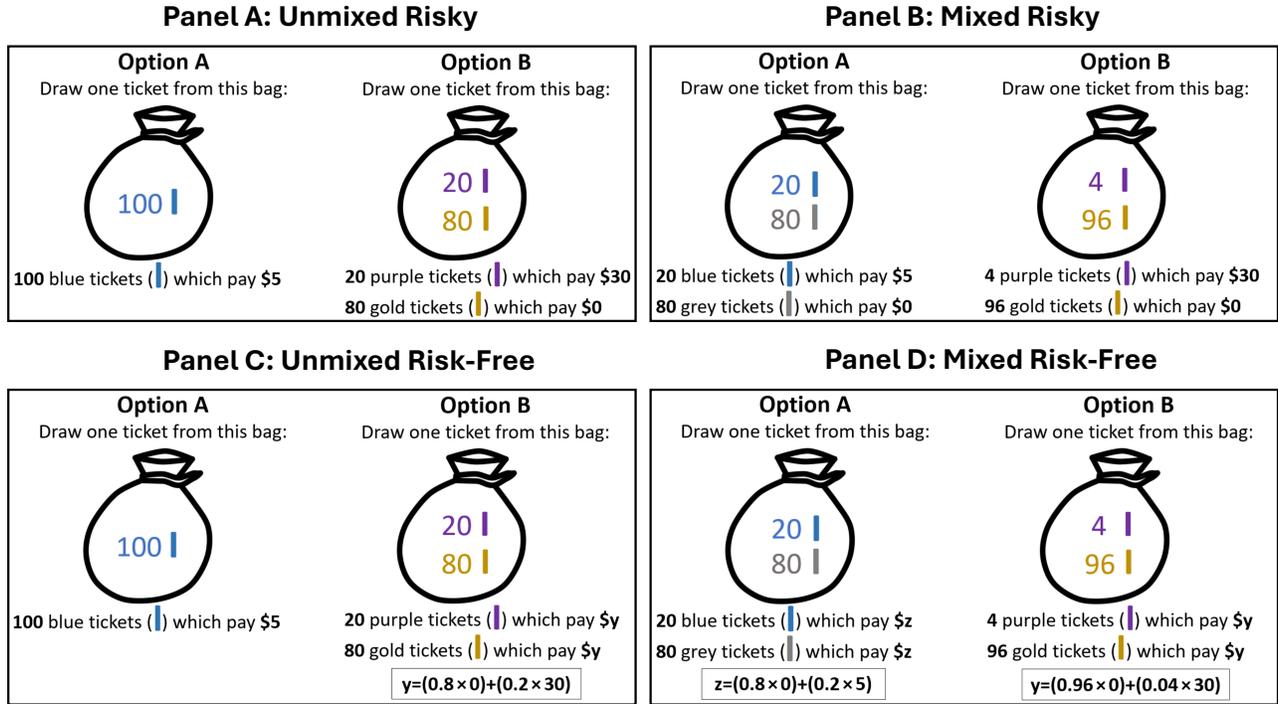
Panel A: Unmixed Risky

Option A
Draw one ticket from this bag:

100 |

100 blue tickets (|) which pay **$5**

Option B
Draw one ticket from this bag:

20 |
80 |

20 purple tickets (|) which pay **$30**
80 gold tickets (|) which pay **$0**

Panel B: Mixed Risky

Option A
Draw one ticket from this bag:

20 |
80 |

20 blue tickets (|) which pay **$5**
80 grey tickets (|) which pay **$0**

Option B
Draw one ticket from this bag:

4 |
96 |

4 purple tickets (|) which pay **$30**
96 gold tickets (|) which pay **$0**

Panel C: Unmixed Risk-Free

Option A
Draw one ticket from this bag:

100 |

100 blue tickets (|) which pay **$5**

Option B
Draw one ticket from this bag:

20 |
80 |

20 purple tickets (|) which pay **$y**
80 gold tickets (|) which pay **$y**
y=(0.8×0)+(0.2×30)

Panel D: Mixed Risk-Free

Option A
Draw one ticket from this bag:

20 |
80 |

20 blue tickets (|) which pay **$z**
80 grey tickets (|) which pay **$z**
z=(0.8×0)+(0.2×5)

Option B
Draw one ticket from this bag:

4 |
96 |

4 purple tickets (|) which pay **$y**
96 gold tickets (|) which pay **$y**
y=(0.96×0)+(0.04×30)

Figure 1: Risky and Risk-Free binary choice tasks

After presenting the confidence measures we use for the DC method and the characterization assessments we use for the CA method, we provide details concerning implementation.

3.1 Risky Tasks

Our Risky tasks consist of standard binary choices between lotteries that map states to outcomes. Subjects choose between two lotteries denoted Options A and B. We describe the lotteries as "bags" containing colored tickets. Each color represents a possible state; the number of tickets per color represents the probability associated with that state; and the ticket-specific prize is the outcome associated with the state. We explain that, if a given task determines the bonus payment, then the subject will receive the payment associated with one ticket drawn randomly from the chosen bag. Figure 1, Panel A shows an example of a Risky task.

Each Risky task takes one of two forms: Unmixed or Mixed. In an Unmixed task, a subject chooses between a certain amount M and a lottery that pays either $30 with probability p or $0 with probability $1 - p$. For each Mixed task, we multiply the probabilities of the positive prizes in an Unmixed task by a common factor r, so that a subject chooses between receiving M with probability r and $30 with probability $p \cdot r$ (where the complementary event pays $0 for both options). The common ratio effect (CRE) is present when, contrary to expected utility theory, choices in Mixed tasks skew toward riskier options relative to choices in Unmixed

tasks.[15] Panel A of Figure 1 shows an Unmixed Risky task and Panel B shows the corresponding Mixed Risky task.

We fix $r = 0.2$ throughout and vary p in $\{0.10, 0.20, 0.50, 0.80, 0.90\}$ so we can examine changing risk tolerance across low and high probabilities. For each p, the value of M is either $p \cdot 30 + 1$ (a "+1 task") or $p \cdot 30 - 1$ (a "−1 task"). We call the difference in expected value between the two options the "value difference." Because the +1 and −1 tasks are balanced, a risk neutral decision-maker would prefer the safe(r) option in 50% of all tasks. Small-stakes risk aversion manifests as a preference for the safe(r) option in more than 50% of tasks. There are 20 possible Risky tasks (5 p values × 2 M values × 2 task types {Unmixed, Mixed}), of which each subject completes a randomly selected subset (see below).

3.2 Risk-Free Tasks

Because we are interested in evaluating the PM method, our experiments include Risk-Free tasks that closely parallel the Risky tasks. In a Risk-Free task, all states map to the same common outcome, which subjects must calculate. The required calculation coincides with determining the expected payoff for the lottery in the corresponding Risky task.

Consider an Unmixed Risky task where a subject chooses between \$$M$ for sure and \$30 with probability p. In the corresponding Unmixed Risk-Free task, a subject chooses between \$$M$ for sure and \$$y = \$p \cdot 30$ for sure, and the value of y appears as a formula. Thus, the "probabilistic p" from the Risky task becomes an "algebraic p" in the Risk-Free task. Similarly, consider a Mixed Risky task where a subject chooses between \$$M$ with probability r and \$30 with probability $p \cdot r$. In the corresponding Mixed Risk-Free task, a subject chooses between \$$z = \$r \cdot M$ for sure and \$$y = \$r \cdot p \cdot 30$ for sure, where both z and y appear as formulas. Panel C of Figure 1 shows an Unmixed Risk-Free task and Panel D shows the corresponding Mixed Risk-Free task.

There are 20 possible Risk-Free tasks, encompassing the same combinations of p, M, and task types (Mixed and Unmixed) as for Risky tasks. Each subject similarly completes a randomly selected subset of the Risk-Free tasks. Critically, in Risk-Free tasks, the value of M fully determines which option is *objectively* best.

The PM method can shed light on the quality of decision making in primary domain only if matching-domain tasks preserve the hypothesized mechanisms that putatively lead people to make improvable choices in the primary domain. While the existing literature does not provide rigorous objective criteria for designing matching-domain tasks that meet this requirement, a natural strategy is to change the substance of the primary-domain task while maintaining a

---

[15]Recent work has discussed whether classic common ratio comparisons provide valid tests of expected utility (see McGranaghan et al., 2024). We focus on the CRE phenomenon itself rather than on testing expected utility, sidestepping these inference concerns.

similar presentation. Accordingly, our main design uses Risk-Free tasks that preserve many elements of our Risky tasks. For both Risky and Risk-Free tasks, there is a common state space—i.e., a set of tickets, one of which is drawn at random. The mapping from states to outcomes is non-degenerate for Risky tasks and degenerate for Risk-Free tasks. However, the visual stimuli we use for these two domains are not *identical*. Specifically, we express ticket-specific payoffs as numbers in Panel A of Figure 1 and as formulas in Panel C.

Given that the literature has not yet outlined or analyzed rigorous design criteria, it is unsurprising that other studies use different matching-domain tasks. For instance, in Oprea (2024), Risky and Risk-Free tasks involve identical visual stimuli, but they do not reference a common state space. In principle, this difference could either increase or decrease the frequency with which subjects mistakenly interpret Risk-Free tasks as Risky.[16] Critically, the best design does not necessarily minimize the incidence of these recognition failures. A more important consideration is whether the design permits the identification of choices for which such failures occur, so that the analyst can focus on the choices subjects properly understand. To this end, a novel aspect of our design is that we include characterizations assessments for each task, which we describe below. Indeed, we settled on our stimuli because that they allow us to perform identical characterization assessments for Risky and Risk-Free tasks. As a robustness check, we also explored a design with identical stimuli for Risky and Risk-Free tasks along the lines of Oprea (2024). We describe this alternative design along with associated results in Section 7.

3.3 Confidence

After a subject chooses their preferred option in a given Risky or Risk-Free task, we ask: "On a scale from 0 (not confident at all) to 100 (completely confident), how confident do you feel about this choice?" They respond using a slider. These judgments enable us to evaluate small-stakes risk aversion and the CRE using the DC method. We can also relate Risk-Free choices to stated confidence and test whether objective mistakes are more prevalent when subjects say they lack confidence in their choices.

In follow-up experiments, we probe the meaning of these responses by asking subjects to explain their less-than-complete confidence. Our objective is to determine whether expressions of low confidence reflect considerations that imply perceived improvability. Additionally, we explore the sensitivity of our results to an alternative formulation of the confidence question, along with associated instructions and a comprehension test, that closely follow Enke et al. (2024). See Section 6 for details.

---

[16]With our design, the randomness of ticket selection may evoke risk even when payoffs are invariant. However, in contrast to Oprea's design, subjects can tell whether a task is Risky or Risk-Free directly from the stimulus.

Figure 2: Characterization Tables for the Unmixed Risky task in Figure 1, Panel A

3.4 Characterization Tables

To implement the CA method, we ask subjects to fill out a "characterization table" for each option in each task. The table elicits information ordinarily required for correct characterization—specifically, the probability of drawing a ticket of each color along with the associated payment. Typically, a subject who misunderstands this information suffers from characterization failure. To promote and evaluate comprehension, we require subjects to fill out two sample characterization tables correctly during the instructions, one for a Risky task and one for a Risk-Free task. Subjects who initially complete these sample tables incorrectly receive feedback.

Figure 2 shows the characterization table associated with the Unmixed Risky lottery in Figure 1, Panel A. Each table includes all ticket colors appearing in both options, and subjects can erroneously assign positive probability to zero-probability colors.[17] For Mixed tasks, there are four possible ticket colors instead of three, so the tables include an additional column.

3.5 Pre-Choice and Post-Choice Conditions

In our Post-Choice condition, subjects first make all choices and provide associated confidence judgments, then they fill out all characterization tables without incentives. This condition allows us to implement the DC and PM methods without potentially contaminating them by probing characterization contemporaneously. A potential drawback is that these ex post assessments may not reflect characterization at the moment of choice. Consequently, we also consider Pre-Choice conditions with contemporaneous characterization: Subjects complete the corresponding characterization table immediately prior to making each decision, and the table remains on the screen while they make their choice. Subjects complete the sample character-

---

[17]Subjects first listed the odds associated with each ticket color, then listed the prize associated with each color to which they ascribed non-zero odds.

ization tables after making all choices for the Post-Choice condition and prior to making any choices for the Pre-Choice conditions. Thus, our treatment manipulation varies the timing of both the characterization assessments and any learning that occurs as a result of feedback.[18] In the Incentivized Pre-Choice condition, eligible subjects receive the bonus payment for a given task only if they complete the corresponding characterization table correctly. In the Unincentivized Pre-Choice condition, the bonus payment is not conditional on correct characterization.

We randomized each subject into one of these three treatments. As outlined in our pre-analysis plan, our main analysis pools the two Pre-Choice conditions (Incentivized and Unincentivized) because the results are similar; see Appendix A.2.

3.6   Subjects' Choices and Logistics

Each subject completed 20 choice tasks randomly selected from the 40 Risky and Risk-Free tasks described above, in random order. Any given subject did not necessarily perform both the Unmixed and Mixed versions of the same task, or both the Risky and Risk-Free versions. The initial instructions included an illustration of the calculation required to solve for the formulaic payments in the Risk-Free tasks as well as practice tasks of both types.[19] To explore whether intermixing Risky and Risk-Free tasks led to confusion, we conducted follow-up experiments in which we separated these tasks into blocks; see Section 7.

3.7   Subjects

For our main experiment (Experiment 1), we recruited 999 subjects through Prolific. We restricted the sample to high-school graduates with high approval ratings who were located in the United States. We recruited an equal number of male and female subjects. The median completion time, 34.9 minutes, was similar across all three conditions. Subjects received a guaranteed payment of $7.50 and had a 20% chance of being eligible for a bonus payment. For each eligible subject, we randomly selected and implemented one of the 20 choice tasks. In the Incentivized Pre-Choice condition, eligible subjects received the resulting payment only if they filled out the corresponding characterization table correctly. The average bonus payment among eligible subjects was $8.70. We recruited roughly 3,000 subjects through Prolific for our follow-up experiments; see subsequent sections for similar details.

---

[18]For detailed analysis of the characterization assessment training data, see Appendix Section A.5.

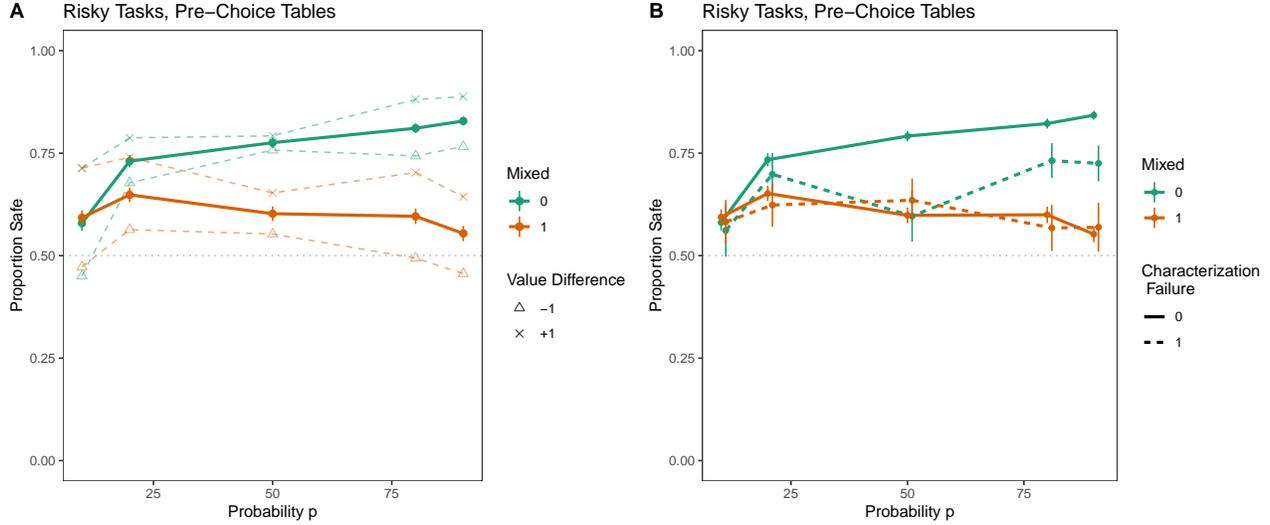[19]Screenshots from the experiment appear in Appendices B.1.1 through B.1.5.

Figure 3: CA Method: Risky Tasks, Pre-Choice Condition

Notes: Figure 3 shows choices in Risky tasks for subjects in the Pre-Choice condition ($N = 6{,}674$). Panel A reports the proportion of choices for the Safe option in Unmixed tasks (green) and Mixed tasks (orange). Light shaded exes and triangles denote $+1$ and -1 tasks, while dark shaded circles indicate sample averages. Bars represent ± 1.96 standard errors (95% CIs). Panel B reports sample averages separately for observations without characterization errors (solid) and with characterization errors (dashed) in the corresponding tables.

4   Implications and Limitations of the CA Method

In this section, we apply the CA method to Risky tasks with the objective of determining whether small-stakes risk aversion and the CRE are manifestations of improvability (Section 4.1). We then assess the reliability of our conclusions by providing evidence on the method's ability to detect objective errors (Section 4.2). Throughout, we focus on the Pre-Choice condition because the associated characterization tables capture subjects' assessments of the tasks at the time of choice.

4.1   Application to Risky Choice

Figure 3, Panel A plots the frequencies with which subjects select the safe option in all Unmixed (solid green) and Mixed (solid orange) tasks, with 95% confidence intervals, for each value of p.[20] The light dashed lines show these frequencies separately for the $+1$ and -1 tasks.

A risk-neutral expected-value maximizer would choose the safe option 100% of the time in $+1$ tasks and 0% of the time in -1 tasks. Because our experiment involves equal proportions of $+1$ and -1 tasks, the risk-neutral benchmark corresponds to choosing the safe option 50% of

---

[20]In a slight abuse of terminology, we refer to the mixed version of the safe option as "safe" and the mixed version of the risky option as "risky."

the time.[21] A higher frequency indicates small-stakes risk aversion. As in most of the literature, we focus on Unmixed tasks to measure small-stakes risk attitudes. The CRE manifests as a higher frequency of choosing the safe option in Unmixed tasks than in Mixed tasks.

Subjects choose the safe option in Unmixed tasks 74% of the time, which indicates substantial small-stakes risk aversion ($t = 32.4$, $p < 0.01$, versus 50%).[22] They choose it significantly less often in Mixed tasks, resulting in an overall CRE (i.e., the difference in choice proportions between all Unmixed and all Mixed tasks) of 15%-age points ($t = 12.9$, $p < 0.01$). Both patterns are apparent in Figure 3, Panel A: Choice probabilities lie above 50%, especially for Unmixed tasks. The CRE is most prevalent (i.e., the gap is largest) for higher values of p and is virtually absent for low values. These results essentially replicate standard findings.[23]

Next, we ask whether small-stakes risk aversion and the CRE reflect improvable choices according to the CA method. Overall, subjects completed 11% of Pre-Choice characterization tables incorrectly for Risky tasks.[24] In Panel B of Figure 3, we plot the frequencies with which subjects select the safe option separately for choices with and without errors in characterization tables. Subjects choose the safe option 75% of the time in Unmixed tasks with correct characterization tables versus 67% in Unmixed tasks with incorrect tables ($t = 3.0$, $p < 0.01$).[25] Similarly, the CRE is 15%-age points in tasks with correct tables versus 8%-age points in tasks with incorrect tables ($t = 2.1$, $p = 0.04$). Thus, the CA method does not attribute either small-stakes risk aversion or the CRE to improvability.

4.2   Comprehensiveness of the CA Method

Especially considering that subjects complete only 11% of characterization tables incorrectly, it is important to examine the possibility that these assessments paint an incomplete picture of improvability. While the tables gauge understanding of the mapping from states to outcomes, any given decision-maker may require additional information to characterize a problem, such as

---

[21]Of the 6,674 Risky Pre-Choice tasks, 3,323 (49.8%) were +1 tasks and 3,351 (50.2%) were −1 tasks.

[22]For simplicity, throughout the text we provide t-tests and correlations without controlling for task characteristics or clustering standard errors. For each statistic, we also estimated corresponding regressions controlling for relevant task characteristics and clustering standard errors at the individual level. The conclusions are almost always unchanged. When there is a meaningful difference, either qualitatively or with respect to statistical significance, we describe it in a footnote.

[23]We obtain virtually identical results in the Post-Choice condition; see Section 5.

[24]The median subject correctly completes 97% of their Risky tables; 50% (333/666) fill out all tables correctly, while 2% (16/666) fill out all incorrectly. Such failures are largely independent of p, task type (Mixed vs. Unmixed), and sign (+1 vs. −1). However, the error rate falls from 13% to 9% with incentives ($t = 5.4$, $p < 0.01$). Notably, tables with errors take longer to complete on average than those without (65.3 vs. 45.4 seconds, $t = 9.5$, $p < 0.01$), suggesting that mistakes are not simply due to excessive speed or inattention.

[25]Notably, the lower frequency of safe choices in tasks with characterization errors does not imply closer conformity to expected value maximization: 58% of choices with correct tables maximized expected value, compared to 54% with incorrect tables ($t = 2.1$, $p = 0.03$).

mean and variance. Undiagnosed instances of improvable choices would potentially undermine our conclusion regarding Risky tasks by contaminating the solid lines in Panel B of Figure 3.

To evaluate the comprehensiveness of the CA method, we turn to the *Risk-Free* domain. This investigation is feasible because, in that domain, there is always an objectively correct selection, and all other selections are objectively improvable. Critically, in our experiment, the Risk-Free domain is also highly favorable to the CA method: A subject who has correctly filled out a Risk-Free task's characterization table simply needs to identify which of two fixed payments is larger, the one associated with all tickets in Option A, or the one associated with all tickets in Option B. Accordingly, the scope for other forms of characterization and optimization failure is far more limited in the Risk-Free setting than in the Risky setting.

Table 1: Characterization Errors and Objective Mistakes in Risk-Free Choice

		Maximization	
		No Failure	Failure
CA	Correct Characterization Table	0.745	0.027
			$\beta = 0.255$
	Incorrect Characterization Table	0.149	0.079
		$\alpha = 0.167$	

$$N = 6{,}646;\ t = 24.8,\ (p < 0.01)$$

Notes: Cross-tabulation of objective mistakes and characterization table errors in Pre-Choice Risk-Free tasks. α denotes the Type-I error rate, i.e., the likelihood of a false positive conditional on "No Failure," and β the Type-II error rate, i.e., the likelihood of a false negative conditional on "Failure." t-statistic corresponds to test of equality in maximization failures for observations with correct and incorrect characterization tables.

Table 1 cross-tabulates CA characterization errors and objective maximization failures for Risk-Free tasks in the Pre-Choice condition. Overall, 11% of Risk-Free choices exhibit objective maximization errors and 23% of Risk-Free characterization tables contain at least one error. Characterization errors strongly predict improvability: 97% of choices with accurate tables are correctly maximized, versus only 65% of choices with incorrect tables ($t = 24.8$, $p < 0.01$). Even so, the Type-II error rate (β) is substantial: The CA method fails to diagnose around 25% of maximization failures.[26] Accordingly, even in this highly favorable setting, the frequency with which this method fails to identify improvable choices is concerning.

Understanding why a subject might complete a characterization table accurately only to select the smaller of two fixed payoffs can help us appreciate the limitations of the CA method.

---

[26]The table also shows that the Type-I error rate (α) is around 17%. For the CA method, we are relatively unconcerned with Type-I error. While people sometimes make the right choices in incorrectly characterized decision problems, they do so purely by chance. If one knows that an individual incorrectly characterized an essential feature of a decision task, then their choice is necessarily an unreliable guide for a social planner, and hence unsuitable for inclusion in the WRD.

While noisy responses often explain anomalous choices in experiments, here there is a natural alternative explanation: These subjects may be *unsure* of their calculations, and their aversion to this subjective uncertainty may justify choosing the "safer" option even when their calculations imply that it yields the lower payoff.[27] This consideration may be especially consequential when one option requires no calculation, as is the case for Unmixed tasks.

Consistent with this hypothesis, subjects tend to avoid the option requiring calculation in tasks they correctly characterize but incorrectly maximize: Within this set, they choose Option A 68% of the time in Unmixed tasks (vs. 50%: $t = 3.9$, $p < 0.01$), compared with only 57% in Mixed tasks (vs. 50%: $t = 1.1$, $p = 0.27$) where both options require calculation.

Thus, it appears that there are meaningful sources of improvability in the Risk-Free domain that the CA method does not capture. Because the Risky domain presents even greater challenges for the CA method, we would expect an even higher rate of Type-II error, which calls the conclusions of Section 4.1 into question.[28]

The limited comprehensiveness of characterization assessments is difficult to remedy for two reasons. First, different people may care about different aspects of a task. To illustrate, suppose the mean of a lottery's return is relevant for some people but not for others. Not asking subjects for the mean limits diagnostic comprehensiveness among those who care about it, but asking them for it can produce false diagnoses of characterization failure among those who do not. Second, even if the analyst could determine every relevant feature of a choice task for each decision-maker,[29] some of this information might be difficult or impossible to elicit (e.g., ex ante assessments of expected disappointment). Additionally, as noted in Section 2, even ideal characterization assessments would not detect certain types of optimization failures.

---

[27]Technically, this possibility is a form of characterization failure: The CDF governing the actual payoff from each option is degenerate, but the individual chooses based on non-degenerate CDFs. In principle, a more discerning characterization assessment could elicit this (counterfactual) feature of beliefs.

[28]We cannot directly measure Type-II error in Risky tasks because no single selection is objectively correct. As a proxy, we consider choice reversals—choosing the lottery in the +1 task but the safer option in the −1 task. Among 1,614 cases in which a subject faced both versions of the same Risky task in our Pre-Choice conditions, 6% exhibit a choice reversal. While such reversals may reflect fuzzy preferences or near-indifference rather than improvability (see Section 6), they nonetheless cast doubt on the normative validity of these choices. Within this subsample, the CA method detects only 25% of reversals, missing 75% (see Appendix Table A1 for details).

[29]One strategy for assessing relevance is to investigate whether making certain features of the decision problem explicit rather than implicit impacts choice (e.g., explicitly stating the mean of a distribution). It is tempting to conclude that a feature is relevant if and only if there is an effect, but caution is warranted. Statements about irrelevant features can influence choices through experimenter demand effects and other framing phenomena. Statements about relevant features can fail to influence choices because the decision-maker considers them even when they are not mentioned. Another possibility is to rely on open-ended statements concerning relevant features, but these are often difficult to interpret and may themselves involve misconceptions.
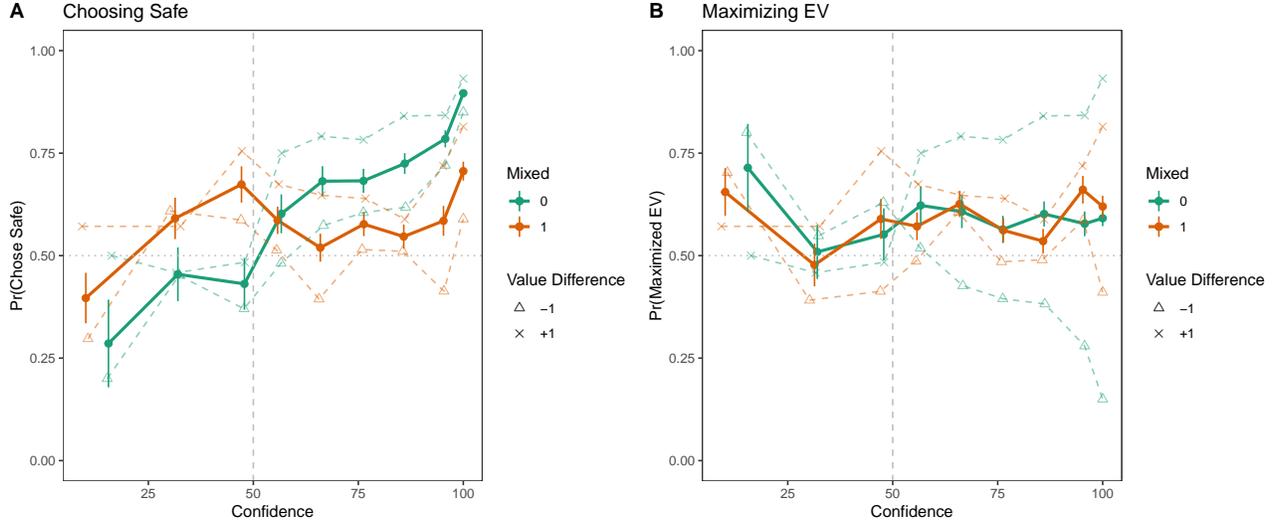
Figure 4: Decision Confidence Method: Risky Tasks and Confidence, Post-Choice Condition

Notes: Figure 4 shows behavior in Risky tasks for subjects in the Post-Choice condition ($N = 3{,}325$). Panel A reports the relationship between stated confidence and the probability of choosing the Safe option. We divide confidence responses into nine bins and, within each bin, calculate both average confidence and the share choosing Safe. Results are shown separately for Unmixed tasks (green) and Mixed tasks (orange). Light-shaded exes and triangles denote $+1$ and -1 tasks, while dark-shaded circles indicate sample averages. Bars represent ± 1.96 standard errors (95% CIs). Panel B repeats this analysis using the proportion of subjects maximizing expected value.

5 Implications of the DC and PM Methods

Because the CA method may fail to identify many improvable risky choices, broader diagnostics are potentially valuable. In this section, we apply the DC and PM methods.

The Decision Confidence (DC) Method

The DC method diagnoses improvability based on expressions of confidence. A pattern that disappears or attenuates as confidence rises is attributed in whole or in part to improvability.

Figure 4, Panel A shows the relationship between confidence and the proportion of safe choices in Mixed and Unmixed Risky tasks. To avoid the possibility that completing characterization tables influences confidence assessments, we restrict attention to the Post-Choice condition. A striking pattern emerges: Confidence is highly correlated with the likelihood of choosing the safer alternative, particularly in Unmixed tasks where that alternative involves no risk ($\rho = 0.30$, $p < 0.01$ in Unmixed tasks, and $\rho = 0.08$, $p < 0.01$ in Mixed tasks). Consequently, when one focuses on the choices in which subjects had greater confidence, the evidence of small-stakes risk aversion and the CRE is *even stronger* than for the full sample.

To underscore this point, we classify a choice as improvable if expressed confidence is below 50 (the vertical dashed line in Panel A). By this criterion, 11% of choices are improvable, which
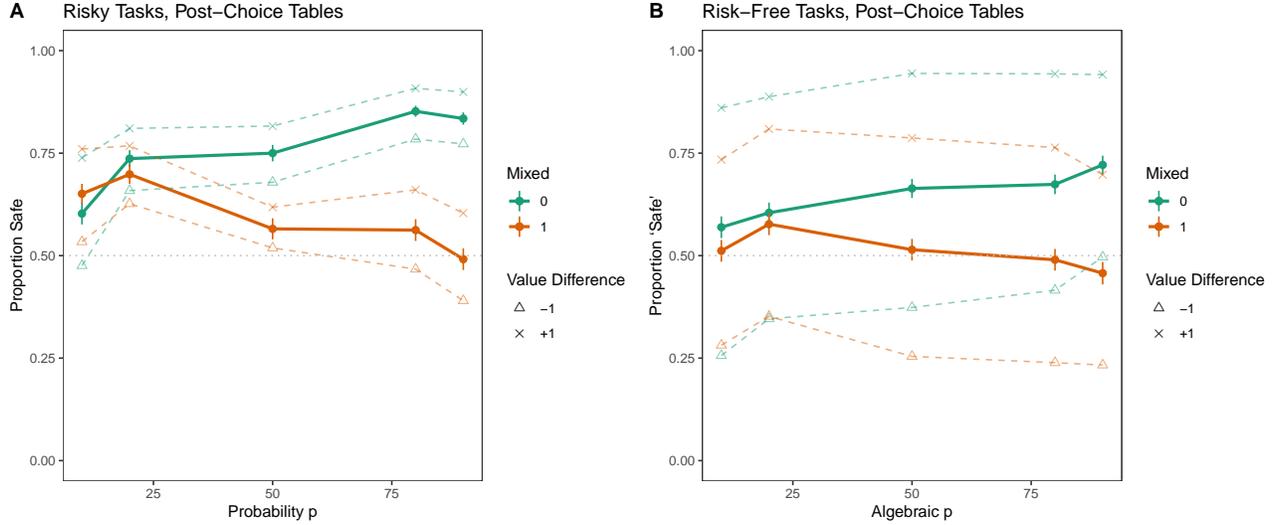
Figure 5: Pattern Matching Method: Risky and Risk-Free Tasks, Post-Choice Condition

Notes: Figure 5 shows the proportion of choices for the Safe/"Safe" option in Unmixed tasks (green) and Mixed tasks (orange). Light-shaded exes and triangles denote $+1$ and -1 tasks, while dark-shaded circles indicate sample averages. Bars represent ± 1.96 standard errors (95% CIs). Panel A reports Risky tasks ($N = 3,325$), and Panel B reports Risk-Free tasks ($N = 3,335$) for subjects in the Post-Choice condition.

is about the same as for the CA method. For those choices, the frequency with which subjects selected the safe option is actually below 50% (43% for improvable choices vs. 78% for other choices, $t = 8.5$, $p < 0.01$), and the CRE is negative (-15%-age points for improvable choices vs. $+19$%-age points for other choices, $t = 6.5$, $p < 0.01$). Thus, the DC method implies that the absence of small-stakes risk aversion and the CRE, rather than their presence, reflects improvability in Risky choice.

Interestingly, Panel B of Figure 4 shows that the overall correlation between confidence and expected value maximization is negligible: $\rho = 0.01$ ($p = 0.85$) in Unmixed tasks, and $\rho = 0.03$ ($p = 0.23$) in Mixed tasks. Consequently, the DC method does not generally associate the absence of expected value maximization with improvability.

The Pattern Matching (PM) Method

The Pattern Matching (PM) method attributes a behavioral pattern in a primary domain to improvability when the pattern also emerges in a matching domain as a result of unambiguous errors. Our primary domain is Risky choice and our matching domain is Risk-Free choice.

Figure 5, Panel A replicates Figure 3, Panel A, except we use data from the Post-Choice condition to avoid possible contamination from the characterization assessments; the patterns are similar. Panel B provides an analogous analysis of Risk-Free choices: The algebraic p replaces the probability p and the "Safe" choice references Option A. For the Post-Choice condition, 24% of Risk-Free choices involve maximization failures. Moreover, these failures

produce patterns that closely resemble those in Panel A. We observe small-stakes risk aversion in Panel A and its mirror image in Panel B: The frequencies with which subjects choose the Safe/"Safe" options in Unmixed problems substantially exceed the benchmark of 50% (76% in Risky tasks, $t = 24.3$, $p < 0.01$, and 64% in Risk-Free tasks, $t = 12.4$, $p < 0.01$, in each case against the null hypothesis of 50%). We also observe the CRE in Panel A and its mirror image in Panel B: subjects choose the Safe/"Safe" options more frequently in the Unmixed tasks than in the Mixed tasks. The magnitudes of these effects are similar—17%-age points in Risky tasks and 14%-age points in Risk-Free tasks ($t = 10.4$, $p < 0.01$, and $t = 8.0$, $p < 0.01$, respectively, against the null hypothesis of 0). Additionally, in both settings, the CRE/"CRE" increases in magnitude with the value of p. Because the deviations from value maximization in Risk-Free tasks roughly match the deviations from expected value maximization in Risky tasks, the PM method attributes small-stakes risk aversion and the CRE in Risky tasks to improvability.[30]

Summary of Comparison

Results based on the DC and PM methods have conflicting implications: While the PM method attributes small-stakes risk aversion and the CRE to improvability, the DC method attributes the *opposite* patterns to improvability. In principle, the two methods may detect different types of errors, in which case the properties of unimprovable choices are unclear. Alternatively, either or both of these methods may have methodological shortcomings. In the next two sections, we spell out the assumptions underlying both methods and investigate their validity.

6 Does low decision confidence reliably diagnose perceived improvability?

The DC method assumes that expressions of low confidence in choices reliably capture perceptions of improvability. The validity of this premise hinges on the assumption that respondents interpret the confidence question to mean precisely what the analyst intends.[31] In this section, we evaluate this assumption by exploring the intended meaning of responses to questions about decision confidence. Our objective is to distinguish between the following eight potential interpretations, some of which imply perceived improvability and some of which do not.

Interpretations that imply perceived improvability:

---

[30]We can also investigate pattern matching on an individual level. For the Post-Choice condition, there are 1,616 instances of a single subject completing a Risky task and its analogous Risk-Free task. The correlation between maximization in the Risk-Free tasks and expected value maximization in the Risky tasks is $\rho = 0.15$ ($p < 0.01$). As emphasized by Oprea (2024, 2025), the PM method construes correlation across domains as evidence for the hypothesis of parallel errors.

[31]We do not address a second assumption, that respondents answer unincentivized questions honestly.

1. Limited comprehension of the available options: The subject is unsure they interpreted the available options correctly or understood their consequences.

2. Potentially mistaken application of decision criteria: The subject thinks the criteria they decided to use along with any simplifying assumptions they made may actually favor an alternative other than the one they chose.

3. Insufficient consideration of objectives: The subject knows they have not thought through what they want to achieve as carefully as they could have.

4. Use of imperfect mental shortcuts: The subject is worried about having used a mental shortcut that might have given the wrong answer.

Interpretations that do not imply perceived improvability:

5. Normative ambiguity: The subject has "fuzzy" preferences.

6. Uncertainty about ex post sensations: The subject is unsure how they will feel about each possible outcome once it materializes.

7. Riskiness of the chosen option: Based on the information they have received, the subject cannot be sure about the outcome their chosen option will yield, and consequently recognizes that it may turn out to be worse than the alternative ex post.

8. Near indifference: The subject is nearly indifferent between their alternatives despite having little uncertainty about the "utility" either would yield.

To be clear, for Risk-Free choices, while explanations 5, 6, and 7 do not imply *perceived* improvability, they may be symptomatic of deeper characterization failure, and hence indicative of *actual* improvability. Someone who correctly characterizes a Risk-Free choice would understand that risk is absent, ruling out explanation 7. Moreover, as long as they robustly prefer more to less, they would not experience low decision confidence as a result of normative ambiguity or uncertainty about ex post sensations (explanations 5 and 6). Such responses likely reflect a tendency, documented in Section 7, for subjects to misconstrue Risk-Free choices as Risky.

Our concern about the DC method is that, regardless of the confidence question's technical meaning, subjects may adopt any of the preceding interpretations when responding. For example, the typical subject may construe questions about ex ante decision quality as probing ex post quality (i.e., interpretations 6 and 7). If subjects favor interpretations 5 through 8 over interpretations 1 through 4, then the DC method will systematically misdiagnose perceived improvability. We designed a follow-up experiment, Experiment 2, to explore this concern.

6.1 Experimental design

As in the Unincentivized Pre-Choice and Post-Choice conditions of our main experiment, Experiment 2 begins with Risky and Risk-Free tasks along with confidence elicitations. We then explain the eight potential reasons for low decision confidence, test subjects' comprehension, and ask them to attribute their less-than-complete confidence to these specific rationales.

Experimental strategy. Drawing inferences about the nature of low decision confidence from subjective responses raises a conceptual difficulty: Just as subjects may misinterpret questions about decision confidence, they may also misinterpret questions about the reasons for less-than-complete confidence. One cannot address this concern simply by phrasing the questions more precisely: Carefully wording the explanation of a concept so that its meaning is clear to an economist does not necessarily make it clear or memorable to an experimental subject, especially to one who exercises only moderate care in reviewing instructions.

Our design mitigates this concern by leveraging lessons from research on concept learning and categorization (e.g., Rosch and Mervis, 1975, Medin and Schaffer, 1978, Nosofsky, 1986, and Ashby and Maddox, 2005), vocabulary learning (e.g., Stahl and Fairbanks, 1986, and Beck et al., 2013), and educational psychology (Sweller, 1988, Renkl, 1997, and Kurtz et al., 2001). A theme of this work is that people learn the meanings of abstract categories most effectively from *examples* and *comparisons* rather than from formal definitions or rules. Leading models posit that people develop operational understandings of categories based on "prototypes" or "exemplars" that clarify conceptual definitions. When classifying a stimulus, they identify the illustrative cases it resembles most closely and assign it to an associated category.

Consistent with these principles, our description of each potential reason for low decision confidence has three parts: a conceptual explanation, a prototypical example, and a memorable label. To avoid priming and experimenter demand effects, we use examples from another choice domain. Each example involves the same decision-maker, Chris, who must choose between attending two social events; we vary the reason he questions his choice. Critically, we define all eight variants of low decision confidence and provide all eight examples before asking the subject to reflect on their own decisions. This procedure allows subjects to identify the examples their experience resembles most closely and rate the associated categories accordingly.

A concrete illustration underscores the collective impact of these design features. Consider a subject for whom "true" risk preferences exist, but who expresses low confidence when choosing a lottery over a safe alternative because they have difficulty bringing those preferences into sharp focus. This possibility is an example of explanation 3. The research cited above implies that, when categorizing the source of their low confidence, the subject would look for the best fit among the various explanations based on the conceptual definitions and especially

the associated examples. Our description of explanation 3 reads: *Didn't Think Carefully About What You Want: You lack confidence in your decision because you haven't thought through what you want to achieve as carefully as you could.* The subject gains an operational understanding of this definition from the associated example: *Chris values experiences with different friends for different reasons but hasn't seriously considered whether one set of reasons is more important than another. He might express less-than-complete confidence in his decision because he hasn't carefully thought through what he wants from this social event.* No other description or example references difficulties bringing preferences into focus. The next closest fit is likely explanation 5 (normative ambiguity), but both the description and the example highlight the individual's belief that there is more than "one right way to think about" their objectives, which rules out the existence of the "true" objectives our hypothesized subject seeks to apprehend.

Certainly, potential sources of confusion remain. Consider explanation 7: The associated description and example posit that the individual experiences low confidence because they have *chosen* an option for which the outcome is unpredictable given the available information. There is no suggestion that the individual experiences low confidence due to the *mere presence* of a risky alternative when the chosen option is safe. However, if our hypothesized subject misses this point, they might interpret explanation 7 as *implicitly* encompassing the possibility that low confidence results from general difficulties with bringing risk preferences into focus. Even then, explanation 3 may match their experience more closely because it *explicitly* highlights difficulties apprehending "true" preferences. In practice, as we mention below, high ratings for explanation 7 are indeed strongly correlated with selection of the risky option, which is consistent with the category's intended meaning, and inconsistent with explanation 3.

Recognizing that it is impossible to convey the boundaries between the various explanations for low decision confidence with formal precision, we allow subjects to express that uncertainty by eliciting a rating for each category rather than a single categorical selection. These ratings reflect the relative importance of each reason for expressing low confidence on a scale from 1 (Very Little) to 7 (Very Much). When analyzing responses to these questions, we treat the top-rated variants as the ones that offer the best matches between the individual's subjective experience and categorical concepts, including the associated prototypical examples.

To evaluate whether subjects understand the eight explanations, we require them to complete a four-question classification test. Each question describes the thought process that leads a decision-maker to express low confidence in a decision. The four decisions all fall within different choice domains. In each case, we list three of the eight potential explanations and ask the subject to classify the example. Subjects receive an additional bonus of $1 if they correctly answer three of the four questions. Critically, because this exercise tests whether subjects understand the rationales well enough to correctly classify instances of less-than-complete

confidence, it provides a sound basis for evaluating whether they accurately classify their own explanations. Our design proved effective: 83% of our subjects answered at least three of the four classification questions correctly.[32]

The accuracy of any method for categorizing subjective motives is of course open to question. However, it is worth bearing in mind that reservations concerning such methods preclude one from ruling out explanations 5 through 8, and hence argue *against* using expressions of decision confidence as a criterion for diagnosing improvability.

Implementation details. Because the new battery of questions is time-consuming, we present each subject with ten randomly selected tasks—five Risky and five Risk-Free—rather than twenty. After subjects make all ten choices and report their associated confidence ratings, we select the three Risky tasks and the three Risk-Free tasks for which they express the least confidence.[33] Then we ask subjects to rate the importance of the eight reasons for expressing less-than-complete decision confidence, as well as a catch-all explanation labeled "Other," on a 1-to-7 scale. We randomize the order of the eight reasons between subjects.

Experiment 2 involved 600 new subjects.[34] Appendices B.1.1 to B.1.6 show screenshots of the instructions, including the prototypical examples and test questions.

6.2 Results

Our analysis of Experiment 2 focuses on the set of tasks for which we asked subjects to explain their less-than-complete decision confidence (henceforth, the *confidence rationalization* sample). We are primarily concerned with the explanations they offer.[35] For each task in this sample, we identify the explanation(s) for less-than-complete confidence the subject rated as most important. Then we calculate the frequency with which each explanation receives the top rating across all such tasks.[36] Figure 6 presents results separately for the Risky and Risk-Free tasks and for the Pre- and Post-Choice conditions. Because our results for the Pre- and

---

[32]The odds of doing so purely by chance are one in nine. Our analysis is based on data from all subjects regardless of their test performance. Restricting the analysis to those who answered at least three of the four test questions correctly does not alter our conclusions; see Appendix A.6.4.

[33]We only select tasks for which the subject reports a confidence level below 100. Overall, 75% of Risky tasks and 47% of Risk-Free tasks satisfy this condition. If fewer than three Risky tasks meet this criterion for a given subject, then we select fewer Risky tasks without adding Risk-Free tasks. We proceed similarly if fewer than three Risk-Free tasks meet this criterion. If the subject reports maximal confidence in all tasks, then we skip the elicitation of reasons for less-than-complete confidence.

[34]The median completion time was 46.4 minutes for the Pre-Choice Unincentivized condition, 38.9 minutes for the Post-Choice condition, and 42.9 minutes overall. Subjects received a guaranteed payment of $8 and had a 20% chance of being selected for a bonus payment.

[35]We analyze patterns of Risky and Risk-Free choices from Experiment 2 in Appendix A.6.1.

[36]In the event of ties, we treat all of the highest-rated explanations as "most important." Consequently, these fractions sum to more than 100%.

Post-Choice conditions are similar, we group them together for the purpose of this discussion.
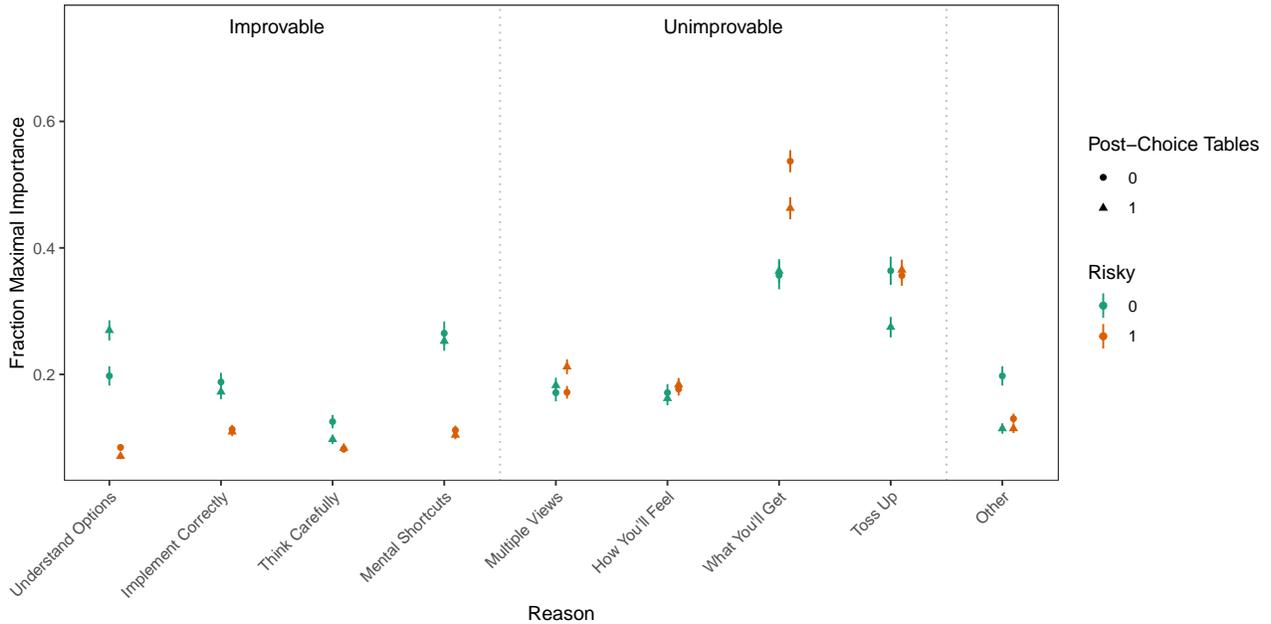
Figure 6: Experiment 2 Confidence Decomposition

Notes: Figure 6 shows the frequencies with which explanations for less-than-complete confidence are rated most important by subjects in the confidence rationalization sample of Experiment 2. Frequencies are reported for Risky choices (orange, $N = 1,547$) and Risk-Free choices (green, $N = 1,001$). The subsamples are unbalanced because subjects more often express complete confidence in Risk-Free than in Risky choices. Results are shown separately for Post-Choice (triangles, $N = 1,364$) and Pre-Choice (circles, $N = 1,184$) conditions. Bars represent ± 1.96 standard errors (95% CIs).

In the Risky tasks, subjects overwhelmingly emphasize explanations for their less-than-complete confidence that do not imply perceived improvability. Aggregating across explanations 1 through 4, we find that subjects' top-ranked explanations imply perceived improvability for only 22% of Risky choices. In contrast, aggregating across explanations 5 through 8, we find that subjects' top-ranked explanations do not imply perceived improvability for 87% of those choices. Indeed, for 50% of choices, riskiness of the chosen option (explanation 7) is the most important explanation. This finding potentially explains the strong correlation between confidence and choice of the safe option in Figure 4, especially for Unmixed tasks: Choosing the risky option creates ex-post uncertainty, which leads to expressions of low confidence. Even allowing for residual confusion concerning the eight categories, these findings provide grounds for doubting the reliability of elicited decision confidence as a measure of perceived improvability.

Given the high frequency with which subjects attribute low confidence in Risky choices to explanation 7, it is important to verify that they interpret the explanation as intended. Because the associated description and example reference riskiness of the *chosen* option, we would expect subjects to express low confidence for this reason significantly more often when they

select the risky option than when they select the safe option. Consistent with this prediction, we estimate that, for Unmixed Risky tasks, the fraction who experience less-than-complete confidence and associate it primarily with riskiness of the chosen option is 50% when choosing the risky option versus only 26% when choosing the safe option.[37] While a non-zero frequency for those who choose the safe option potentially reflects misunderstandings, there are other possible explanations: When answering questions about reasons for low confidence, subjects may forget which option they chose and overlook our reminder; they may be uncertain about the emotional consequences of selecting the safe option such as experiencing regret; they may be uncertain about the monetary consequences of selecting the safe option due to concerns about experimenter reliability; or they give noisy answers due to fatigue and limited attentiveness.

Significantly, subjects offer much different explanations for less-than-complete confidence in Risk-Free choices. Those associated with perceived improvability (1 through 4) receive the highest ratings for 52% of Risk-Free choices in the confidence rationalization sample—in other words, roughly two-and-a-half times as frequently as for Risky choices. This difference is primarily attributable to the importance of explanation 1 (limited comprehension of the available options) and explanation 4 (use of imperfect mental shortcuts), both of which subjects rate as most important for 20 to 30 percent of these choices.

The contrast between the distributions of explanations for low confidence in Risky and Risk-Free choices has two implications. First, the meaning of responses to broad questions about decision confidence depends on the choice domain. This instability suggests that diagnostic methods based on such responses may not be portable across domains. Second, considering our results for Risk-Free choices, the low frequency with which responses in the Risky domain indicate perceived improvability is not a consequence of some general bias in our design.

Variation in subjects' interpretation of the general confidence question across domains po-tentially explains why, in the main experiment, the correlation between confidence and the proclivity to choose the safe/"safe" option is positive and substantial in Risky tasks but essen-tially zero in Risk-Free tasks; compare Figure 4 with Appendix Figure A3. The same figures show that, while confidence is uncorrelated with expected value maximization in Risky tasks,

---

[37]To construct these estimates, we multiplied the fraction of Unmixed Risky choices for which the subject chose the risky (respectively, safe) option and expressed less-than-complete confidence by the fraction of Unmixed Risky choices in the confidence rationalization subsample for which the subject chose the risky (respectively, safe) option and gave the highest rating to explanation 7. This calculation assumes that the confidence rationalization sample is representative with respect to reasons for low confidence. While this sample is selected based on the subject's degree of low confidence, we see no reason to expect that this selection biases our estimate of the relative frequencies. Our finding adds to existing evidence from cognitive psychology on ex-post uncertainty and confidence: da Silva Castanheira et al. (2021) also find that confidence is higher in risky choices when the chosen alternative is a deterministic amount rather than a risky lottery. They reference accumulating evidence that "confidence ratings, to varying extents, also incorporate a sense of certainty (or precision) about choice-relevant variables, over and above a perceived probability of being correct" (p. 1021).

it is highly correlated with value maximization in Risk-Free tasks. We discuss the relationship between confidence, choice, and characterization among Risk-Free tasks in Appendix A.3.

6.3 Robustness with respect to wording

Next, we ask whether our results are robust with respect to the wording of the confidence question. A natural alternative is to use the following question from Enke et al. (2024): "How certain are you (in %) that choosing [X] is actually your best decision, given your preferences and the available information?" Compared with our question, this alternative adds technical precision which subjects may or may not appreciate. In principle, it may focus their attention more effectively on evaluating ex ante decision quality. To investigate this possibility, we conducted an additional experiment with 349 subjects—Experiment 3—that replicated Experiment 2 using the decision confidence question for binary decisions in subjective tasks from Enke et al. (2024). We also included the comprehension check used by Enke et al. (2024) to assess subjects' understanding of the intended meaning of the confidence question.[38]

Experiment 3 essentially replicates our findings from Experiment 2. Even with the alternative confidence question, subjects' top-ranked explanations imply perceived improvability for only 38% of Risky choices in the confidence rationalization sample. In contrast, subjects' top-ranked explanations do not imply perceived improvability for 81% of those choices. We likewise find that the alternative measure of overall decision confidence is strongly correlated with choosing the safe option in Unmixed Risky tasks ($\rho = 0.21$, $p < 0.01$), and is substantially less correlated with choosing the safe option in Mixed tasks ($\rho = 0.02$, $p = 0.46$). Additionally, confidence is uncorrelated with maximizing expected value ($\rho = 0.001$, $p = 0.97$ for Unmixed tasks, $\rho = 0.002$, $p = 0.93$ for Mixed tasks). See Appendix A.7 for details.

The possibility remains that further refinement of the DC method might yield a reliable diagnostic for perceived improvability. Building on our methods, one might attempt to detect perceived improvability by posing questions that seek to identify the source of low confidence. For example, it may be appropriate to exclude choices from the Welfare-Relevant Domain when people express low confidence in their understanding of the available options and their consequences (i.e., self-reported characterization failure). We are, however, unsure that questions about more subtle aspects of decision processes would prove sufficiently reliable given the inherent difficulty of ensuring that subjects consistently interpret them as intended.

---

[38]We recruited 349 new participants for this study. Each participant received a $7.50 completion payment and each participant had a 20% chance to be randomly selected to receive a bonus payment as in our main study. The median completion time for Experiment 3 was 57.3 minutes. A full set of instructions can be found in Appendices B.1.5, B.1.7 and B.1.8.

7 Does pattern matching (PM) reliably diagnose improvability?

Does pattern matching between behavior in a primary domain and objective errors in a matching domain imply that the primary-domain choices are improvable? The answer depends on the mechanism that generates the matching-domain errors and its relevance for the primary-domain choices. Consequently, we propose evaluating any given application of the Pattern Matching method through the following sequence of steps: (1) Identify potential sources of improvability in the matching domain (here, Risk-Free choices)—i.e., specific instantiations of characterization failure or optimization failure—that may account for the matching-domain patterns; (2) for each potential source of matching-domain improvability identified in step 1, identify compatible explanations for parallel patterns in the primary domain (here, Risky choices); (3) for each explanation of the patterns' existence in the primary domain identified in step 2, determine the implications for the improvability of primary domain choices; and (4) determine which of the resulting explanations for pattern matching the data support.

Table 2 summarizes the analysis presented in this section. The first four columns correspond to the four steps listed above. We perform the first three steps in Section 7.1. The first two steps generate a list of hypotheses concerning the reasons for pattern matching, and the third step clarifies the implications of these hypotheses for the improvability of Risky choices. In subsequent subsections, we discuss evidence that speaks to these hypotheses (Step 4).

7.1 Hypotheses

Table 2 lists the following four potential sources of improvability in the Risk-Free domain.

I. Subjects do not understand Risk-Free tasks. If misunderstandings of Risk-Free tasks lead subjects to commit systematic errors, then why might the same patterns appear across Risky and Risk-Free tasks? Two possibilities merit consideration.

First, subjects may mistakenly interpret Risk-Free tasks as involving risk (hypothesis Ia in Table 2). This classification error is a special type of characterization failure, which we call "recognition failure." The PM method inherently introduces the potential for recognition failure because it requires the analyst to design matching tasks that closely resemble the primary tasks. Recognition failure could naturally lead to pattern matching by inducing subjects to deploy their risk preferences in both domains. However, in that case, the resulting parallelism would *not* imply that choices in Risky tasks are improvable.

Second, subjects may misinterpret Risk-Free and Risky tasks in a similar manner (hypothesis Ib in Table 2). For example, someone might fail to understand the significance of p in either

Table 2: Evaluating the Pattern Matching Method

1. Possible Source of Risk-Free Task Improvability	2. Explanation for Pattern Matching	3. Implication of Risky Task Improvability?	4. Support in the Data?	Key Facts

I. Subjects do not understand Risk-Free tasks	a. Because the PM method requires Risk-Free tasks that resemble Risky tasks, subjects confuse the two; recognition failure	No	Yes	Pattern matching is:
				-reduced by "enhanced recognition" treatments (7.2)
				-driven by recognition failures in Risk-Free tasks (7.3)
				-related to comprehension errors (7.4)
				-prevalent among subjects who:
				  view calculations as unnecessary in Risk-Free (7.5);
				  report riskiness in Risk-Free shortcuts (7.5)

	b. Subjects have similar misunderstandings about Risky tasks	Yes	No	Domains are cognitively dissimilar (7.5) Stronger patterns in Risky tasks when characterization is perfect (7.3) No error in Risky parallels recognition failure in Risk-Free (7.3)
II. Subjects understand Risk-Free tasks but knowingly take shortcuts	a. Subjects use similar shortcuts in both domains.	Yes	No	Domains are cognitively dissimilar (7.5) Limited overlap between domains regarding (i) avoidance of calcs. viewed as necessary (ii) use of shortcuts (7.5)
	b. Subjects deploy Risky task decision criteria as Risk-Free task shortcut	No	Possible	Robust patterns in Risk-Free only when report (i) or (ii) (7.5) Robust patterns in Risky whether or not report (i) or (ii) (7.5) When calculation is enforced, deviations from calcs. yield the systematic errors in Risk-Free choices (7.6)
III. Subjects understand Risk-Free tasks and try to perform them, but make mistakes	a. Task performance in Risk-Free and Risky tasks involves cognitively similar operations	Yes	No	Domains are cognitively dissimilar (7.5) Little pattern matching when use calcs. in Risk-Free (7.5) Risk-Free calcs. are generally accurate and unbiased (7.6)
IV. Subjects understand Risk-Free tasks but are uncertain about the accuracy of their calculations	a. Subjects deploy their risk preferences in both domains	No	Possible	When calculation is enforced, deviations from calcs. yield the systematic errors in Risk-Free choices (7.6) Subjects penalize options that call for calcs. (7.6)

domain and proceed as if the dollar amounts associated with each option were either equally likely (for Risky choices) or equally weighted (for Risk-Free choices). Similar misconceptions could give rise to similar patterns. In this case, pattern matching would imply that Risky choices are also improvable.

II. Subjects understand Risk-Free tasks, but take shortcuts. If subjects understand Risk-Free tasks but knowingly take shortcuts, then why might the same patterns appear in Risky and Risk-Free tasks? There are two main possibilities.

First, subjects might use similar shortcuts in Risk-Free and Risky tasks (hypothesis IIa in Table 2). For example, a subject who is risk-neutral with respect to small-stakes gambles, and who thinks that the ideal decision strategy involves calculating a lottery's expected payoff, might intentionally approximate that calculation the same way in both domains. More generally, subjects may deploy comparable simplifications in the two settings if the processes of "setting up" the tasks are cognitively similar. In such cases, improvability in the Risk-Free domain would likely imply improvability in the Risky domain.

Second, subjects may deploy their Risky decision criteria as Risk-Free shortcuts (hypothesis IIb in Table 2). For example, subjects who have easy "mental access" to preferences over lotteries and who dislike mathematical calculations may knowingly decide to evaluate Risk-Free options as if they were lotteries. In such cases, improvability in the Risk-Free domain does not imply improvability in the Risky domain.

III. Subjects understand Risk-Free tasks and try to perform them, but make mistakes. If subjects understand Risk-Free tasks and try to perform them but make mistakes, then why might the same patterns appear in Risky and Risk-Free tasks? We are aware of only one reasonable possibility: Properly understood Risky and Risk-Free tasks may involve cognitively similar operations that trigger similar systematic errors (hypothesis IIIa in Table 2). For example, a subject who is risk-neutral with respect to small-stakes gambles, and who evaluates lotteries by calculating their expected payoffs, might make the same systematic errors when attempting to calculate the mean accurately in both domains. In such cases, improvability in the Risk-Free domain would imply improvability in the Risky domain.

IV. Subjects understand Risk-Free tasks and try to perform them, but are uncertain about the accuracy of their calculations. If subjects understand the Risk-Free tasks but experience subjective uncertainty about the accuracy of their calculations, then why might the same patterns appear in Risky and Risk-Free tasks? Aversion to subjective uncertainty provides a possible explanation for "risk aversion" in Unmixed Risk-Free tasks because, by construction, the "Risky" option requires a calculation, which introduces uncertainty under

this hypothesis, while the "Safe" option does not. Subjects may exhibit the "CRE" because, in Mixed tasks, both options require calculations, which means aversion to uncertainty may skew choices toward the "Safe" option to a smaller degree. For Risky tasks, these same patterns may simply be features of risk preferences. Consequently, under this hypothesis, improvability in the Risk-Free domain would not imply improvability in the Risky domain.

Next, we present evidence that addresses these hypotheses. Since some types of evidence speak to multiple hypotheses, we organize our discussion around the evidence rather than around hypotheses. Our key findings, summarized in Table 2, point to recognition failure in Risk-Free tasks as the primary explanation for pattern matching (hypothesis Ia). For Risky tasks, risk aversion and the CRE are highly robust across all treatments and subgroups. In contrast, for Risk-Free tasks, we primarily observe parallel patterns when the design does not highlight the absence of risk (Section 7.2), as well as among subjects who incorrectly complete characterization tables (Section 7.3), answer comprehension questions incorrectly (Section 7.4), claim incorrectly that the tasks do not require calculations (Section 7.5), and report using self-described shortcuts that reference risk (Section 7.5). Evidence involving self-reported cognitive strategies provides more limited support for two other hypotheses (Sections 7.5 and 7.6): aversion to subjective uncertainty concerning the accuracy of calculations (hypothesis IVa), and the use of risk preferences in Risk-Free tasks as a shortcut to avoid performing calculations (hypothesis IIb). The data fail to support, and sometimes contradict, the three remaining hypotheses, which hold that pattern matching reflects the deployment of similar shortcuts (hypothesis IIa), similar calculation errors (hypothesis IIIa), or similar misconceptions about the nature of Risky and Risk-Free choices (hypothesis Ib). Importantly, none of the supported hypotheses imply that classical risk-taking patterns in Risky tasks are improvable. Thus, we conclude that the PM approach does not credibly diagnose improvable choices.

7.2   The robustness of Pattern Matching

We begin our empirical evaluation of the Pattern Matching method by investigating the robustness of the results shown in Figure 5, which are based on the Post-Choice condition of our main experiment. Finding that subjects exhibit less small-stakes "risk aversion" and a smaller "CRE" in Risk-Free tasks for treatments that highlight the absence of risk would support hypothesis Ia, that pattern matching is a consequence of recognition failure. Our Pre-Choice condition may serve as one such "enhanced recognition" treatment if it encourages subjects to think more carefully about the nature of each option at the time of choice.

Figure 7 replicates Figure 5 using data from the Pre-Choice condition. Visually, pattern matching is much less apparent in Figure 7 than in Figure 5. Lines 1 and 2 of Table 3 make

A   Risky Tasks, Pre Choice Tables   **B**   Risk–Free Tasks, Pre Choice Tables
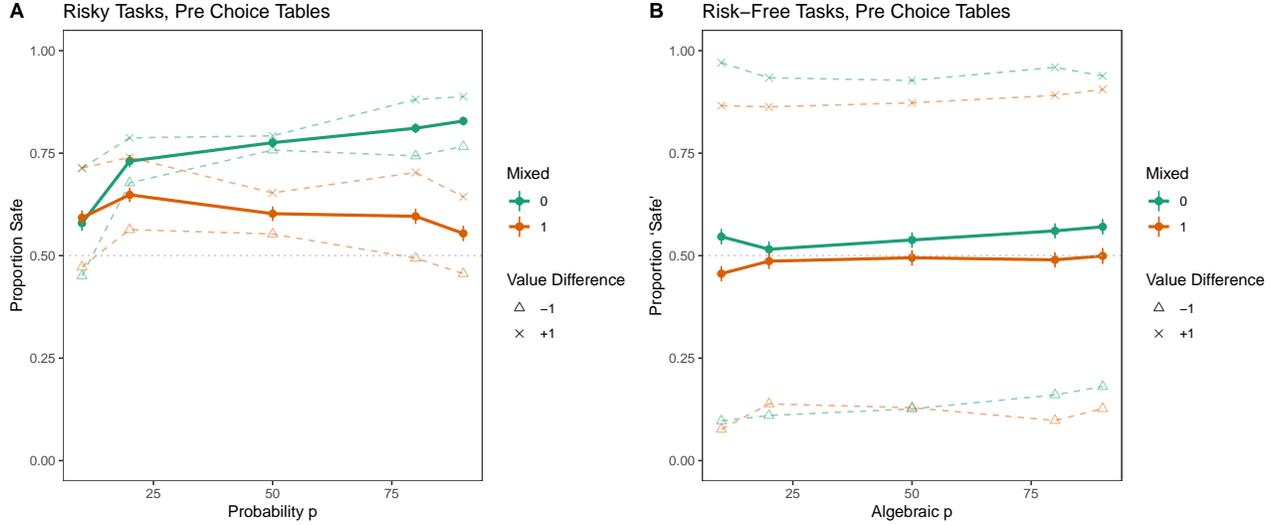
Figure 7: Pattern Matching Method: Risky and Risk-Free Tasks, Pre-Choice Condition

Notes: Figure 7 shows the proportion of choices for the Safe/"Safe" option in Unmixed tasks (green) and Mixed tasks (orange). Light-shaded exes and triangles denote $+1$ and -1 tasks, while dark-shaded circles indicate sample averages. Bars represent ± 1.96 standard errors (95% CIs). Panel A reports Risky tasks ($N = 6{,}674$), and Panel B reports Risk-Free tasks ($N = 6{,}646$) for subjects in the Pre-Choice condition.

the same point quantitatively. For the Risky domain, there is just as much evidence of risk aversion and the CRE in the Pre-Choice condition as in the Post-Choice condition. In contrast, for the Risk-Free domain, their counterparts are much weaker in the Pre-Choice condition, and the rate of maximization errors is much lower.[39]

The Pre-Choice intervention described above combats recognition failure by nudging subjects to think about their options more carefully. However, it might also implicate other hypotheses, for example by reducing subjects' use of shortcuts. We therefore fielded an additional experiment involving 300 new subjects that adopts a more direct approach by highlighting the absence of risk in Risk-Free tasks explicitly (Experiment 4). For each Risk-Free task, we prominently displayed the following text, highlighted in a red box: "Remember, this task DOES NOT INVOLVE ANY RISK. When a bag contains tickets of more than one color, all of them give the same payment." To further reduce the likelihood that subjects might conflate different types of tasks, we separated Risky and Risk-Free decisions into two blocks, with the Risky block always appearing first. We also omitted the characterization assessments.[40]

---

[39]Patterns also become less similar at the individual level. For the 3174 cases in the Pre-Choice conditions where a single subject completed a Risky task and the corresponding Risk-Free task, the raw correlation between maximization in the Risk-Free domain and expected value maximization in the Risky domain falls by about 50%, to 0.08.

[40]Each subject received a $4 completion payment. 20 percent of subjects, chosen at random, also received a bonus payment following the same procedures as in our main study. The median completion time for Experiment 4 was 17.1 minutes. A full set of instructions appears in Appendices B.1.9 through B.1.12.

Table 3: Behavioral Patterns by Experiment, Condition, & Subgroup

Sample	Risky tasks		Risk-Free tasks		
	Unmixed Safe	CRE	Unmixed "Safe"	"CRE"	Max. Error
Experiment 1					
1. Post-Choice	0.76 (0.01)	0.17 (0.02)	0.64 (0.01)	0.14 (0.02)	0.24 (0.01)
2. Pre-Choice	0.74 (0.01)	0.15 (0.01)	0.55 (0.01)	0.06 (0.01)	0.11 (0.00)
3. Pre-Choice Recognition in RF?					
3a. Recog. Fail.	-	-	0.71 (0.02)	0.23 (0.03)	0.35 (0.02)
3b. No Recog. Fail.	-	-	0.53 (0.01)	0.04 (0.01)	0.07 (0.00)
3c. CA Error & No Recog. Fail.	-	-	0.55 (0.03)	0.07 (0.04)	0.33 (0.02)
4. CA Error & No Recog. Fail. (Sim.)	-	-	0.52 (0.03)	0.01 (0.04)	0.33 (0.02)
5. Pre-Choice Comprehension errors?					
5a. None	0.77 (0.01)	0.16 (0.01)	0.52 (0.01)	0.04 (0.02)	0.06 (0.00)
5b. Some	0.71 (0.01)	0.13 (0.02)	0.58 (0.01)	0.09 (0.02)	0.17 (0.01)
Experiment 4					
6. Recog. Enhanced	0.77 (0.01)	0.14 (0.02)	0.56 (0.01)	0.08 (0.02)	0.20 (0.01)
7. Math calculations in RF?					
7a. Made calculations	0.78 (0.01)	0.14 (0.02)	0.55 (0.01)	0.06 (0.02)	0.14 (0.01)
7b. Not needed	0.69 (0.04)	0.11 (0.06)	0.58 (0.04)	0.10 (0.06)	0.50 (0.03)
7c. Too taxing	0.76 (0.04)	0.16 (0.06)	0.65 (0.05)	0.23 (0.07)	0.47 (0.03)
Experiment 5 (Block 1)					
8. Conditions					
8a. Baseline	0.78 (0.01)	0.14 (0.02)	0.60 (0.01)	0.04 (0.02)	0.31 (0.01)
8b. Recog. Enhanced	0.76 (0.01)	0.10 (0.02)	0.53 (0.01)	0.02 (0.02)	0.27 (0.01)
9. Comprehension errors?					
9a. None	0.81 (0.01)	0.16 (0.02)	0.49 (0.02)	0.00 (0.03)	0.08 (0.01)
9b. Some	0.75 (0.01)	0.10 (0.01)	0.58 (0.01)	0.04 (0.01)	0.34 (0.01)
10. Math calculations in RF?					
10a. Made calculations	0.79 (0.01)	0.14 (0.01)	0.54 (0.01)	0.02 (0.02)	0.21 (0.01)
10b. Not needed	0.72 (0.02)	0.04 (0.03)	0.64 (0.02)	0.05 (0.03)	0.47 (0.02)
10c. Too taxing	0.76 (0.02)	0.11 (0.03)	0.56 (0.03)	-0.01 (0.04)	0.45 (0.02)
11. Shortcuts?					
11a. Yes	0.78 (0.01)	0.13 (0.02)	0.61 (0.02)	0.08 (0.03)	0.36 (0.01)
11b. No	0.76 (0.01)	0.11 (0.01)	0.54 (0.01)	0.01 (0.02)	0.26 (0.01)
12. RF Shortcut cites risk?					
12a Yes	-	-	0.72 (0.03)	0.13 (0.04)	0.45 (0.02)
12b No	-	-	0.55 (0.02)	0.05 (0.03)	0.31 (0.01)

Notes: Table reports aggregate results across values of p and $+1/-1$ tasks for groups of observations separately for Risky and Risk-Free tasks (standard errors in parentheses). Unmixed Safe/"Safe" presents evidence of small-stakes risk/"risk" aversion (benchmark of 0.5). CRE/"CRE" presents evidence of the common ratio effect (benchmark of 0). Max. Error corresponds to the proportion of maximization errors in Risk-Free tasks. RF denotes Risk-Free tasks.

Results for Experiment 4, reported in line 6 of Table 3, resemble those for the Pre-Choice condition of Experiment 1 (line 2). Comparing line 6 to line 1, we see that, for the Risky domain, there is just as much evidence of risk aversion and the CRE as in the Post-Choice condition; in contrast, for the Risk-Free domain, there is much less evidence of "risk aversion" and the "CRE."[41] See Appendix A.8 for additional details, including Appendix Figure A12, which bears a striking similarity to Figure 7. Wu (2025) provides contemporaneous evidence involving a similar treatment that corroborates our finding.

Next, we address a potential objection: The format we use for depicting Risk-Free tasks may deactivate a mechanism that would otherwise produce pattern matching for reasons other than recognition failure. Specifically, as explained in Section 3.2, to accommodate our characterization assessments, we described Risk-Free options as involving the selection of a single colored ticket from a bag and expressed the associated payments, y and z, as algebraic formulas. If subjects struggle to "set up" decision problems in which they must perform calculations to determine the payoffs that certain options yield, then providing them with formulas may eliminate the need for shortcuts, thereby suppressing behavior associated with hypothesis II. At the same time, our stimuli may inadvertently encourage subjects to think of Risk-Free tasks as stochastic, thereby introducing behavior associated with hypothesis Ia.

To evaluate this potential concern, we fielded an additional experiment involving 1200 new subjects (Experiment 5). Following Oprea (2024), we use the same visual stimulus for both Risky and Risk-Free tasks. As in our main experiment, the stimulus depicts options as bags of colored tickets. Each color is associated with a label that states a dollar value. For Risky tasks, we explain that we will select a single ticket from the chosen bag and that the subject will receive the amount on the label. We call these "Select One Ticket" tasks. For Risk-Free tasks, we explain that we will select all the tickets from the chosen bag and that the subject will receive the total amount for all of the labels divided by 100. Importantly, in contrast to our main experiment, we do not reduce this calculation to a simple formula (i.e., frequency times payment summed over colors) on the decision screen. We call these "Select All Tickets" tasks. Following Oprea (2024), subjects completed the Risky and Risk-Free tasks in two randomly ordered blocks. We also administered Oprea's four-question comprehension quiz along with corrective feedback before both the Risky and Risk-Free blocks.[42] Notably, Oprea (2025) argues that these elements of his design mitigate recognition failures. In a separate "enhanced recognition" treatment, we also include a reminder for Risk-Free tasks similar to the

---

[41]The frequency of maximization errors remains high but the errors are much less systematic than in the Post-Choice condition.

[42]We adjusted one comprehension question: Instead of using a two-color example in which tickets of both colors were equally numerous, we used one in which tickets of one color were more numerous than tickets of the other color. Our objective was to catch potential errors associated with computing arithmetic means rather than weighted averages.

one we introduced in Experiment 4: "Remember, this task DOES NOT INVOLVE ANY RISK. When you choose an option, there's only one possible amount you can receive." Along with several other instructional elements that distinguish Risky and Risk-Free tasks, this reminder may further reduce the tendency for subjects to conflate tasks of different types.[43]

Results for first-block responses, reported in lines 8a and 8b of Table 3 for the baseline and enhanced recognition conditions, respectively, corroborate the preceding findings.[44] For the Risky domain, we see just as much evidence of risk aversion and the CRE as in the previous conditions (lines 1, 2, and 6). The baseline condition of Experiment 5 (line 8a) exhibits less pattern matching than the Post-Choice condition of Experiment 1 (line 1), but there is still significant "risk aversion" in the Risk-Free domain. For the enhanced recognition treatment (line 8b), we find no evidence of pattern matching: "Risk aversion" and the "CRE" are virtually absent (see Appendix A.9 for details). Thus, we again observe that treatments designed to avoid recognition failure significantly reduce apparent "risk aversion" in Risk-Free tasks.

It is unclear why this design does not yield a stronger "CRE" in Risk-Free tasks. Notably, however, the absence of a "CRE" does not imply that subjects maximize correctly: The frequency of Risk-Free maximization failures is *higher* in Experiment 5 than in our main experiment (compare lines 8a and 8b to lines 1 and 2). Possibly other types of errors mask the recognition failures that would otherwise produce the "CRE." Indeed, as we document below, even in experiments for which pattern matching is more limited (such as Experiment 5), Risk-Free maximization errors tend to produce the classical risk-taking patterns primarily among subjects for whom there is evidence of recognition failures.

Taken together, the preceding results demonstrate the empirical fragility of pattern matching in the context of Risky and Risk-Free choices. Design adjustments that enhance recognition of Risk-Free tasks substantially weaken patterns normally associated with risk-taking in Risk-Free choices. These results are consistent with hypothesis Ia, which attributes anomalous Risk-Free patterns to recognition failures. Yet even the designs that enhance recognition do not remove all Risk-Free anomalies. The residual small-stakes "risk aversion" and "CRE" may indicate that our conditions did not entirely eliminate recognition failure, or alternatively that some of the other hypotheses listed in Table 2 are also at work. Resolving that issue requires other types of evidence, which we present below.

---

[43]Each treatment condition in Experiment 5 involved 600 subjects. Each subject received a $5 completion payment. Twenty percent of subjects, chosen at random, also received a bonus payment following the same procedures as in our main study. The median completion time, 21.9 minutes, was similar across the two conditions. The full instructions for Experiment 5 appear in Appendices B.1.13 through B.1.18.

[44]We focus on first block observations because our data show substantial evidence of order effects between blocks; see Appendix A.9.5.

7.3 Characterization assessments

The characterization assessments in the Pre-Choice conditions of our main experiment provide direct evidence concerning the prevalence and nature of characterization failures for both Risky and Risk-Free tasks. For Risk-Free tasks, *recognition* failure manifests as any non-degenerate mapping from states to outcomes.[45]

Characterization failures are present in 23% (1514 observations) of all Pre-Choice Risk-Free tables (Incentivized and Unincentivized). Of these, the majority (55%) are recognition failures.[46] As one might expect, for 77% of recognition failures in Unmixed Risk-Free tasks, the erroneous characterization indicates that there is one safe option and one risky option, and for 72% of recognition failures involving Mixed Risk-Free tasks, the erroneous characterization indicates that there are two risky options. Furthermore, options erroneously characterized as risky are not characterized as *almost* safe.[47]

Figure 8 shows the relationship between Risk-Free behavior and contemporaneously measured recognition failures in our Pre-Choice conditions. Recall that there is not much aggregate pattern matching in the Pre-Choice condition; however, Risk-Free choices with recognition failures exhibit the classical risk-taking patterns, while Risk-Free choices without such failures—even those with other types of characterization failures—deviate only modestly on average from choices that objectively maximize payoffs (see Table 3, lines 3a, 3b, and 3c). Thus, in addition to being the most common source of characterization failure, recognition failure drives residual risk-taking patterns in Risk-Free choice. This evidence supports hypothesis Ia.

Our characterization assessments also falsify hypothesis Ib, which posits that parallelism between the Risky and Risk-Free domains arises because subjects misinterpret the two types of tasks in similar ways. First, we can rule out misunderstandings for the 89% of Risky tasks that subjects characterized perfectly. And yet, as reported in Section 4.1, the patterns of interest are especially pronounced in those tasks. Second, as we have just demonstrated, in the Risk-Free

---

[45]A non-degenerate mapping may also reflect the use of a shortcut, such as treating Risk-Free options as Risky, or uncertainty about the accuracy of one's calculations. However, for subjects who understand that Risk-Free tasks objectively involve no risk, such reports are strictly dominated in the Pre-Choice Incentivized condition because they forfeit the bonus with certainty, whereas reporting a degenerate mapping does not.

[46]Completed Risk-Free characterization tables imply one of three structures: (1) two safe options (the correct interpretation, even if payoffs are miscalculated); (2) one safe and one risky option (recognition failure via conflation with Unmixed Risky tasks); or (3) two risky options (recognition failure via conflation with Mixed Risky tasks). We infer that a subject considers an option to be safe (risky) if, according to the table, the payoff's standard deviation is below (above) 0.001.

[47]See Appendix Figure A1 for examples of incorrectly completed characterization tables with and without recognition failures. We find little support for the hypothesis that these recognition failures, and characterization failures more generally, arise from inattentive subjects answering quickly without serious thought. For Risk-Free tasks, subjects spent more time on tables they characterized incorrectly than on those they characterized correctly (89.7 seconds vs. 69.3 seconds, $t = 6.9$, $p < 0.01$). They also spent more time on tables exhibiting recognition failures than on those without such failures (92.9 seconds vs. 71.2 seconds, $t = 6.8$, $p < 0.01$).
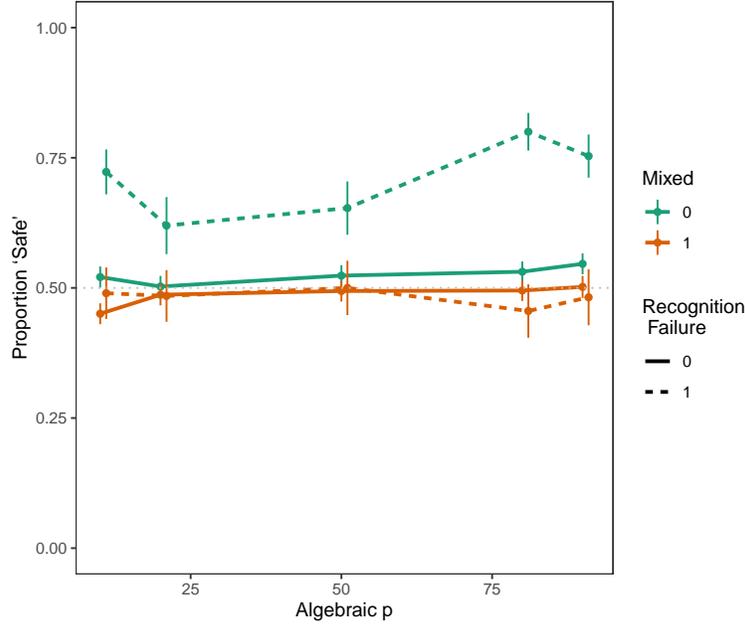
Figure 8: Risk-Free Tasks and Recognition Failures, Pre-Choice Condition

Notes: Figure 8 shows the proportion of choices for the "Safe" option in Unmixed tasks (green) and Mixed tasks (orange). Bars represent ±1.96 standard errors (95% CIs). Solid and dashed series indicate sample averages without and with recognition failures, respectively, in the corresponding tables.

domain, the patterns are most prominent among tasks subjects mistakenly characterized as Risky. There is no parallel error for Risky choices: Characterizing a Risky choice as Risky is not a mistake. In principle, symmetric frequencies of recognition failure for Risky and Risk-Free choices—i.e., characterizing Risky choices as Risk-Free and Risk-Free choices as Risky—could homogenize average behavior across domains, especially if both frequencies were 50%. However, recognition failures occur in only 1% of Risky tasks and 12% of Risk-Free tasks.

In Section 7.2, we noted a potential objection to the format used in our main experiment for depicting Risk-Free tasks: It may inadvertently suppress a mechanism that would otherwise generate pattern matching for reasons unrelated to recognition failure. The same objection could apply here. Experiment 5 overcomes this objection, but it does so by adopting a format unsuited to our characterization assessments. As a result, parallel analyses of Experiment 5 must rely on other indicators of recognition failure such as comprehension questions and open-ended descriptions of cognitive strategies. We consider that evidence next.

7.4 Comprehension questions

As noted previously, Experiment 5 includes comprehension checks similar to those of Oprea (2024) prior to decision making. Across baseline and enhanced recognition conditions, 75% of

subjects answer at least one comprehension question incorrectly, and the error rate is substantially higher for Risk-Free tasks than for Risky tasks.[48] Our main experiment also includes comprehension checks with corrective feedback for sample characterization tables. Similar conclusions follow: In our Pre-Choice condition, 58% of subjects answer at least one comprehension question incorrectly, and the error rate is substantially higher for the Risk-Free domain.[49] Critically, in both experiments, maximization failures in Risk-Free tasks are 3 to 4 times as common for subjects with comprehension errors than for those without; moreover, evidence of classical risk-taking patterns in Risk-Free tasks is largely confined to subjects with comprehension errors (see lines 5a and 5b, 9a and 9b of Table 3). These findings are consistent with hypothesis Ia.

Banki et al. (2025) conduct a similar analysis of comprehension quiz errors and choice patterns using data from Oprea (2024).[50] Results from their paper and our experiments are mutually corroborative: Risk-Free choices that exhibit classical risk-taking patterns are disproportionately attributable to subjects who make mistakes in the comprehension quiz, consistent with recognition failure. In responding to Banki et al. (2025), Oprea (2025) offers two arguments along with various supporting analyses. First, he attributes the correlations between comprehension quiz errors and pattern matching to characteristics such as inattentiveness that predispose subjects to take common shortcuts in all tasks. Second, he disputes the inference that comprehension quiz errors are associated with subsequent recognition failures in Risk-Free tasks, noting that accuracy improves as subjects proceed through the four questions. From this pattern, he infers that feedback corrects the pertinent misconceptions.

To evaluate the first argument using our data, we focus on the behavioral patterns that emerge in *Risky* tasks. If the argument were correct, we should likewise find much less evidence of classical risk-taking patterns in Risky tasks for subjects without comprehension errors than for those with comprehension errors. Yet in both experiments, those patterns are a bit *stronger* for subjects without comprehension errors, not weaker (see Table 3 lines 5a and 5b, 9a and 9b). This finding implies that comprehension errors do not simply identify inattentive respondents who deploy parallel shortcuts in both domains.[51]

To evaluate the second argument, we ask whether comprehension quiz errors predict subse-

---

[48]Conditional on making at least one comprehension error, 95% (respectively, 49%) of subjects answer at least one question about Risk-Free (respectively, Risky) tasks incorrectly. The enhanced recognition treatment reduces the overall frequency of making at least one comprehension error from 80% to 71% ($t = 4.9$, $p < 0.01$); for questions about Risk-Free tasks, the frequency falls from 76% to 68% ($t = 4.9$, $p < 0.01$).

[49]Conditional on having at least one comprehension error, 80% of subjects answer at least one question about tables associated with Risk-Free tasks incorrectly, versus 44% for tables associated with Risky tasks.

[50]As Banki et al. (2025) acknowledge, we independently and contemporaneously observed that classical risk-taking patterns in the Risk-Free tasks of Oprea (2024) are more pronounced among subjects with comprehension quiz errors. To avoid overlap, we do not report those findings.

[51]We note that Oprea (2025) finds somewhat stronger evidence of classical risk-taking patterns in Risky tasks for subjects with comprehension errors.

quent characterization failures in our main experiment. If error-related feedback is corrective as Oprea (2025) posits, then quiz errors pertaining to Risk-Free tasks should not be good predictors of measured characterization failure for Risk-Free choices. On the contrary, we find that they are highly predictive. Indeed, comprehension errors are associated with the types of subsequent recognition failures that produce the classical risk-taking patterns in Risk-Free tasks.Thus, in our data, errors in the comprehension quiz correlate with subsequent characterization failures even when subjects receive corrective feedback. Appendix A.5 provides further details.[52]

7.5   Self-reported cognitive strategies

The various explanations for pattern matching listed in Table 2 involve different assumptions concerning underlying cognitive processes. The next portion of our analysis seeks to illuminate those processes through an examination of responses to questions about the strategies subjects deployed, or believed they should have deployed, to make their selections.

Use of mathematical calculations.   After each block of tasks in Experiment 4—which reduced recognition failure through explicit reminders that Risk-Free tasks involve no risk—we asked the following question: "When you were making your decisions in Block X, did you make any mathematical calculations to arrive at your choices? By 'mathematical calculation' we mean any use of mathematical operations such as addition, subtraction, multiplication, or division." In the Risk-Free domain, 82% of subjects said they made mathematical calculations, compared with only 48% in the Risky domain. We followed up by asking those who said they performed calculations to provide details. We then categorized their responses according to whether they referenced an actual mathematical calculation. To minimize subjectivity, we prompted ChatGPT-5 Pro to perform these categorizations. More than 93% of responses in the Risk-Free tasks referenced a calculation, compared with only 57% in the Risky tasks.[53]

In light of the potential objection to the design of our main experiment mentioned in Section 7.2—that our format for depicting Risk-Free tasks may deactivate a pertinent pattern-matching mechanism—we asked the same question about the use of mathematical calculations after each

---

[52]As noted previously, our main experiment involves a different presentation of Risk-Free tasks and a different quiz than Oprea's. Experiment 5 follows his analysis more closely but uses stimuli that are unsuited to our characterization assessments. Consequently, we cannot replicate our analysis of the predictive relationship between quiz errors and subsequent characterization failures using those data. However, as discussed in Section 7.5, subjects who incorrectly answer quiz questions about Risk-Free tasks in Experiment 5 disproportionately report using cognitive strategies that reference risk when making Risk-Free choices. This evidence is consistent with our finding from Experiment 1.

[53]For Risky tasks, subjects who reported using mathematical calculations often described purely qualitative criteria rather than actual calculations. Appendix B.2.1 reproduces the prompt supplied to ChatGPT-5 Pro.

task block in Experiment 5.[54] In Block 1 Risk-Free tasks, 70% of subjects said they made calculations, compared with 50% in Block 1 Risky tasks.

The tendency to report using mathematical calculations far more often in the Risk-Free domain than in the Risky domain casts doubt on the premise, implicit in hypotheses Ib, IIa and IIIa, that subjects deploy similar cognitive processes in both domains. Moreover, in both Experiments 4 and 5, for subjects who reported making mathematical calculations in Risk-Free tasks, classical risk-taking patterns are weak in Risk-Free tasks but strong in Risky tasks (see Table 3, lines 7a and 10a). Consequently, one cannot attribute their risk-taking behavior to cognitive errors that their Risk-Free choices reveal.

Next, we ask whether the minority of subjects who said they did not make calculations in Risk-Free tasks exhibit pattern matching. We divide these subjects into two groups based on their answers to the following question concerning their avoidance of calculations: "Is this because mathematical calculations are not necessary for you to figure out which option you prefer? Or, would mathematical calculations have helped you figure out which option you prefer, but these calculations require too much time/effort/energy etc.? Or, was it for some other reason?"[55] For the Risk-Free domain, 10% of subjects in Experiment 4 and 18% in Experiment 5 said calculations were unnecessary. In contrast to those who said they made calculations, these subjects exhibit pattern matching in both experiments (see Table 3, lines 7b and 10b). However, their failure to acknowledge the relevance of calculations for Risk-Free choices is indicative of recognition failure. Consequently, results for this group are consistent with hypothesis Ia, which does not impugn Risky choices.

Those who said calculations are relevant but too taxing to perform constitute 7% of subjects in Experiment 4 and 10% in the Block 1 Risk-Free tasks of Experiment 5. These subjects exhibit pattern matching involving the classical risk-taking patterns only in Experiment 4 (see Table 3, lines 7c and 10c). The results from Experiment 4 are therefore consistent with both versions of hypothesis II concerning the use of shortcuts. However, only 18% of this group—just 1% of all subjects—gave the same response in both domains (i.e., that they did not make calculations because doing so was too taxing), again implying that the tasks are cognitively dissimilar, contrary to hypothesis IIa.[56] Hypothesis IIb—according to which calculation-averse subjects use their risky-choice decision criteria as shortcuts in the Risk-Free domain—is consistent with the emergence of pattern matching in the remaining 82% of this group.

---

[54] Augmenting Experiment 4, we provided examples of mathematical calculations from a different domain and asked subjects to complete comprehension questions related to the use of calculations.

[55] Only 1% of subjects in Experiment 4 and 2% in Experiment 5 cited some other reason.

[56] Another 50% reported not using calculations in the Risky domain because they were not relevant. In principle, members of this group might have used the same shortcut for Risky and Risk-Free choices (consistent with hypothesis IIa), despite recognizing the relevance of mathematical calculations for the latter.

Use of shortcuts. As part of Experiment 5, following the questions about mathematical calculations in each block, we explicitly asked subjects whether they used mental shortcuts. The instructions included a clarifying example and comprehension questions.

In first-block tasks, 26% of subjects report using shortcuts for Risk-Free choices versus 36% for Risky choices. Table 3 summarizes choice patterns separately according to whether or not subjects said they used shortcuts in the pertinent domain (lines 11a and 11b). In the Risk-Free domain, the classical risk-taking patterns are reasonably strong for shortcut users but nearly absent for non-users. This finding is directionally consistent with both versions of hypothesis II.[57] However, in the Risky domain, the classical risk-taking patterns are equally strong regardless of whether subjects said they used shortcuts. That finding contradicts hypothesis IIa, which attributes pattern matching to parallel shortcuts, but is consistent with hypothesis IIb.

We also examined open-ended responses describing the shortcuts subjects reported using in the Risk-Free tasks.[58] We used ChatGPT-5 Pro to identify subjects whose descriptions explicitly referenced risk.[59] According to these classifications, 43% of the descriptions referenced risk in the baseline condition, compared with 29% in the enhanced recognition condition.[60] Subjects whose descriptions referenced risk committed maximization failures more frequently and exhibited more pronounced classical risk-taking patterns in the Risk-Free tasks (see Table 3, lines 12a and 12b). These findings point toward hypotheses Ia (recognition failure), IIb (deployment of risk preferences as a shortcut), or IVa (deployment of risk preferences due to uncertainty of calculations), none of which imply perceived improvability for Risky choices. Consistent with hypothesis Ia and contrary to hypotheses IIb and IVa, risk-referencing subjects were also more likely than other shortcut users to have made at least one error in the comprehension questions for Risk-Free tasks (100% vs. 79%; $t = 5.0$, $p < 0.01$); see Appendix A.9 for details.

7.6 Ability to perform requisite calculations

As a final step in our empirical evaluation of the PM method, we ask whether subjects make systematic calculation errors when attempting to perform the requisite computations for Risk-Free tasks (which speaks to hypothesis IIIa), and whether they exhibit risk-taking patterns

---

[57]The finding does not rule out other explanations. For instance, as discussed below, there is suggestive evidence that those who reported using shortcuts in Risk-Free tasks experienced recognition failure.

[58]To aid recall when eliciting these descriptions, we reproduced one Unmixed task. Due to a coding error, roughly half of these subjects viewed a randomly selected Unmixed task they had not previously seen.

[59]We prompted ChatGPT-5 Pro to classify a description as referencing risk if it described the options as lotteries or indicated that subjective uncertainty about calculations influenced choices. The first possibility is consistent with recognition failure but could also reflect shortcuts. Appendix B.2.2 reproduces the prompt.

[60]For example, ChatGPT-5 Pro classified the following description of a Risk-Free task as referencing risk: "In Option A, there are no tickets labeled \$0, so I am guaranteed to get something. Option B, there are a significantly large number of tickets labeled \$0 and thus I am more likely to end up with nothing. Option A is more lucrative for me in the moment than the gamble."

in their Risk-Free choices because they are uncertain about the accuracy of their calculations (which speaks to hypothesis IVa).

First, we present evidence based on characterization assessments for Risk-Free tasks in the Pre-Choice condition of our main experiment. Conditional on recognizing these tasks to be Risk-Free, 88% of characterization tables involve no calculation errors.[61] Thus, calculation errors are relatively uncommon.[62] They are also largely unsystematic. To establish this point, we use subjects' calculated values to simulate the Risk-Free choices of hypothetical subjects who make the same errors and always choose the option with the higher calculated value. Among correctly recognized Risk-Free tasks with calculation errors, simulated choices reveal no anomalous patterns (see Table 3 line 4), contradicting hypothesis IIIa. Nevertheless, when we focus on actual choices for this same set of tasks, 14% of which differ from the simulated choices, we find modest small-stakes "risk aversion" and a larger "CRE" (Table 3 line 3c). This discrepancy between simulated and actual choices suggests that these patterns could be partly attributable to uncertainty about the accuracy of calculations, consistent with hypothesis IVa.

Because the preceding analysis uses the characterization assessments from our main experiment, it is necessarily conditional on successful task recognition. Furthermore, it assumes that, for correctly recognized tasks, incorrect entries in a characterization table reflect calculation errors rather than other possible mistakes such as failing to associate the label "Value" with the correct formula. In a supplemental experiment (Experiment 6), we therefore explicitly ask subjects to calculate the payments y and z in Risk-Free tasks. This feature allows us to study calculation errors directly and without conditioning on successful task recognition. We again find that calculation errors are not systematic. However, consistent with hypothesis IVa, departures from calculated-value maximization exhibit a modest but systematic bias toward the "Safe" option in Unmixed tasks as well as an apparent "CRE;" see Appendix A.10. This evidence could also be consistent with hypothesis Ia if subjects ignore the calculations and act on their risk preferences because they think the Risk-Free tasks are Risky. In contrast, hypothesis II (concerning shortcuts) cannot plausibly account for the anomalous Risk-Free choices.[63]

As a final step, we perform a targeted test of hypothesis IVa in another experiment with 375 additional subjects (Experiment 7).[64] Our objective is to determine whether the tendency to

---

[61]Additionally, conditional on recognizing the task to be Risk-Free, subjects select the option with the higher calculated value 95% of the time.

[62]Oprea (2024) arrives at a similar conclusion based on a condition requiring simpler calculations than those in his main experiment.

[63]Because this treatment instructs subjects to incur the cognitive costs of calculating y and z, it potentially removes the reason for deploying shortcuts. Even subjects who expend enough effort to perform the calculations correctly, which suggests they did not use shortcuts, exhibit the same patterns.

[64]Each subject received a $4 completion payment. Twenty percent of subjects, chosen at random, also received a bonus payment following the same procedures as in our main study. The median completion time for Experiment 7 was 18.9 minutes. A full set of instructions appears in Appendix B.1.19. Appendix A.11 provides

avoid options that require calculations reflects a causal relationship. The design parallels that of Experiment 6 for Unmixed tasks, except there are three variants of the Risk-Free tasks: one in which the payment appears as a number for Option A and as a formula for Option B (as before), one in which it appears as a formula for Option A and as a number for Option B, and one in which it appears as a formula for both. Most calculations are accurate (91% of two-calculation tasks and 93% of one-calculation tasks) and miscalculations are again unsystematic: Simulated subjects who always follow the calculations select Option A in 51% of tasks. However, in 9% of cases, subjects choose the option with the lower calculated value. The departures from calculated-value maximization exhibit a reasonably strong bias towards uncalculated options for tasks in which Option A has the smaller value: Subjects choose Option B 93% of the time when only Option A requires calculation, 87% of the time when both options require calculation, and 84% of the time when only Option B requires calculation ($t = 3.7$, $p < 0.01$ for equality of the first and second figures, $t = 2.5$, $p = 0.01$ for equality of the second and third). The effect is muted for tasks in which Option B has the smaller value: Subjects choose Option A 94% of the time when only Option B requires calculation, 95% of the time when both options require calculation, and 93% of the time when only Option A requires calculation; only the difference between the second and third figures is statistically significant ($t = 2.0$, $p = 0.05$).[65] These findings are consistent with hypothesis IVa, that subjects are averse to subjective uncertainty regarding the accuracy of their calculations. Under that hypothesis, pattern matching does not imply that Risky choices are improvable.

8   Discussion

Given the rising interest among policymakers in harnessing insights from Behavioral Economics and Psychology to help people make better decisions (see, e.g., Sunstein, 2020), there is an urgent need for diagnostic methods that can reliably detect when choices are improvable. This paper contributes to an emerging literature that seeks to replace ad hoc and impressionistic normative judgments with conceptually sound criteria that pass rigorous empirical tests.

We have evaluated three diagnostic methods: the Characterization Assessment (CA) method, the Decision Confidence (DC) method, and the Pattern Matching (PM) method. The first of these has become the leading tool for identifying improvable choices in Behavioral Public Economics. Its main shortcoming is that it identifies a limited range of decision-making

---

additional details and results.

[65]Notice that, when both options require calculation, the frequency with which subjects choose Option A when it yields the higher payoff (95%) is greater than the frequency with which they choose Option B when it yields the higher payoff (87%). A possible explanation is that, because Option A presents one ticket color while Option B presents two, some subjects suffer from recognition failure despite making the calculations. Subjects may also penalize options with multiple ticket colors for other reasons.

imperfections. The second and third methods are newer proposals; they are of interest because they cast broad nets and may therefore detect improvable choices that the CA method misses.

We find that the three methods have starkly contrasting implications concerning the normative significance of two widely documented risk-taking patterns: small-stakes risk aversion and the Common Ratio Effect. These disparate implications motivate a deeper dive into the validity of the assumptions each method implicitly invokes. We provide evidence that the CA method misses many instances of improvability, which underscores the potential value of broader tools. However, our analysis casts doubt on the reliability of the DC and PM methods. In our application, both classify many choices as improvable for reasons unrelated to improvability.

Future research could usefully build on our analysis by testing these three methods in other choice domains, developing more reliable refinements, and subjecting other improvability-detection methods to rigorous empirical tests.

References

Abaluck, J. and Adams-Prassl, A. (2021). What do consumers consider before they choose? identification from asymmetric demand responses*. *The Quarterly Journal of Economics*, 136(3):1611–1663.

Allcott, H., Lockwood, B. B., and Taubinsky, D. (2019). Regressive sin taxes, with an application to the optimal soda tax. *Quarterly Journal of Economics*, 134(3):1557–1626.

Allcott, H. and Taubinsky, D. (2015). Evaluating behaviorally motivated policy: Experimental evidence from the lightbulb market. *American Economic Review*, 105(8):2501–38.

Ambuehl, S., Bernheim, B. D., and Lusardi, A. (2022). Evaluating deliberative competence: A simple method with an application to financial choice. *American Economic Review*, 112(11):3584–3626.

Ashby, F. G. and Maddox, W. T. (2005). Human category learning. *Annu. Rev. Psychol.*, 56(1):149–178.

Banki, D., Simonsohn, U., Walatka, R., and Wu, G. (2025). Decisions under risk are decisions under complexity: Comment. *SSRN Working Paper*.

Barseghyan, L., Molinari, F., and Thirkettle, M. (2021). Discrete choice under risk with limited consideration. *American Economic Review*, 111(6):1972–2006.

Beck, I. L., McKeown, M. G., and Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction*. Guilford Press.

Bernheim, B. D. (2025). What is a mistake? guardrails for behavioral public policy. *Social Research: An International Quarterly*, 92:35–74.

Bernheim, B. D. and Rangel, A. (2004). Addiction and cue-triggered decision processes. *American Economic Review*, 94(5):1558–1590.

Bernheim, B. D. and Taubinsky, D. (2018). Behavioral public economics. *Handbook of behavioral economics: Applications and Foundations 1*, 1:381–516.

Bronnenberg, B. J., Dubé, J.-P., Gentzkow, M., and Shapiro, J. M. (2015). *The Quarterly Journal of Economics*, 130(4):1669–1726.

Butler, D. and Loomes, G. (1988). Decision difficulty and imprecise preferences. *Acta Psychologica*, 68(1):183–196.

Cason, T. N. and Plott, C. R. (2014). Misconceptions and game form recognition: Challenges to theories of revealed preference and framing. *Journal of Political Economy*, 122(6):1235–1270.

Chetty, R., Looney, A., and Kroft, K. (2009). Salience and taxation: Theory and evidence. *American Economic Review*, 99(4):1145–77.

da Silva Castanheira, K., Fleming, S. M., and Otto, A. R. (2021). Confidence in risky value-based choice. *Psychonomic Bulletin & Review*, 28:1021–1028.

Ellis, K., Filiz-Ozbay, E., and Ozbay, E. Y. (2024). Consistency and heterogeneity in consideration and choice. *Working Paper*.

Enke, B. and Graeber, T. (2023). Cognitive uncertainty. *Quarterly Journal of Economics*, 138(4):2021–2067.

Enke, B., Graeber, T., and Oprea, R. (2023). Confidence, self-selection, and bias in the aggregate. *American Economic Review*, 113(7):1933–1966.

Enke, B., Graeber, T., and Oprea, R. (2025). Complexity and time. *Journal of the European Economic Association*.

Enke, B., Graeber, T., Oprea, R., and Yang, J. (2024). Behavioral attenuation. Technical report, National Bureau of Economic Research.

Fleming, S. M. and Lau, H. C. (2014). How to measure metacognition. *Frontiers in human neuroscience*, 8:443.

Jin, S., Verhaeghen, P., and Rahnev, D. (2022). Across-subject correlation between confidence and accuracy: A meta-analysis of the confidence database. *Psychonomic bulletin & review*, 29(4):1405–1413.

Kurtz, K. J., Miao, C.-H., and Gentner, D. (2001). Learning by analogical bootstrapping. *The Journal of the Learning Sciences*, 10(4):417–446.

Maniscalco, B. and Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and cognition*, 21(1):422–430.

Martínez-Marquina, A., Niederle, M., and Vespa, E. (2019). Failures in contingent reasoning: The role of uncertainty. *American Economic Review*, 109:3437–3474.

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

Medin, D. L. and Schaffer, M. M. (1978). Context theory of classification learning. *Psychological review*, 85(3):207.

Nielsen, K. and Rehbeck, J. (2022). When choices are mistakes. *American Economic Review*, 112(7):2237–68.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of experimental psychology: General*, 115(1):39.

O'Donoghue, T. and Rabin, M. (1999). Doing it now or later. *American Economic Review*, 89:103–124.

Oprea, R. (2024). Simplicity equivalents. *Working Paper*.

Oprea, R. (2025). Initial reply to banki, simonsohn, walatka, and wu (2025). *Working Paper*.

Peirce, C. and Jastrow, J. (1884). On small differences in sensation. *National Academy of Sciences*, 3:73–83.

Rahnev, D., Desender, K., Lee, A. L., Adler, W. T., Aguilar-Lleyda, D., Akdoğan, B., Arbuzova, P., Atlas, L. Y., Balcı, F., Bang, J. W., et al. (2020). The confidence database. *Nature human behaviour*, 4(3):317–325.

Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive science*, 21(1):1–29.

Rosch, E. and Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7(4):573–605.

Stahl, S. A. and Fairbanks, M. M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of educational research*, 56(1):72–110.

Sunstein, C. R. (2020). *Behavioral Science and Public Policy*. Elements in Public Economics. Cambridge University Press.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science*, 12(2):257–285.

Vieider, F. (2024). Decisions under uncertainty as bayesian inference on choice options. Technical report.

Wu, G. (2025). There is no smoke with mirrors when instructions are clear. *Working Paper*.

A Online Appendix

A.1 Additional Tables and Figures

Table A1: Main Experiment: CA Method vs. Choice Reversals, Risky Tasks

(a) CA diagnosis: Error in one or more tables   (b) CA diagnosis: Errors in both tables

		$+1/-1$ **Choice**	
		No Reversal	Reversal
CA	0 Diagnosed	0.777	0.043
			$\beta = 0.754$
	≥ 1 Diagnosed	0.165	0.014
		$\alpha = 0.175$	

$N = 1{,}614;\ t = 1.6\ (p = 0.12)$

		$+1/-1$ **Choice**	
		No Reversal	Reversal
CA	< 2 Diagnosed	0.906	0.051
			$\beta = 0.895$
	2 Diagnosed	0.036	0.006
		$\alpha = 0.038$	

$N = 1{,}614;\ t = 2.1\ (p = 0.04)$

Notes: The tables cross-tabulate objective mistakes with observations diagnosed as improvable by the CA method in Pre-Choice data. Panel (a) relates reversals across $+1$ and -1 versions of a Risky task to cases with errors in at least one of the corresponding characterization tables. Panel (b) restricts to cases with errors in both characterization tables. α denotes the Type-I error rate, i.e., the likelihood of a false positive conditional on "No Reversal," and β the Type-II error rate, i.e., the likelihood of a false negative conditional on "Reversal." The t-statistic corresponds to test of equality in reversal rate for diagnosed and undiagnosed observations.
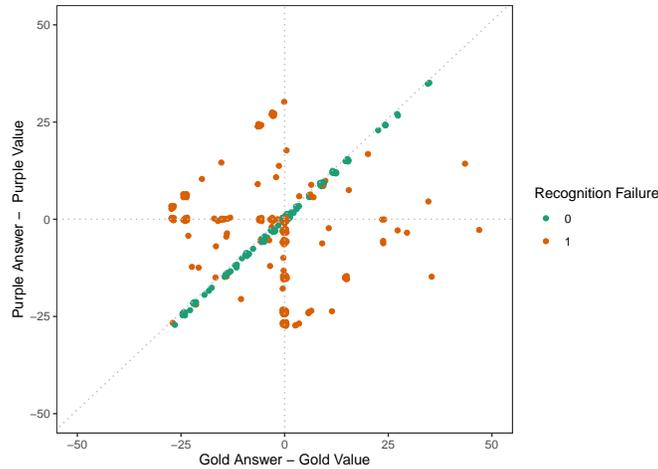
Figure A1: Main Experiment: Recognition and Table Entries Among Characterization Failures, Unmixed Risk-Free Tasks

Notes: Figure A1 shows responses for the values of gold and purple tickets in Option B relative to true values among subjects with characterization failures in Pre-Choice Risk-Free Unmixed tasks. The sample is restricted to cases with positive chances of gold and purple tickets and zero chance of blue tickets, so that distance to equal value corresponds to distance from recognizing the task as Risk-Free (532/637 characterization failures). Orange points denote recognition failures ($N = 293$), and green points denote recognition successes ($N = 239$). Eight extreme observations are excluded. Recognition failures far from the 45-degree line correspond to subjects assigning substantially different values to gold and purple tickets, while points far from $(0, 0)$ indicate large errors in characterizing the task.

Table A2: Overview of Follow-Up Experiments

	Experiment 2	Experiment 3	Experiment 4	Experiment 5	Experiment 6	Experiment 7
Task structure	5 Risky and 5 Risk-Free tasks in random order	5 Risky and 5 Risk-Free tasks in random order	10 Risky tasks followed by 10 Risk-Free tasks	10 Risky tasks and 10 Risk-Free tasks, block-randomized	10 Risky and 10 Risk-Free tasks in random order	20 tasks among Risky and Risk-Free; three versions of Risk-Free tasks
Characterization tables	Pre- and Post-Choice (unincentivized)	N/A	N/A	N/A	N/A	N/A
Confidence wording	"On a scale from 0 (not confident at all) to 100 (completely confident), how confident do you feel about this choice?"	"How certain are you (in %) that choosing [X] is actually your best decision, given your preferences and the available information?" w/ comprehension check	N/A	N/A	N/A	N/A
Stimuli	different	different	different	identical	different	different
Confidence decomposition	yes	yes	no	no	no	no
Reminders	no	no	yes	in one treatment	no	no
Explicit calculation of y and z	no	no	no	no	yes	yes
Follow-up questions	no	no	Math calculation	Math calculation & shortcuts w/ comprehension checks	no	no

A.2 Behavioral Differences Within Pre-Choice Conditions

We review behavioral differences between the Incentivized and Unincentivized Pre-Choice conditions in Experiment 1. Figure A2, Panels A and B, report the Risky tasks.
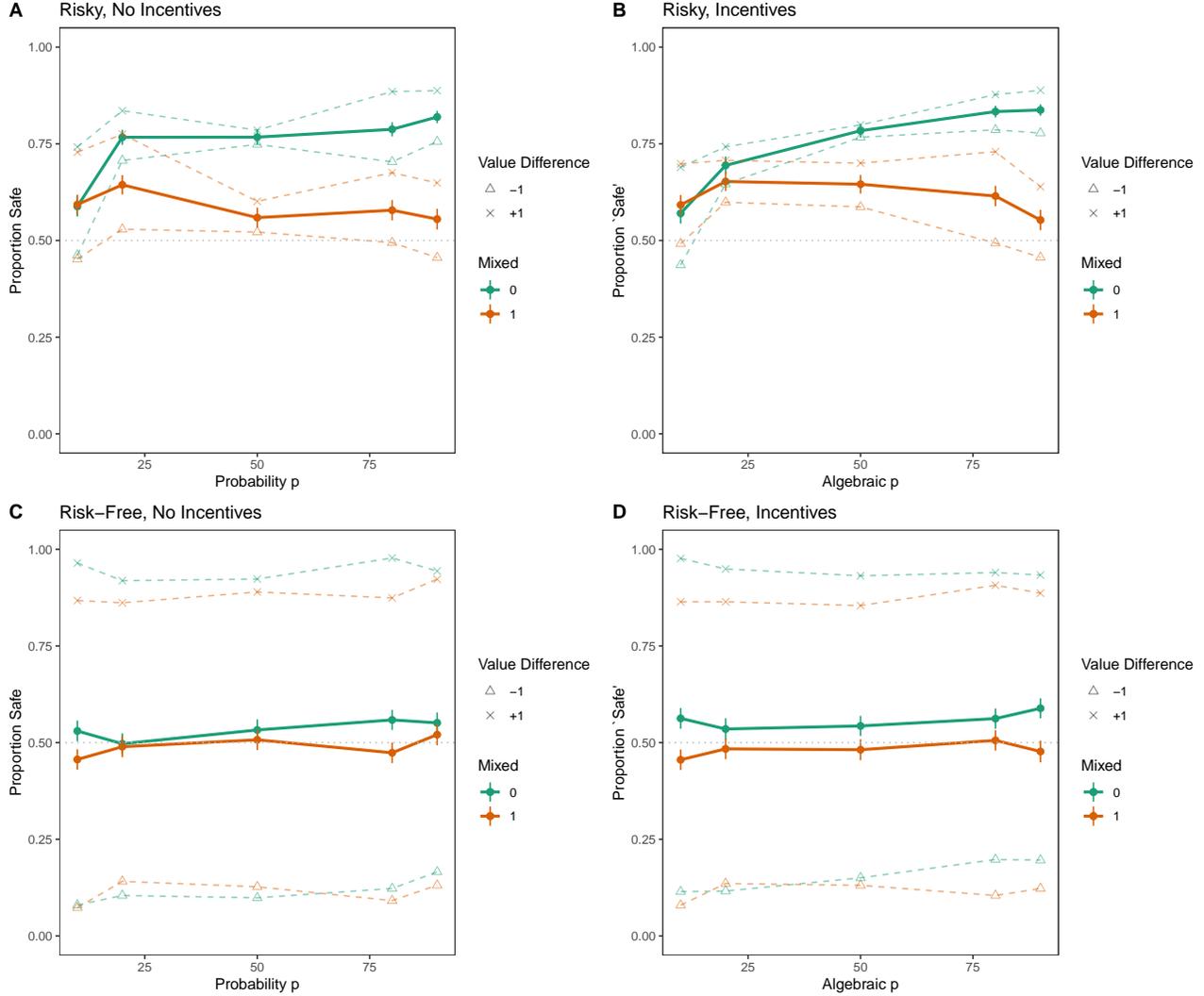
Figure A2: Experiment 1: Pre-Choice, No Incentives vs. Incentives

Notes: Figure A2 shows the proportion of choices for the Safe/"Safe" option in Unmixed tasks (green) and Mixed tasks (orange) under Pre-Choice conditions. Light shaded exes and triangles denote $+1$ and -1 tasks, while dark shaded circles indicate sample averages. Bars represent ± 1.96 standard errors (95% CIs). Panel A reports Risky tasks without incentives ($N = 3,328$), Panel B Risky tasks with incentives ($N = 3,346$), Panel C Risk-Free tasks without incentives ($N = 3,332$), and Panel D Risk-Free tasks with incentives ($N = 3,314$).

Subjects show substantial risk aversion in both: those without incentives choose the safe option in 75% of Unmixed cases and those with incentives in 74% ($t = 0.1$, $p = 0.92$). They also display substantial CRE: 16%-age points without incentives and 13%-age points with incentives ($t = 1.2$, $p = 0.23$). Panels C and D report Risk-Free tasks. Behavior largely accords with

maximization in both: subjects without incentives (with incentives) maximize in 90% (89%) of cases ($t = 1.6$, $p = 0.10$). Maximization errors are similarly unsystematic, with limited evidence of pattern matching to Risky behavior. In Unmixed cases, subjects without incentives (with incentives) choose the "Safe" option in 53% (56%) of cases ($t = 1.4$, $p = 0.16$) and exhibit a CRE of 5%-age (8%-age) points ($t = 1.3$, $p = 0.23$).

The only incentive-related differences appear in characterization. In Risky tasks, subjects without incentives (with incentives) correctly characterized 87% (91%) of cases ($t = 5.4$, $p < 0.01$). In Risk-Free tasks, they did so in 76% (78%) of cases ($t = 2.0$, $p = 0.05$).[66]

A.3   Confidence and Risk-Free Choice

Figure A3 reproduces Figure 4 for Risk-Free Post-Choice tasks in Experiment 1.
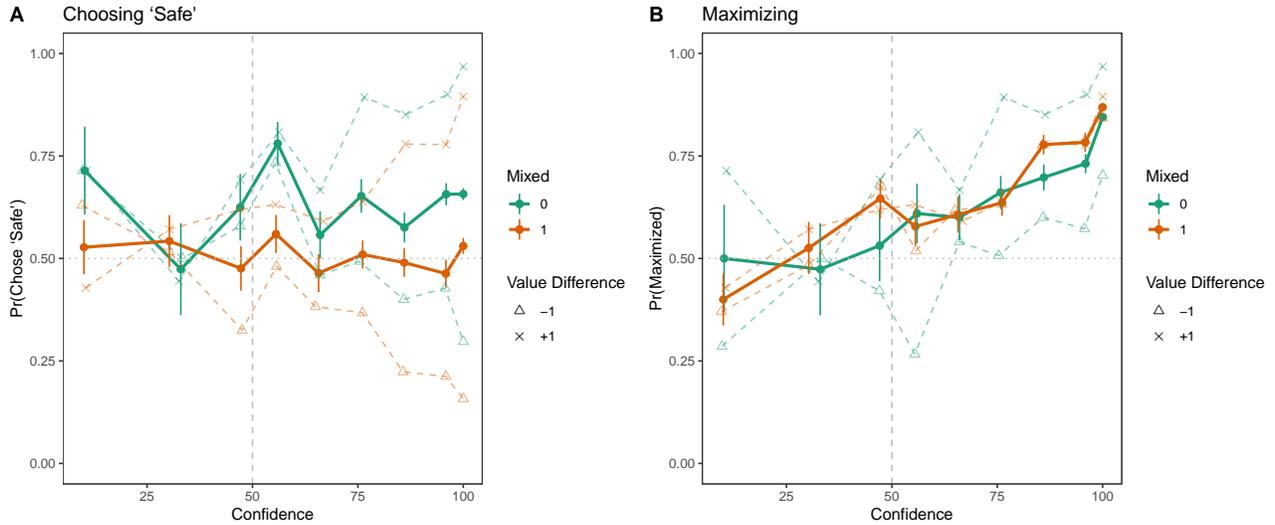
Figure A3: Main Experiment: Risk-Free Tasks and Confidence, Post-Choice Condition

Notes: Figure A3 shows behavior in Risk-Free tasks for subjects in the Post-Choice condition ($N = 3,335$). Panel A relates choice of the "Safe" option to stated confidence. Confidence is divided into nine bins, and within each bin both average confidence and the share choosing "Safe" are calculated. Results are shown separately for Unmixed tasks (green) and Mixed tasks (orange). Light shaded exes and triangles denote $+1$ and -1 tasks, while dark shaded circles indicate sample averages. Bars represent ± 1.96 standard errors (95% CIs). Panel B repeats this analysis using the proportion of subjects maximizing value.

Contrary to Risky choice, confidence is essentially uncorrelated with choosing the "Safe" option ($\rho = 0.04$, $p = 0.02$) but strongly correlated with maximizing expected value ($\rho = 0.25$, $p < 0.01$). Finally, confidence is slightly negatively correlated with accurate characterization tables in Risky choice ($\rho = -0.05$, $p < 0.01$) but positively correlated with accurate characterization in Risk-Free choice ($\rho = 0.20$, $p < 0.01$).

---

[66]This result does not hold once task characteristics and clustered errors are accounted for. Controlling for p, Mixed/Unmixed, and ± 1 task features, the incentive coefficient is 0.02 (0.02), $t = 0.8$, $p = 0.41$.

A.4 Decision Confidence and Behavioral Attenuation

Enke and Graeber (2023) and Enke et al. (2024) link decision confidence to responsiveness to changes in problem fundamentals, generally finding that lower confidence is associated with reduced responsiveness ("attenuation"). Our +1 and −1 tasks shift fundamentals, allowing us to examine the relationship between confidence and attenuation. Figures 4 and A3 show heterogeneous responses to changes in the value of the Safe/"Safe" alternative: The most confident are the most responsive in Risk-Free tasks but the least responsive in Risky tasks. This latter result for Risky tasks is inconsistent with prior findings on attenuation.

To probe further, we study the link between +1/ − 1 responsiveness (change in choice) and confidence at the individual level in the Post-Choice condition. Figure A4 reports the results.
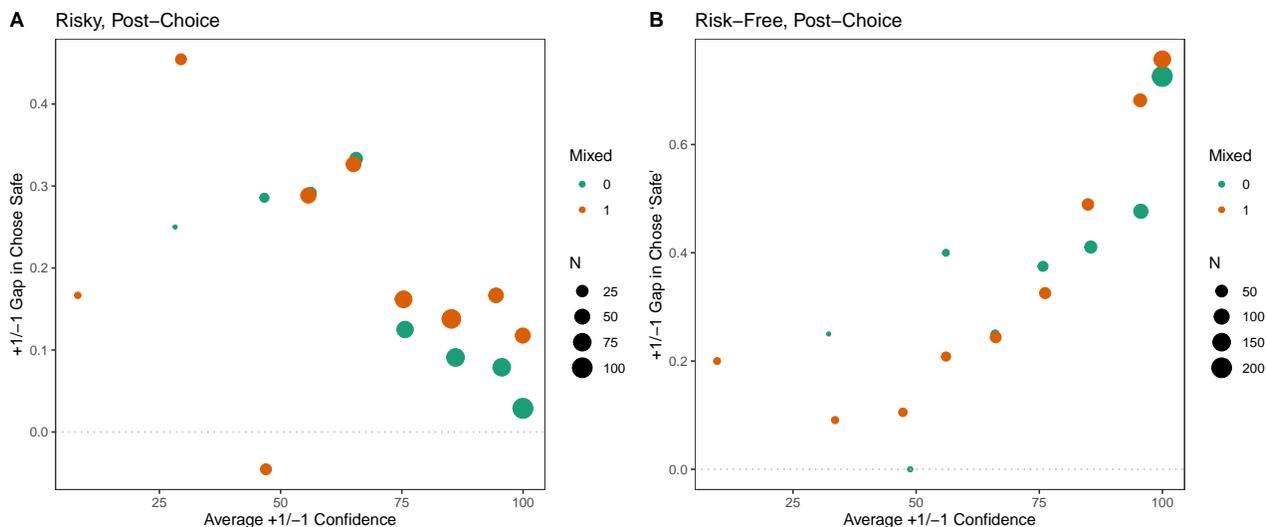
Figure A4: Experiment 1: Confidence and Attenuation, Post-Choice

Notes: Figure A4 shows Post-Choice Unmixed (green) and Mixed (orange) observations in which subjects completed both +1 and −1 versions of a task. Average confidence across the two versions is related to responsiveness, measured as the difference in choosing the "Safe" option in Risky (Panel A) and Risk-Free (Panel B) tasks. Confidence is divided into nine bins, with average confidence and responsiveness calculated within each.

Once again, we observe behavioral attenuation in Risk-Free tasks: Unmixed ($N = 417$, $\rho = 0.28$, $p < 0.01$) and Mixed ($N = 393$, $\rho = 0.35$, $p < 0.01$). On the contrary, we observe behavioral amplification in Risky tasks: Unmixed ($N = 396$, $\rho = -0.19$, $p < 0.01$) and Mixed ($N = 405$, $\rho = -0.11$, $p < 0.05$).

An important reason why our results may differ from those in the cognitive-uncertainty agenda is the wording of the confidence question. We assess this in Experiment 3, which adopts the confidence wording of Enke et al. (2024). As in Experiment 1, we find attenuation in Risk-Free tasks: Unmixed ($\rho = 0.31, p < 0.01$) and Mixed ($\rho = 0.26, p < 0.01$). We also find amplification in Risky tasks: Unmixed ($\rho = -0.12, p < 0.05$) and Mixed ($\rho = -0.02, p = 0.67$).

A.5 Evaluating the Effectiveness of Corrective Comprehension Questions

Recent debate in the literature (see, e.g., Banki et al., 2025; Oprea, 2025) focuses on comprehension questions with corrective feedback. For our purposes, the key issue is whether subjects who first err—potentially due to recognition failures—learn from correction and subsequently grasp the distinction between Risky and Risk-Free tasks. To examine this, we analyze evidence from Experiments 1 and 5, the latter using the same comprehension questions as Oprea (2025).

Experiment 1 includes a comprehension exercise with feedback when introducing the characterization tables. Because characterization is measured in every task, we can assess whether subjects who initially erred and were corrected learned from the experience. In the Pre-Choice condition, where the quiz precedes any decisions, 42% of subjects make at least one error (34% in the Risk-Free component, 19% in the Risky component). These errors strongly predict subsequent characterization and recognition failures, as well as choice errors. Subjects with a Risk-Free comprehension error exhibit characterization failures in 42% of Risk-Free tables (vs. 13% without errors, $t = 24.4$, $p < 0.01$), recognition failures in 27% (vs. 5%, $t = 21.9$, $p < 0.01$), and maximization failures in 18% (vs. 7%, $t = 12.9$, $p < 0.01$). They also display greater small-stakes "risk aversion" (60% vs. 52% Safe in Unmixed tasks, $t = 4.2$, $p < 0.01$) and larger "CRE" (11 vs. 4 %-age points, $t = 2.7$, $p < 0.01$) in Risk-Free tasks.

The findings in Experiment 5 are consistent with those in Experiment 1. Comprehension quiz errors in Experiment 5 strongly predict maximization errors in Risk-Free tasks, even though mistakes were corrected. Moreover, subjects who err on the quiz are more likely to invoke risk considerations when asked to describe their decision-making approaches after their choices. Overall, corrective comprehension quizzes of this form seem unlikely to eliminate characterization failures, recognition failures, maximization failures, and the corresponding systematic choice errors that emerge in Risk-Free tasks.

A.6 Additional Results from Experiment 2

In what follows, Appendix A.6.1 reports subjects' behavior in Experiment 2, Appendix A.6.2 links it to recognition failures, Appendix A.6.3 examines its relationship with confidence, and Appendix A.6.4 analyzes the reported rationales for less-than-complete decision confidence.

A.6.1 Experiment 2: Behavioral Patterns

Experiment 2 studies rationales for low confidence using the same Unincentivized Pre-Choice and Post-Choice designs as in Experiment 1, but with 5 Risky and 5 Risk-Free tasks (10 total) instead of 20. Figure A5 shows results by condition and replicates the main behavioral patterns.
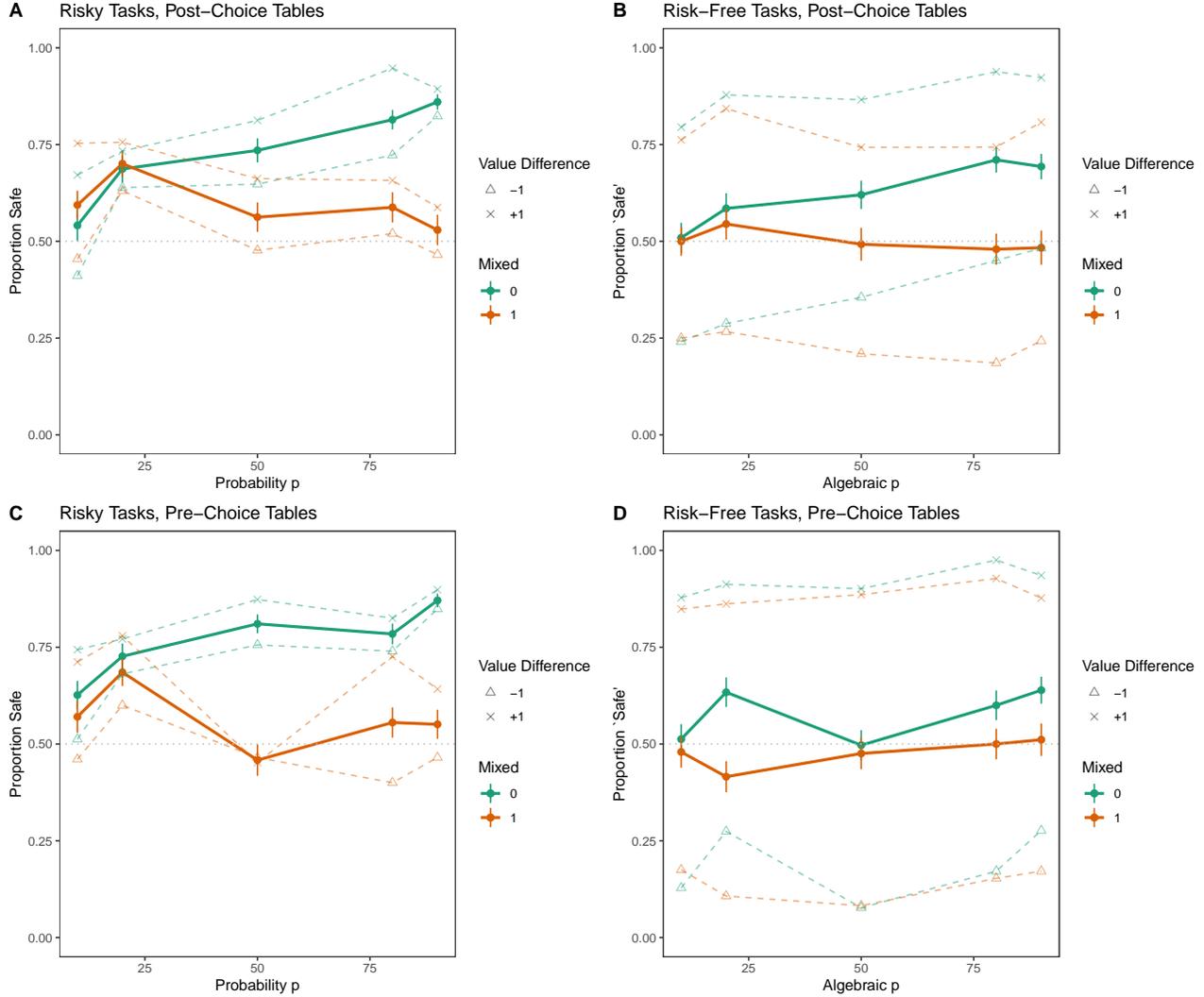
Figure A5: Experiment 2: Post-Choice and Pre-Choice conditions

Notes: Figure A5 shows the proportion of choices for the Safe/"Safe" option in Unmixed tasks (green) and Mixed tasks (orange) in Experiment 2. Light shaded exes and triangles denote +1 and −1 tasks, while dark shaded circles indicate sample averages. Bars represent ±1.96 standard errors (95% CIs). Panels A and C report Risky tasks in the Post-Choice and Pre-Choice conditions, respectively; Panels B and D report Risk-Free tasks in the Post-Choice and Pre-Choice conditions, respectively. The sample size in each panel is $N = 1,500$.

We find small-stakes risk aversion (73% Safe in Unmixed) and a CRE of 13%-age points in the Post-Choice Risky tasks (Panel A). Panel B shows similar patterns in Risk-Free choice (62% "Safe" in Unmixed; CRE = 12%-age points). As in Experiment 1, Pre-Choice characterization tables substantially affect Risk-Free behavior: Objective maximization failures occur in 23% of Post-Choice vs. 13% of Pre-Choice observations ($t = 7.7$, $p < 0.01$). Accordingly, Pre-Choice observations display less pattern matching. In Risky choice, we again find small-stakes risk aversion (76% Safe in Unmixed; CRE = 20%-age points), but these effects are weaker in

Risk-Free choice (58% "Safe"; CRE = 10%-age points). Panels C and D illustrate these results.

Mean behavioral differences between Pre- and Post-Choice in Risk-Free tasks are modest (4%-age points for "Safe" in Unmixed; 2%-age points for the CRE), but become clearer once we control for task characteristics. Regressing "Safe" choice on condition interacted with Mixed/Unmixed, and controlling for p and ± 1 task features, yields a Post-Choice coefficient of 0.07 (clustered s.e. = 0.02; $t = 3.2$, $p < 0.01$). The interaction of Mixed and Post-Choice is -0.06 (clustered s.e. = 0.03; $t = 2.2$, $p = 0.03$).

A.6.2 Experiment 2: Recognition Failures

As in Experiment 1, recognition failures remain a key driver of errors in Risk-Free choice in Experiment 2. Figure A6 reproduces Figure 8 using the Pre-Choice tasks of Experiment 2.
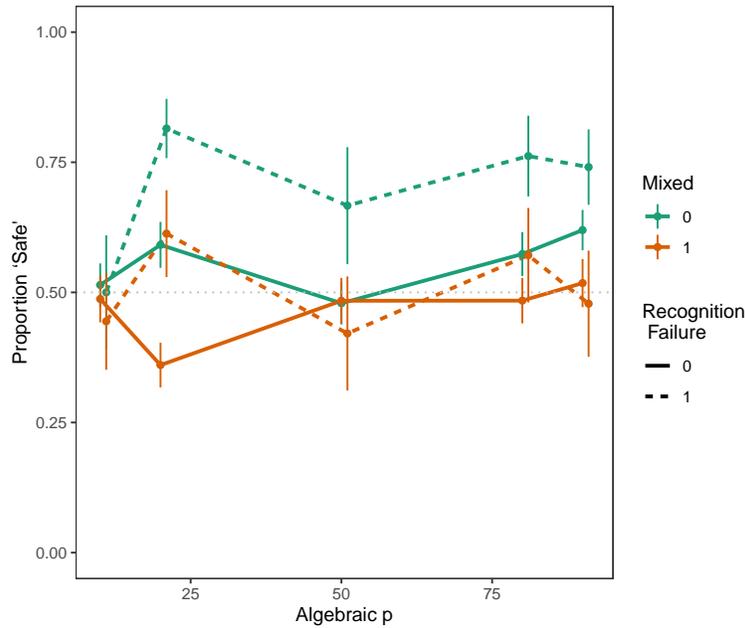
Figure A6: Experiment 2: Risk-Free Tasks and Recognition Failures, Pre-Choice Condition

Notes: Figure A6 shows the proportion of choices for the "Safe" option in Unmixed tasks (green) and Mixed tasks (orange). Bars represent ± 1.96 standard errors (95% CIs). Solid and dashed lines indicate averages without and with recognition failures, respectively.

In total, 27% of Risk-Free characterization tables contain errors in Pre-Choice tasks, with 58% of these errors being recognition failures—closely matching Experiment 1. Recognition failures again relate to Risk-Free choice: observations without (with) failures choose the "Safe" option in 55% (71%) of Unmixed tasks ($t = 3.0$, $p < 0.01$), exhibiting a "CRE" of 9%-age points (19%-age points) ($t = 1.5$, $p = 0.13$).[67]

---

[67]Regressing "Safe" choice on recognition interacted with Mixed/Unmixed, controlling for p and ± 1 task

A.6.3 Experiment 2: Confidence and Behavior

The relationship between confidence and choice in Experiment 2 echoes that in Experiment 1. Figure A7, Panels A and B, reproduce Figure 4 for Post-Choice Risky tasks.
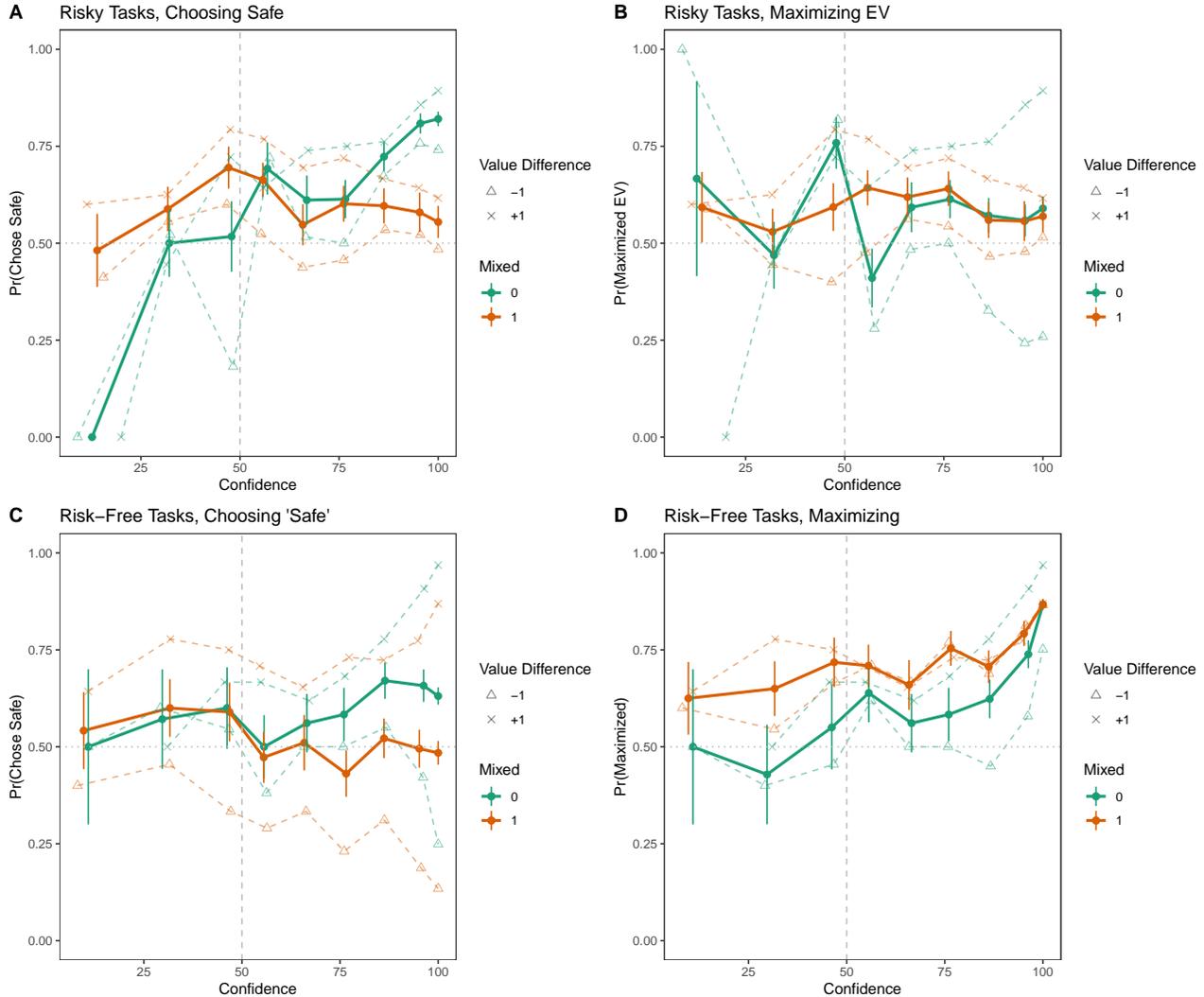
Figure A7: Experiment 2: Confidence and Behavior, Risky and Risk-Free Post-Choice Tasks

Notes: Figure A7 shows behavior in Risky tasks ($N = 1,500$) and Risk-Free tasks ($N = 1,500$) for subjects in the Post-Choice condition. Panels A and C relate choice of the Safe/"Safe" option to stated confidence in Risky and Risk-Free tasks, respectively. Confidence is divided into nine bins, with average confidence and the share choosing Safe computed within each. Results are shown separately for Unmixed tasks (green) and Mixed tasks (orange). Light shaded exes and triangles denote $+1$ and -1 tasks, while dark shaded circles indicate sample averages. Bars represent ± 1.96 standard errors (95% CIs). Panels B and D repeat the analysis of Panels A and C using the proportions of subjects maximizing expected value.

Confidence correlates positively with choosing the Safe option in Unmixed tasks ($\rho = 0.24$,

---

features, the recognition coefficient is -0.16 (clustered s.e.=0.05; $t = 3.0$, $p < 0.01$). The Mixed×recognition interaction is 0.12 (s.e.=0.07; $t = 1.7$, $p = 0.10$).

$p < 0.01$) but not in Mixed tasks ($\rho = -0.03$, $p = 0.44$). By contrast, expected value maximization is uncorrelated with confidence in both Unmixed tasks ($\rho = 0.02$, $p = 0.62$) and Mixed tasks ($\rho = -0.01$, $p = 0.73$). Figure A7, Panels C and D, reproduce Figure A3 for Risk-Free tasks in the Post-Choice condition. Relative to Risky choice, the correlations with confidence reverse: in Risk-Free choice, confidence is essentially uncorrelated with choosing the "Safe" option ($\rho = 0.02$, $p = 0.44$) but strongly correlated with maximizing value ($\rho = 0.19$, $p < 0.01$).

A.6.4   Experiment 2: Rationales for less-than-complete confidence

The results in the main text for Experiment 2 on rationales for less-than-complete confidence hold for the 83% of subjects who passed the classification test on these rationales. Figure A8 replicates Figure 6 for this subsample, showing that improvable rationales are ranked as most important in 49% of cases in Risk-Free choice, but only 19% of cases in Risky choice.
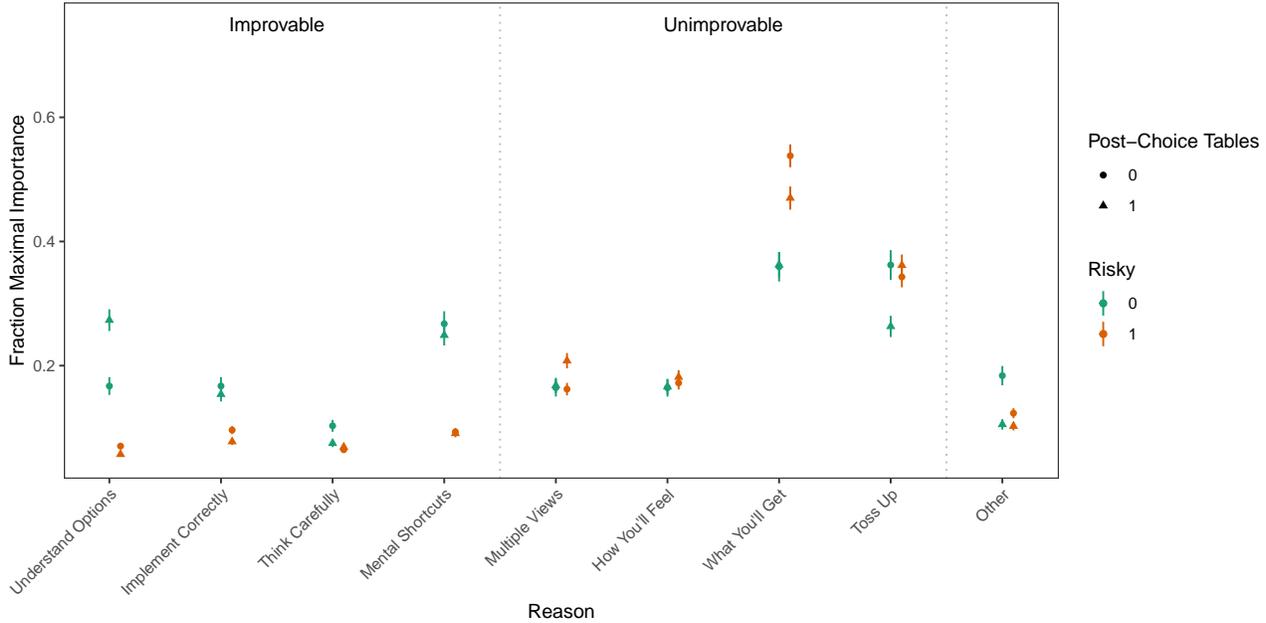
Figure A8: Experiment 2: Confidence Decomposition, Passed Classification Test

Notes: Figure A8 shows how frequently rationales for less-than-complete confidence are rated most important by subjects in the rationalization sample of Experiment 2 who passed the related quiz. Results are shown separately for Risky (orange, $N = 1,380$) and Risk-Free (green, $N = 853$) choices, and for Post-Choice (triangles, $N = 1,364$) and Pre-Choice (circles, $N = 1,184$) conditions; bars represent ±1.96 standard errors (95% CIs). Unbalanced subsamples reflect complete confidence being more frequent in Risk-Free than in Risky choices.

A.7   Additional Results from Experiment 3

Experiment 3 followed the design of Experiment 2 in eliciting rationales for less-than-complete confidence, but adopted the confidence wording and the comprehension question on the purpose

of confidence elicitation from Enke et al. (2024). Because it excluded characterization tables, it is most comparable to the Post-Choice condition of Experiments 1 and 2. Figure A9 reproduces pattern-matching across Risky and Risk-Free tasks: Subjects choose the safe option in 68% of Unmixed cases and exhibit a CRE of 15%-age points; in Risk-Free tasks, they choose "Safe" in 61% of Unmixed cases and exhibit a CRE of 12%-age points.
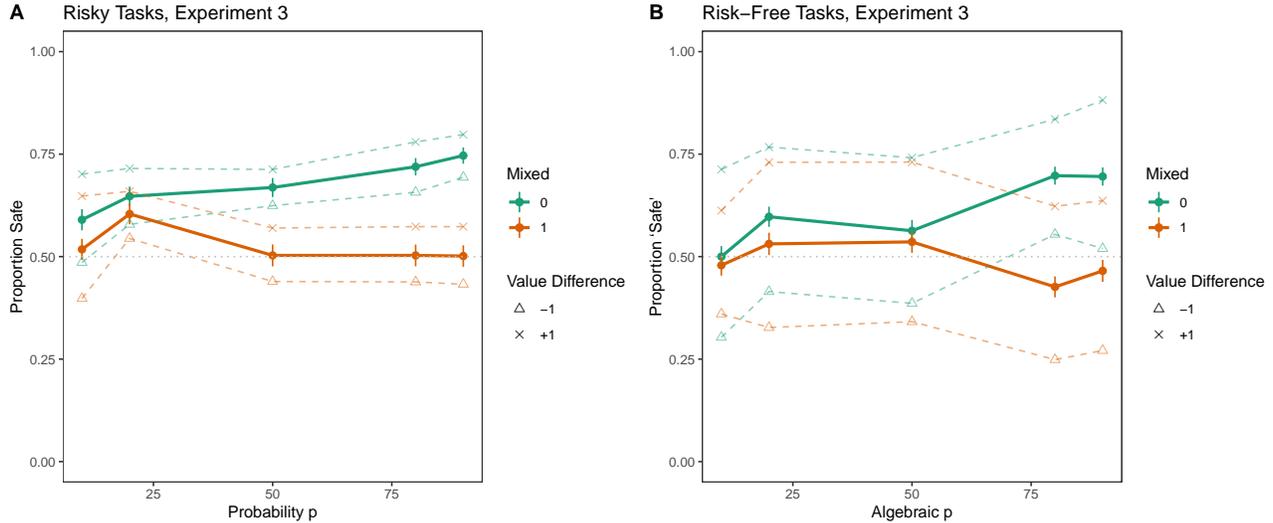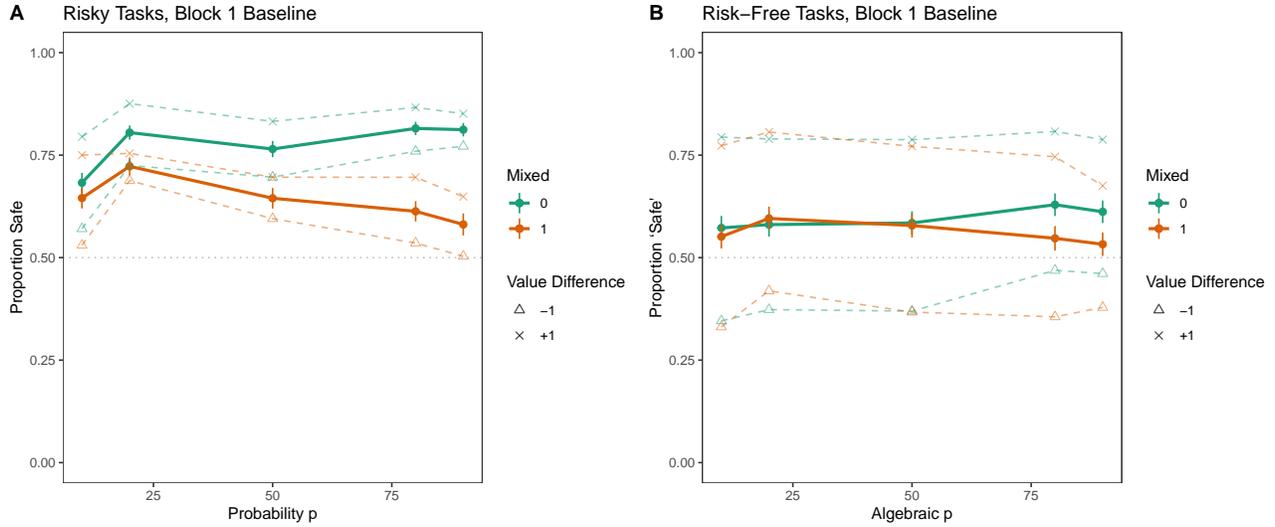

Figure A9: Experiment 3: Risky and Risk-Free Tasks

Notes: Figure A9 shows the proportion of choices for the Safe/"Safe" option in Unmixed tasks (green) and Mixed tasks (orange). Light shaded exes and triangles denote +1 and −1 tasks, while dark shaded circles indicate sample averages. Bars represent ±1.96 standard errors (95% CIs). Panel A reports Risky tasks ($N = 3{,}490$) and Panel B Risk-Free tasks ($N = 3{,}490$) for subjects in Experiment 3.

Figure A10 relates the alternative confidence measure to choice. As in Experiments 1 and 2, correlations differ between by task type. In Risky tasks, confidence correlates with choosing the Safe option in Unmixed tasks ($\rho = 0.21$, $p < 0.01$) but not in Mixed tasks ($\rho = 0.02$, $p = 0.46$), and the correlation with expected value maximization is negligible in both Unmixed ($\rho = 0.0001$, $p = 0.97$) and Mixed tasks ($\rho = 0.002$, $p = 0.92$). In Risk-Free tasks, by contrast, confidence is only weakly correlated with choosing "Safe" ($\rho = 0.05$, $p < 0.01$) but strongly correlated with maximizing expected value ($\rho = 0.20$, $p < 0.01$).

Figure A11 reproduces Figure 6 for Experiment 3, presenting the confidence decomposition. As in Experiment 2, subjects often attribute less-than-complete confidence in Risky choice to unimprovable rationales; even with the Enke et al. (2024) training question clarifying that confidence should not reflect payment likelihood (97% passed on the first attempt), they frequently cite "Can't Know For Sure What You'll Get" as the most important rationale, selecting it in 45% of such tasks (closely paralleling Experiment 2).
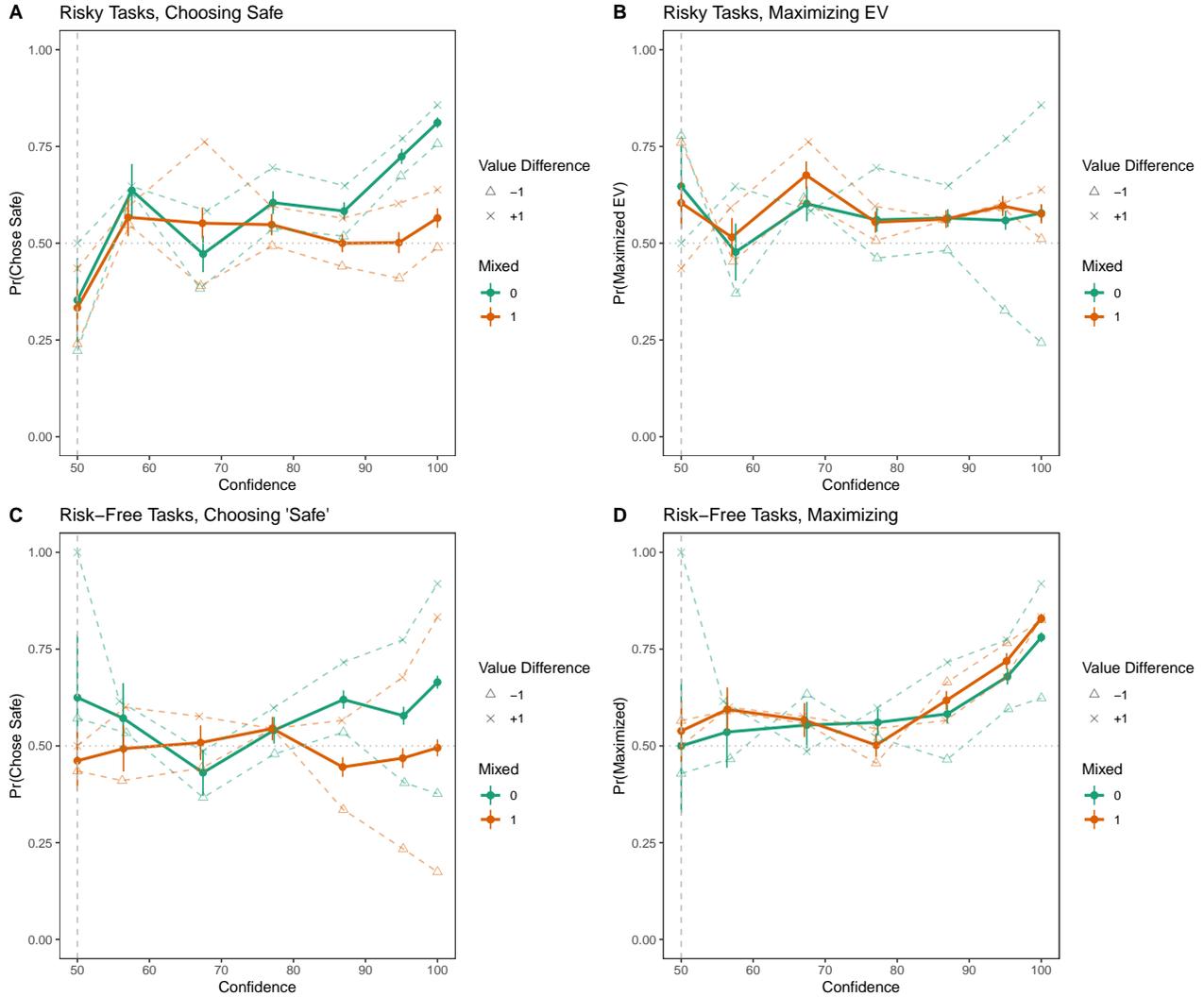
Figure A10: Experiment 3: Confidence and Behavior, Risky and Risk-Free Post-Choice Tasks

Notes: Figure A10 shows Experiment 3 behavior in Risky and Risk-Free tasks (each $N = 3,490$). Panels A and C relate choice of the Safe/"Safe" option to stated confidence in Risky and Risk-Free tasks, respectively. Confidence is binned into nine intervals; within-bin averages of confidence and the Safe share are shown separately for Unmixed (green) and Mixed tasks (orange). Light shaded exes and triangles denote $+1$ and -1 tasks, while dark shaded circles indicate sample averages. Bars represent ± 1.96 standard errors (95% CIs). Panels B and D repeat the analysis of Panels A and C using the proportion of subjects maximizing expected value.

A.8 Additional Results from Experiment 4

Experiment 4 used a blocked design (10 Risky tasks, then 10 Risk-Free tasks) and explicitly distinguished Risk-Free tasks by stating that there was no risk and all tickets paid the same amount. Figure A12 reports the basic pattern-matching results (described in the main text)
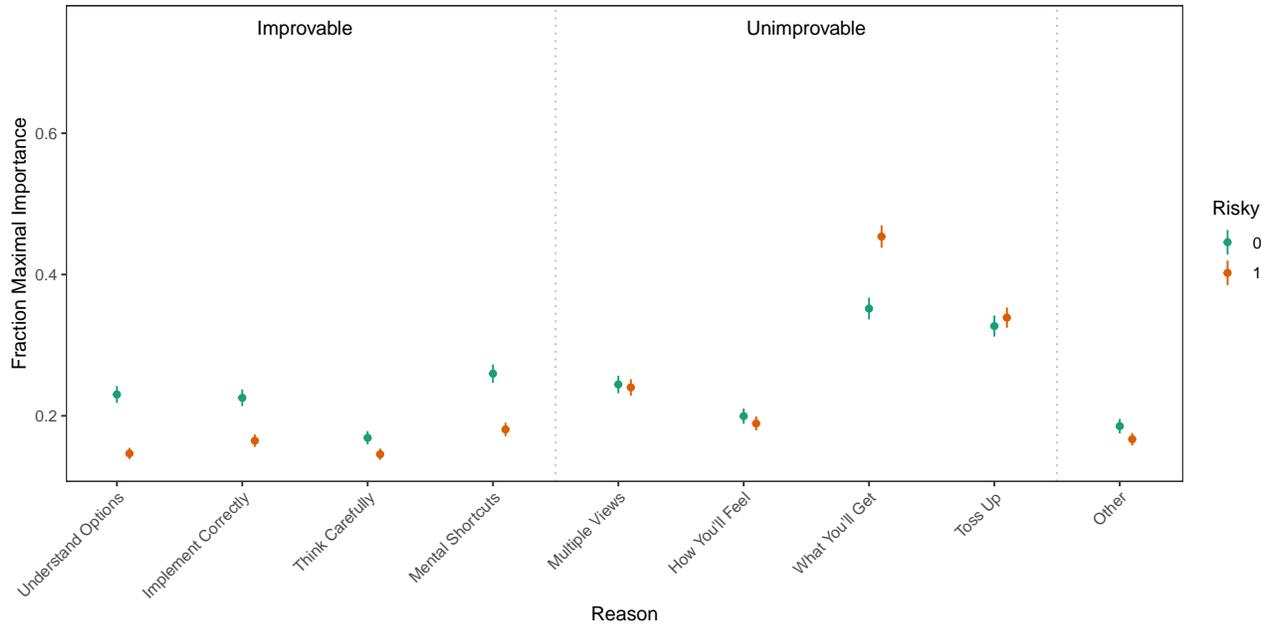
Figure A11: Experiment 3: Confidence Decomposition

Notes: Figure A11 shows the most important rationales for less-than-complete confidence in subjects' three least confident Risky choices (orange, $N = 941$) and Risk-Free choices (green, $N = 847$). Unbalanced subsamples reflect complete confidence being more frequent in Risk-Free than in Risky choices. Bars represent ± 1.96 standard errors (95% CIs).

and shows notably less pattern matching than Experiment 1's Post-Choice condition.[68]

After each block, subjects indicated whether they used mathematical calculations and, if not, why not. Figure A13 shows that the 245 subjects (82% of the sample) reporting calculations in Risk-Free tasks exhibit no anomalous Risk-Free pattern, while their Risky choices display the familiar small-stakes risk aversion and CRE observed in our other studies.

Of the 55 subjects reporting no calculations, 29 (10% of the sample) viewed them as unnecessary in Risk-Free tasks (consistent with recognition failure) and 22 (7% of the sample) as too effortful. Figure A14, Panels A and B, show behavior in Risky and Risk-Free tasks for the first subgroup, and Panels C and D for the second. In both subgroups, we observe evidence of anomalous choice patterns in Risk-Free tasks alongside pattern matching in Risky tasks.

A.9   Additional Results from Experiment 5

In what follows, Appendix A.9.1 reports behavior in the first block of Experiment 5 and contrasts the enhanced recognition and baseline treatments. Appendix A.9.2 analyzes answers to calculation questions and their correlation with choice. Appendix A.9.3 examines self-reported

---

[68]Experiment 4 did not collect confidence judgments, so we cannot study the relationship between confidence and Risky/Risk-Free choice as in Experiments 1, 2 and 3.
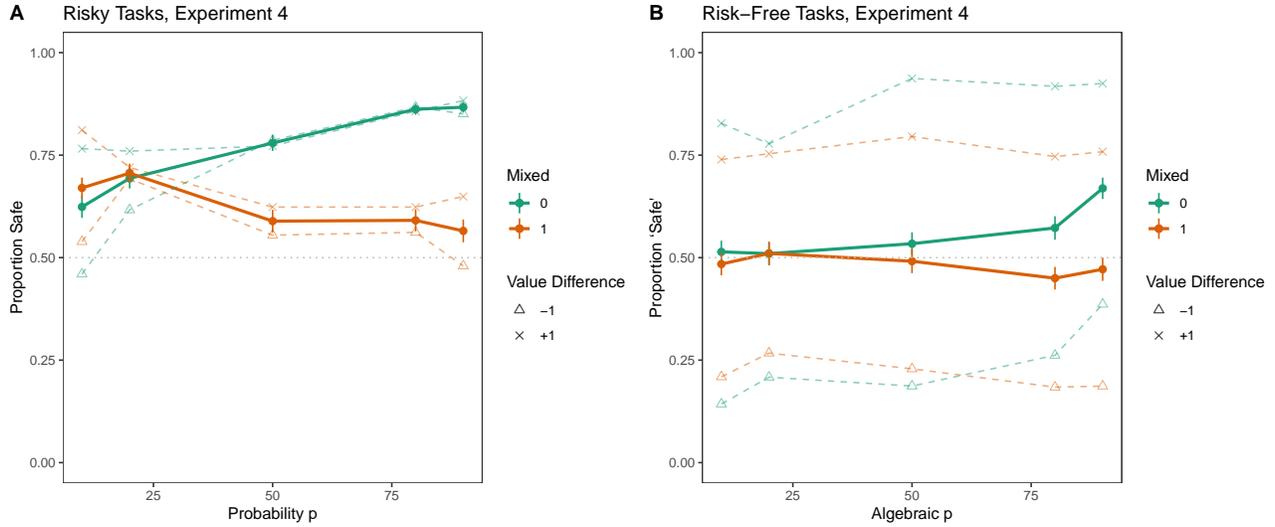
Figure A12: Experiment 4: Risky and Risk-Free Tasks

Notes: Figure A12 shows the proportion of choices for the Safe/"Safe" option in Unmixed (green) and Mixed tasks (orange). Light shaded exes and triangles denote +1 and −1 tasks, respectively; dark shaded circles indicate sample averages. Bars represent ±1.96 standard errors (95% CIs). Panels A and B show Risky and Risk-Free tasks (each $N = 3{,}000$) for Experiment 4 subjects.
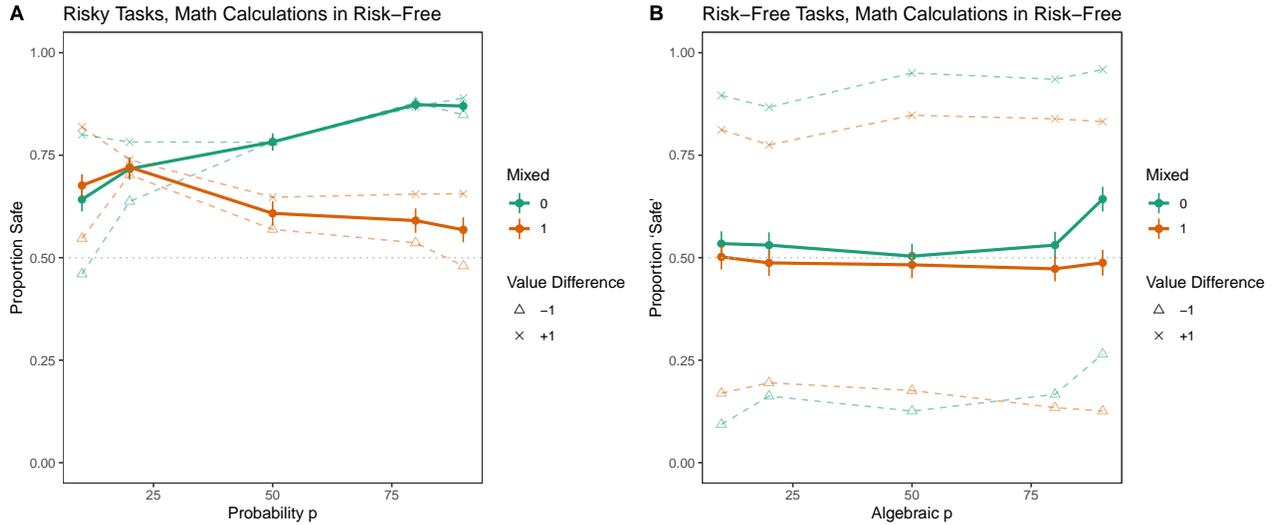
Figure A13: Experiment 4: Subjects Reporting Math Calculations in Risk-Free Tasks

Notes: Figure A13 shows the proportion of choices for the Safe/'Safe' option for Unmixed (green) and Mixed tasks (orange). Light shaded exes and triangles correspond to +1 and −1 tasks, respectively; dark shaded circles correspond to sample averages. Bars represent ±1.96 standard errors (95% CIs). Panels A and B show Risky and Risk-Free tasks (each $N = 2{,}450$) for Experiment 4 subjects reporting calculations in Risk-Free tasks.

reliance on shortcuts and its correlation with choice. Appendix A.9.4 explores whether subjects' shortcut descriptions reveal risk-preferences considerations in Risk-Free tasks. Appendix A.9.5 reports behavior in the second block of Experiment 5, highlighting potential order effects.
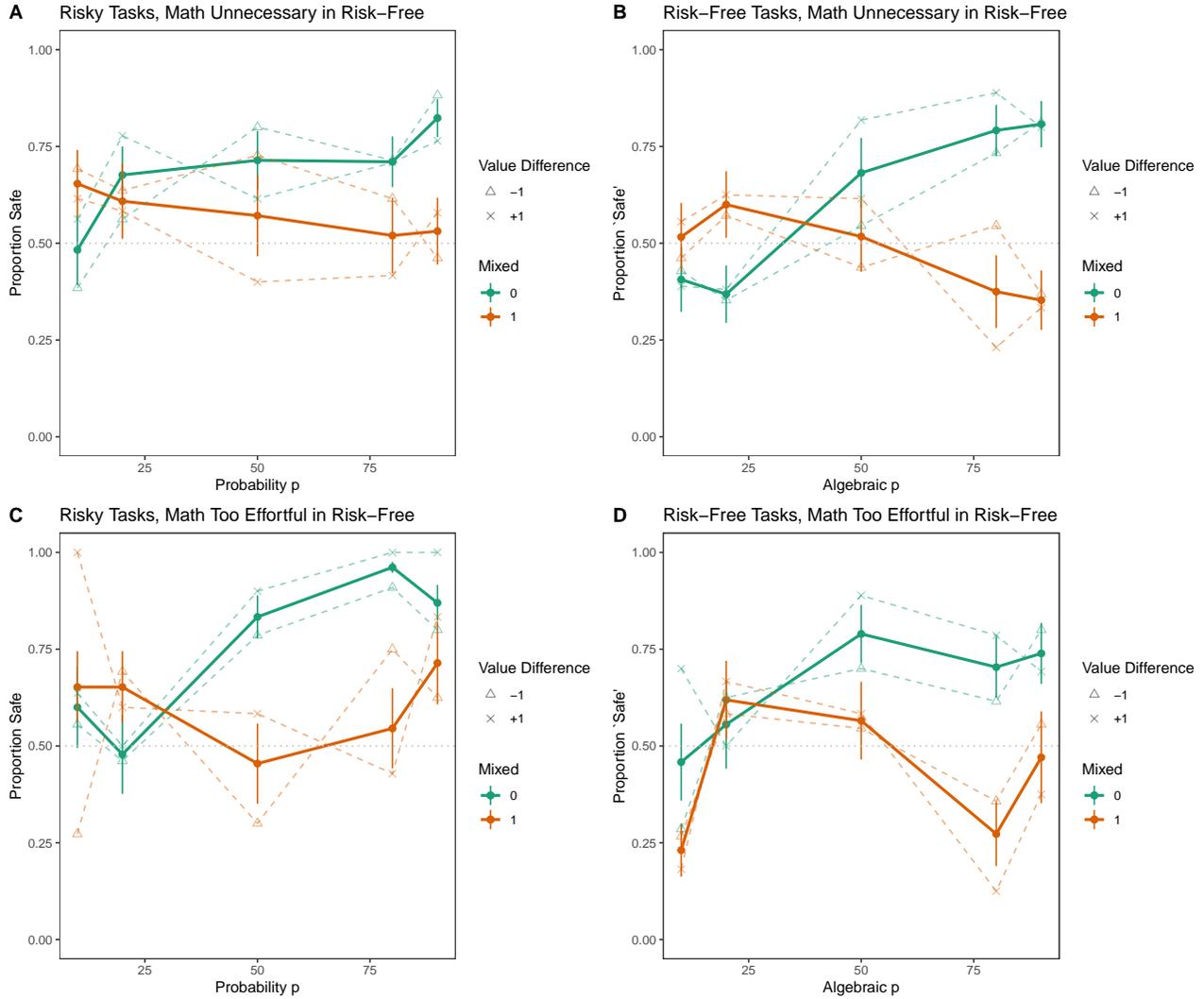
Figure A14: Experiment 4: No Calculations in Risk-Free Tasks

Notes: Figure A14 shows the proportion of choices for the Safe/"Safe" option in Unmixed (green) and Mixed tasks (orange). Light shaded exes and triangles denote +1 and −1 tasks; dark shaded circles indicate sample averages. Bars represent ±1.96 standard errors (95% CIs). Panels A and B show Risky and Risk-Free tasks (each $N = 290$) for subjects reporting no calculations in Risk-Free tasks because unnecessary. Panels C and D show Risky and Risk-Free tasks (each $N = 220$) for subjects reporting no calculations because too effortful.

A.9.1 Experiment 5: Treatment Comparison

Experiment 5 included two dimensions of randomization: subjects were assigned either Risky or Risk-Free tasks in their first block and, independently, to an enhanced recognition treatment and a baseline treatment. In the main text, we reported results from first block observations in these two conditions, showing limited evidence of pattern matching in a between-subjects analysis. Figures A16 and A15 report the corresponding pattern matching figures.

Figure A15: Experiment 5, Block 1 Baseline: Risky and Risk-Free Tasks

Notes: Figure A15 shows the proportion of choices for the Safe/"Safe" option for Unmixed (green), and Mixed tasks (orange). Light shaded exes and triangles correspond to +1 and −1 tasks; dark shaded circles correspond to sample averages. Bars represent ±1.96 standard errors (95% CIs). Panels A and B show Risky ($N = 2{,}760$) and Risk-Free ($N = 3{,}240$) Block 1 tasks for subjects in the baseline condition of Experiment 5.
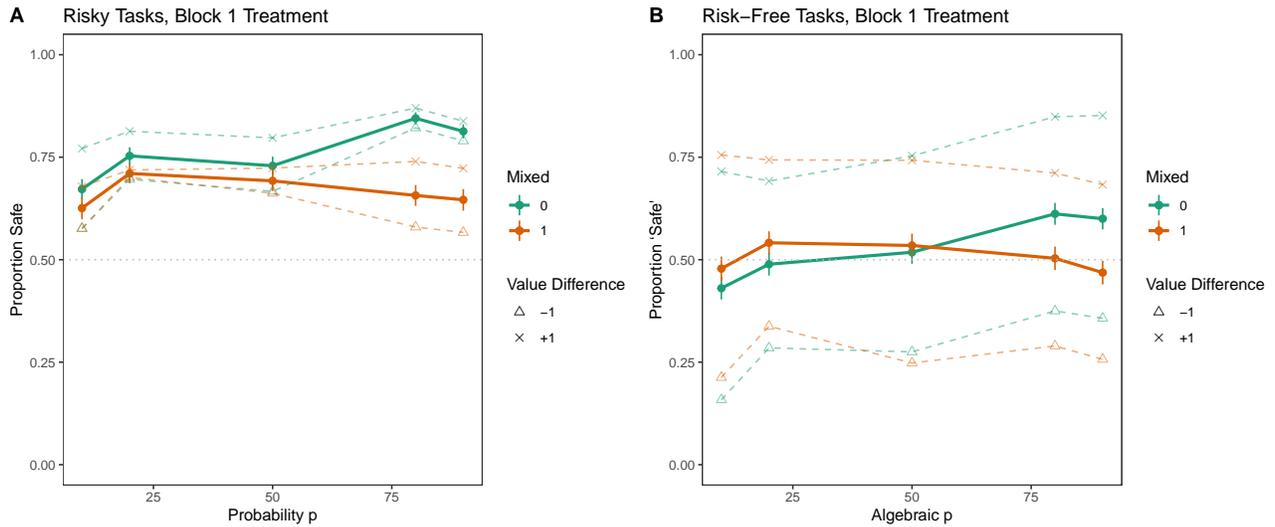
Figure A16: Experiment 5, Block 1 Enhanced Recognition: Risky and Risk-Free Tasks

Notes: Figure A16 shows the proportion of choices for the Safe/'Safe' option for Unmixed (green) and Mixed tasks (orange). Light shaded exes and triangles correspond to +1 and −1 tasks; dark shaded circles correspond to sample averages. Bars represent ±1.96 standard errors (95% CIs). Panels A and B show Risky ($N = 2{,}990$) and Risk-Free ($N = 3{,}010$) tasks for Block 1 subjects in the enhanced recognition condition of Experiment 5.

A.9.2 Experiment 5: Mathematical Calculations

As in Experiment 4, after each block of tasks in Experiment 5 we asked subjects whether they engaged in mathematical calculations and, if not, why not. Figure A17 presents Risk-Free Block 1 data for the 70% of subjects who reported engaging in calculations (404/577), the 18% who reported not engaging because calculations were unnecessary (101/577), and the 10% who reported not engaging because they required too much effort (58/577).
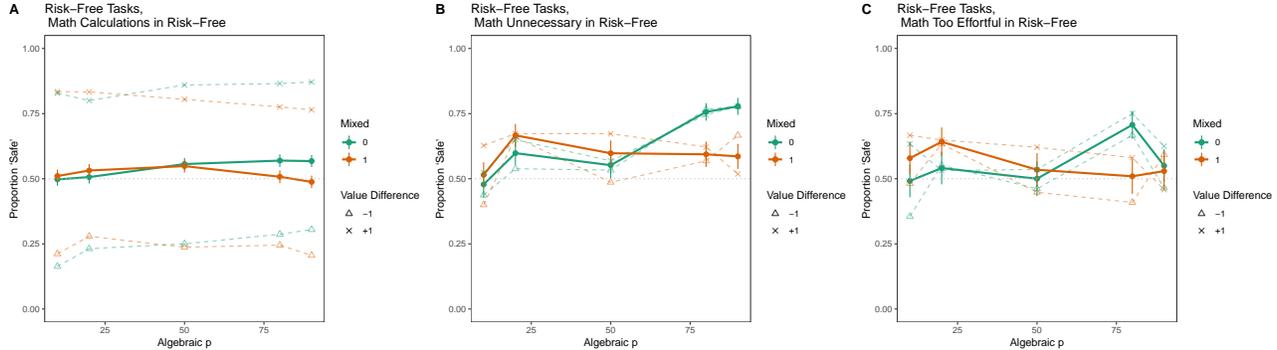
Figure A17: Experiment 5: Risk-Free Tasks and Mathematical Calculations, Block 1

Notes: Figure A17 shows the proportion of choices for the "Safe" option in Unmixed tasks (green) and Mixed tasks (orange). Light shaded exes and triangles denote +1 and −1 tasks, while dark shaded circles indicate sample averages. Bars represent ±1.96 standard errors (95% CIs). Panel A reports subjects who engaged in mathematical calculations in Block 1 Risk-Free tasks of Experiment 5 ($N = 4,040$). Panel B reports subjects who did not calculate because they considered it unnecessary ($N = 1,010$). Panel C reports subjects who did not calculate because it was too effortful ($N = 580$).

Subjects who reported making calculations in Risk-Free tasks do not exhibit anomalous patterns: they chose "Safe" in 54% of Unmixed tasks and exhibited a CRE of 2%-age points. Subjects who reported not making calculations because they were unnecessary exhibit more errors and systematic anomalies: they chose "Safe" in 64% of Unmixed tasks and exhibited a CRE of 5%-age points. Subjects who reported not making calculations because they required too much effort often erred in their Risk-Free choices, but the errors were not systematic: they chose "Safe" in 56% of Unmixed tasks and exhibited a CRE of –0.5%-age points.

A.9.3 Experiment 5: Shortcuts

In Experiment 5, after each block subjects also reported whether they had used mental short-cuts. In Block 1 Risk-Free tasks, 26% (151/577) reported doing so. Figure A18 presents the Risk-Free data separately for those who did and did not report shortcut use.

Subjects who did not report shortcuts show no anomalies: they choose "Safe" in 54% of Unmixed tasks and exhibit a CRE of 1%-age point. By contrast, subjects who reported shortcuts make more errors and display systematic anomalies: they choose "Safe" in 61%
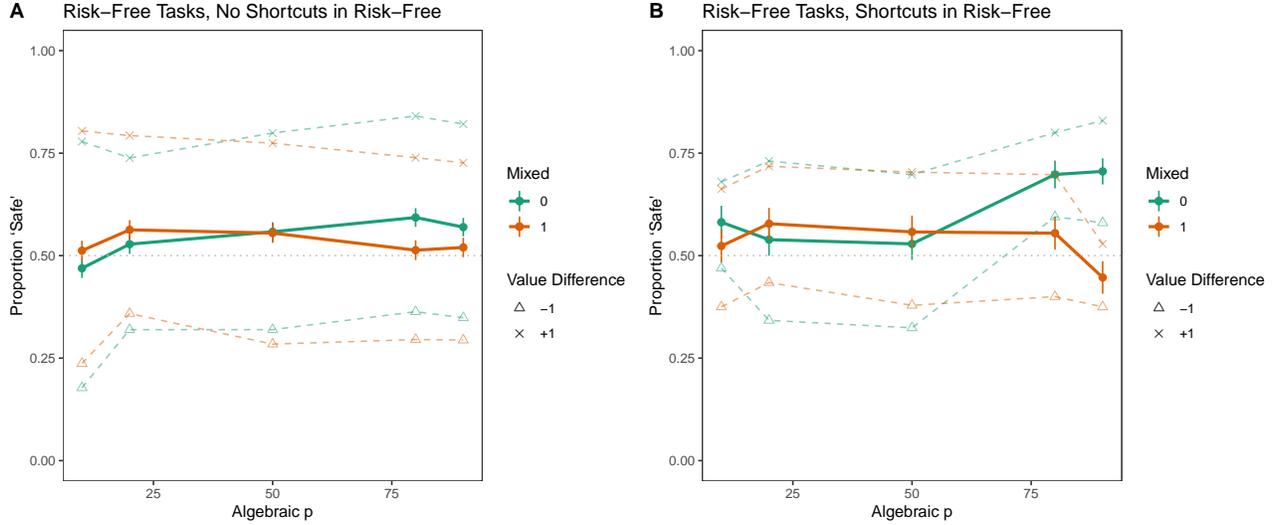
Figure A18: Experiment 5: Risk-Free Tasks and Mental Shortcuts, Block 1

Notes: Figure A18 shows the proportion of choices for the "Safe" option in Unmixed tasks (green) and Mixed tasks (orange). Light shaded exes and triangles denote +1 and −1 tasks, while dark shaded circles indicate sample averages. Bars represent ±1.96 standard errors (95% CIs). Panel A reports subjects who did not use mental shortcuts in Block 1 Risk-Free tasks of Experiment 5 ($N = 4{,}260$), and Panel B reports subjects who did ($N = 1{,}510$).

of Unmixed tasks and exhibit a CRE of 8%-age points. These subjects are thus plausible candidates for pattern matching. However, as noted in the text, they do not exhibit similar patterns in their Block 2 Risky behavior.

A.9.4 Experiment 5: Text Analysis of Subjects' Descriptions of their Shortcuts

In Experiment 5, when subjects reported using shortcuts, we asked them to describe their shortcuts using an example task.[69] To explore the nature of these shortcuts, we examined subjects' explanations. A striking feature of the text data for Risk-Free tasks is that subjects often referred to their attitudes toward risk when describing their shortcuts—either by characterizing the options as lotteries or by expressing uncertainty about the guaranteed payment. We refer to such descriptions in Risk-Free tasks as "risky shortcuts."

To identify risky shortcuts, we provided ChatGPT-5 Pro with the description of each shortcut reported after Block 1 and asked whether perceived risk or uncertainty in Risk-Free tasks affected subjects' decision-making.[70] ChatGPT-5 Pro could respond "Yes," "Uncertain," or "No," and we coded a shortcut as risky whenever the response was "Yes." Using this strategy,

---

[69] We had intended the example task to be one the subject had previously completed, but due to a coding error all subjects saw the same fixed task. As a result, for about half the sample the example task was one they had completed, and for the other half it was not.

[70] Appendix B.2.2 provides the prompt used to classify the 151 shortcuts reported in Block 1 Risk-Free tasks.

we find that 36% of subjects who described shortcuts in Block 1 Risk-Free tasks had risky shortcuts. In the main text, we show that subjects with inferred risky shortcuts are more likely to fail to maximize expected value in Risk-Free tasks, and more likely to exhibit small-stakes risk aversion and CRE. Figure A19 illustrates these results.
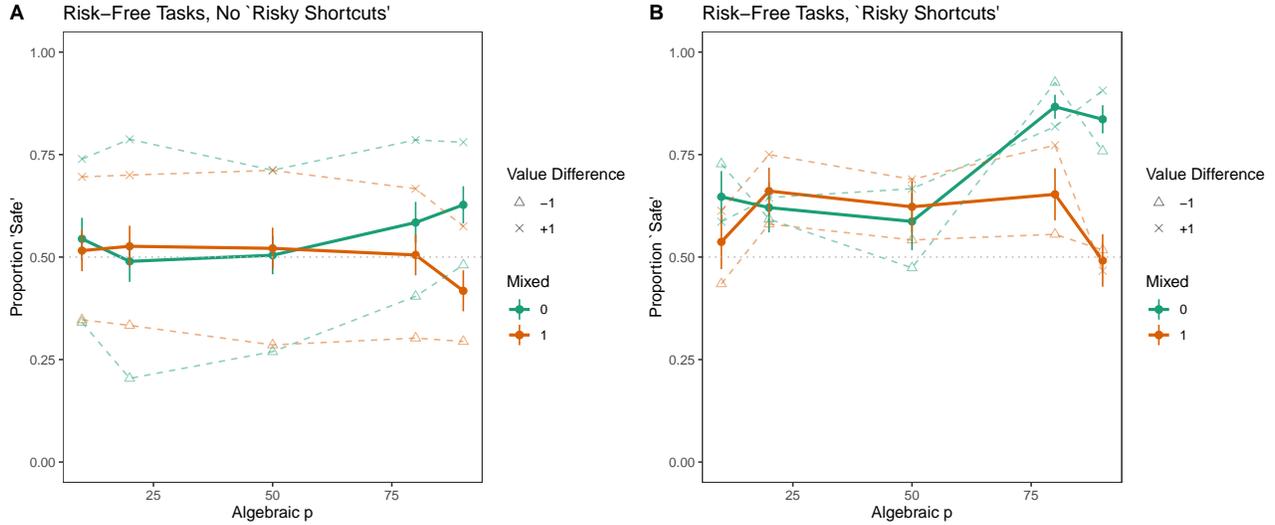
Figure A19: Experiment 5: Risk-Free Behavior and "Risky Shortcuts," Block 1

Notes: Figure A19 shows the proportion of choices for the "Safe" option in Unmixed (green) and Mixed tasks (orange). Light-shaded Xs and triangles denote +1 and −1 tasks, while dark-shaded circles indicate sample averages. Bars represent ±1.96 standard errors (95% CIs). Panel A reports subjects without an inferred "risky shortcut" from their Block 1 descriptions ($N = 960$), and Panel B reports subjects with an inferred "risky shortcut" ($N = 550$) in Experiment 5.

A.9.5   Experiment 5: Order Effects

Turning to behavior in the second block of Experiment 5, we find substantial order effects: completing Risky tasks in Block 1 influences subsequent Risk-Free behavior in Block 2, and vice versa. Figure A20 presents Block 2 results. Risky choice differs markedly by order. In Block 1, subjects show the standard patterns of small-stakes risk aversion and the CRE, but in Block 2 these behaviors are much less prevalent: subjects choose the Safe option in 66% of Unmixed tasks and exhibit a CRE of only 0.7%-age points. Risk-Free choice is also strongly affected by order. While subjects make objective errors in 29% of Block 1 Risk-Free tasks, the error rate rises to 50% in Block 2 ($t = 24.1$, $p < 0.01$). Despite the higher frequency of mistakes, Block 2 Risk-Free errors are largely unsystematic: subjects choose Safe in 55% of Unmixed tasks and exhibit a CRE of 3%-age points. These order effects suggest possible roles for decision fatigue and task contamination in blocked experimental designs.
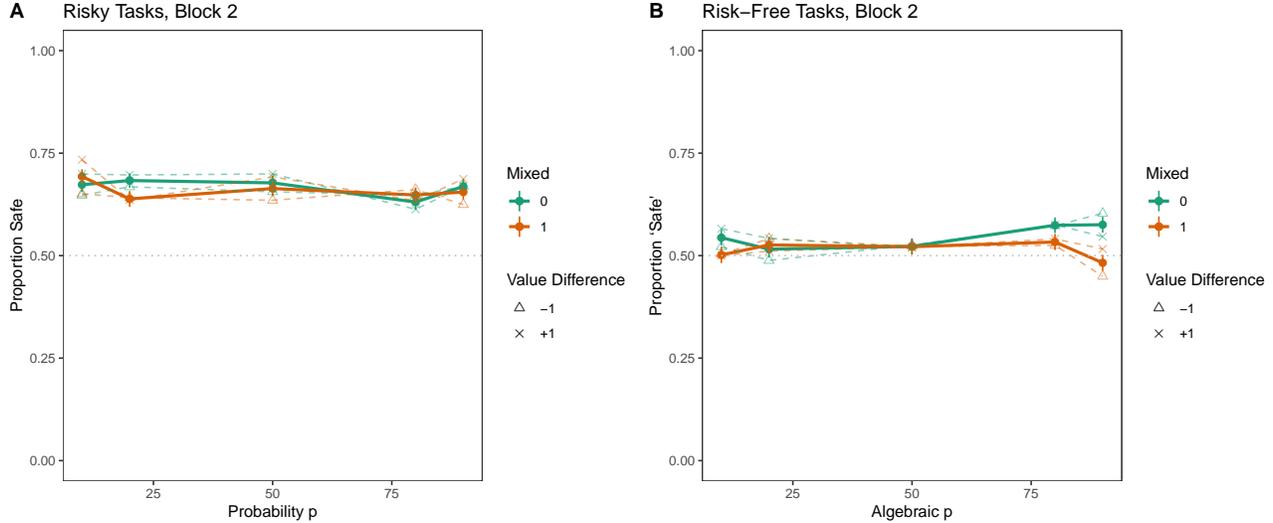
Figure A20: Experiment 5: Risky and Risk-Free Tasks, Block 2

Notes: Figure A20 shows the proportion of choices for the Safe/"Safe" option in Unmixed (green) and Mixed tasks (orange). Light shaded exes and triangles denote +1 and −1 tasks, while dark shaded circles indicate sample averages. Bars represent ±1.96 standard errors (95% CIs). Panel A reports Risky tasks ($N = 5{,}770$), and Panel B reports Risk-Free tasks ($N = 6{,}230$) for subjects in Block 2 of Experiment 5.

A.10   Additional Results from Experiment 6

We recruited 300 new subjects for Experiment 6. The design followed our Post-Choice condition except that, for every Risk-Free task, we explicitly asked subjects to calculate y and z (the latter for Mixed tasks).[71] This elicitation allows us to evaluate directly whether lack of mathematical proficiency can potentially account for pattern matching, and to do so for all subjects without conditioning on successful task recognition. Although we did not include characterization tables or reminders that Risk-Free tasks were risk-free, requiring calculations *only* in Risk-Free tasks may have both distinguished them from Risky tasks and clarified their risk-free nature.[72]

Figure A21 shows behavioral data from Experiment 6. In Risky tasks, subjects display standard patterns, choosing the Safe option in 70% of Unmixed cases and exhibiting a CRE of 10%-age points. In Risk-Free tasks, we find only limited evidence of pattern matching: Subjects choose Safe in 60% of Unmixed cases and exhibit a CRE of 7%-age points.

Subjects calculate y correctly in 91% of Unmixed tasks, and they calculate both y and z correctly in 85% of Mixed tasks. These findings confirm that calculation errors are relatively uncommon. Moreover, contrary to hypothesis IIIa, when calculations are incorrect, they do

---

[71]Each subject received a $4 completion payment. Twenty percent of subjects, chosen at random, also received a bonus payment following the same procedures as in our main study. The median completion time for Experiment 6 was 15.2 minutes. Full instructions appear in Appendices B.1.5 and B.1.19.

[72]Experiment 6 did not collect confidence judgments, so we cannot examine the relationship between confidence and Risky/Risk-Free choice as in the main experiment and Experiments 2 and 3.
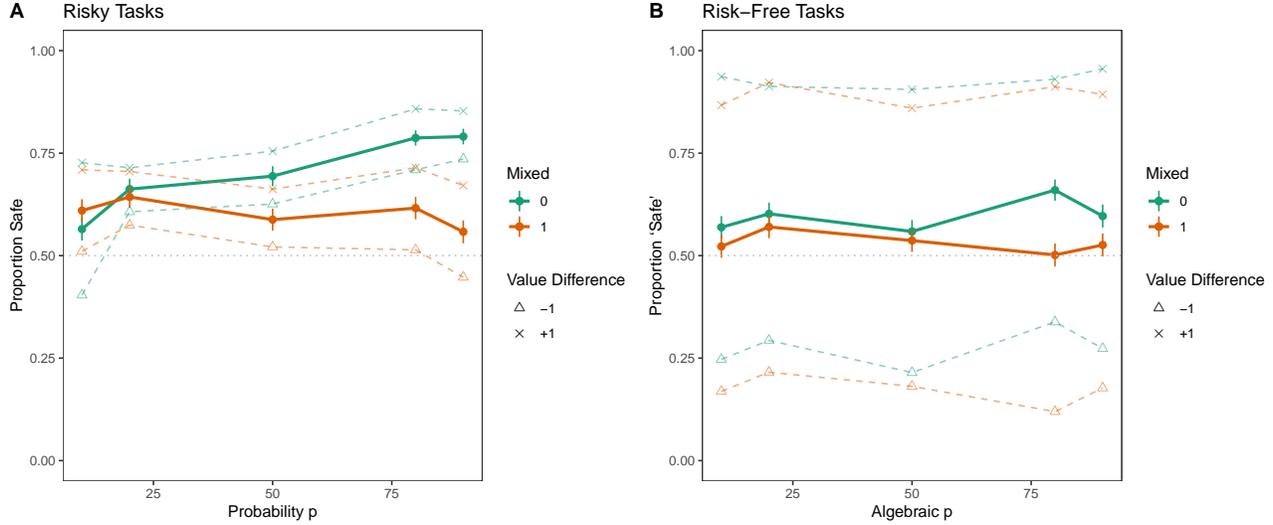
Figure A21: Experiment 6: Risky and Risk-Free Tasks

Notes: Figure A21 shows the proportion of choices for the Safe/"Safe" option in Unmixed (green) and Mixed tasks (orange). Light shaded exes and triangles denote +1 and −1 tasks, while dark shaded circles indicate sample averages. Bars represent ±1.96 standard errors (95% CIs). Panel A reports Risky tasks ($N = 2,977$) and Panel B Risk-Free tasks ($N = 3,023$) for subjects in Experiment 6.

not exhibit a systematic bias that accounts for the patterns of interest. We again use subjects' calculated values to simulate the Risk-Free choices of hypothetical subjects who make the same errors and always choose the option with the higher calculated value. We find no evidence of small-stakes "risk aversion" (45% "Safe" selections in Unmixed tasks; $t = 1.2$, $p = 0.23$ versus 50%) or a significant "CRE" (2%-age points; $t = 0.35$, $p = 0.73$ versus 0) in the simulated Risk-Free choices for which subjects committed calculation errors.

Subjects choose the option with the lower calculated value in 13% of all Risk-Free tasks, Unmixed and Mixed. Furthermore, the departures from calculated-value maximization exhibit a modest but systematic bias toward the "Safe" option in Unmixed Risk-Free tasks: When subjects' calculations show that Option A (which requires no calculation) has a higher value than Option B, the frequency with which they choose Option A is 94%. In contrast, when their calculations show that Option B has a higher value than Option A, the frequency with which they choose Option B is only 74% ($t = 11.1$, $p < 0.01$, test for equality). This differential creates the appearance of small-stakes "risk aversion." Additionally, while the same bias is present for Mixed Risk-Free tasks, it is smaller: The corresponding frequencies are 93% and 86% ($t = 4.6$, $p < 0.01$, test for equality). This differential creates the appearance of a "CRE." Figure A22 separates Risk-Free choice for these two groups. When subjects followed their calculations (87% of cases), errors and anomalous patterns are rare: Subjects chose "Safe" in 55% of Unmixed tasks and exhibited a CRE of 4%-age points. In contrast, when subjects did

not follow their calculations (13%), errors are frequent, proportions for $+1$ and -1 tasks are inverted, and anomalous patterns are evident: Subjects chose "Safe" in 82% of Unmixed tasks and exhibited a CRE of 15%-age points.

The patterns discussed above are plainly not attributable to calculation errors. Moreover, because this treatment forces subjects to incur the cognitive costs of calculating y and z, it removes the reason for deploying shortcuts.[73] Consequently, neither hypothesis II nor hypothesis III provides a plausible explanation for the resulting pattern matching. In contrast, hypothesis IVa, which posits aversion to the subjective uncertainty that calculations entail, easily rationalizes these patterns. Hypothesis Ia may also play a role if subjects continue to commit recognition failures despite performing the calculations needed for characterization.
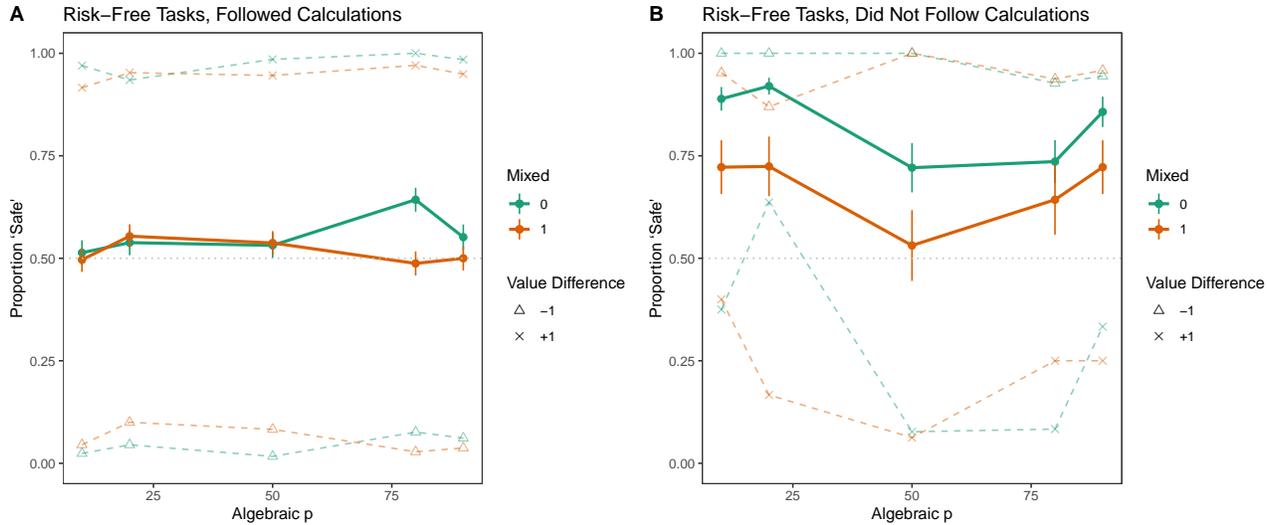
Figure A22: Experiment 6: Risk-Free Tasks and Following Calculations

Notes: Figure A22 shows the proportion of choices for the "Safe" option in Unmixed (green) and Mixed tasks (orange). Light shaded exes and triangles denote $+1$ and -1 tasks, while dark shaded circles indicate sample averages. Bars represent ± 1.96 standard errors (95% CIs). Panel A reports observations that followed their calculations by choosing the higher calculated value ($N = 2{,}629$), and Panel B reports observations that did not follow their calculations ($N = 394$) in Experiment 6.

A.11   Additional Details for Experiment 7

For Experiment 7 we recruited 375 subjects to make choices in four different types of tasks: 1) Unmixed Risky tasks following our main design; 2) Unmixed Risk-Free tasks where only Option B required calculation; 3) Unmixed Risk-Free tasks where only Option A required calculation;

---

[73]Technically, we only remove the reason for deploying shortcuts if subjects try to perform the calculation correctly. It is therefore notable that the same patterns are present for Risk-Free tasks with correctly calculated values.

and 4) Unmixed Risk-Free tasks where both Option A and Option B required calculation. For task types 2, 3, and 4, Table A3 presents the potential calculations required. The tasks were designed such that for every one-calculation task, there was a corresponding two-calculation task with the same required calculation for the same option.

Subjects faced a random selection of 20 tasks from a question bank of 50 tasks including the 40 tasks noted in Table A3 and the 10 Unmixed Risky tasks from our main design. Given that most of the tasks faced by subjects are Risk-Free, it is relevant to ensure that Risky behavior continues to deliver evidence of small stakes risk aversion. Indeed, 76% of Unmixed Risky choices are for the Safe alternative Option A, closely reproducing our prior results. Figure A23 presents the behavioral results for Experiment 7. Risky and Risk-Free choice data differ substantially; overall, subjects choose Option A in only 53% of Risk-Free tasks.

As discussed in the text, in Figure A23 one can see that options requiring calculation are modestly penalized: when only Option A needs to be calculated, subjects choose Option A in 52% of cases; and when only Option B needs to be calculated, subjects choose Option A in 55% of cases. The difference is more substantial when considering -1 tasks in isolation. There, when only Option A needs to be calculated, subjects choose Option A in 8% of cases; and when only Option B needs to be calculated, subjects choose Option A in 16% of cases.
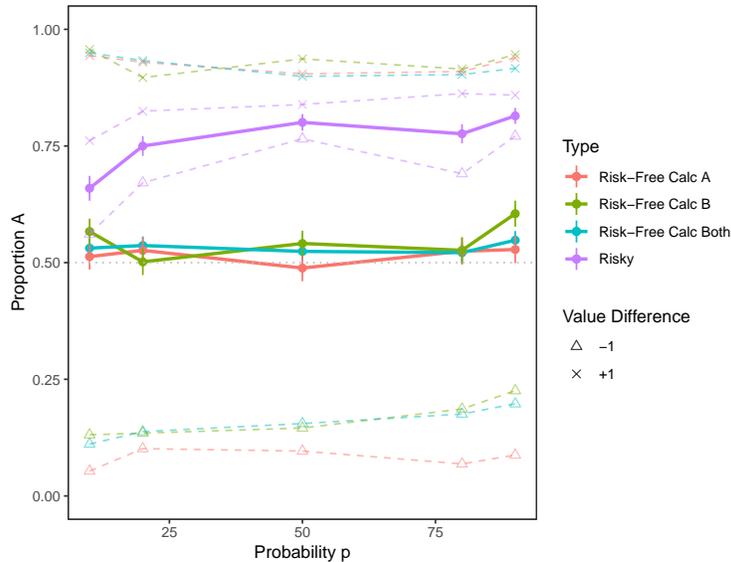
Figure A23: Experiment 7: Risky and Risk-Free Tasks

Notes: Figure A23 shows the proportion of choices for Option A in Unmixed tasks for Risky choice (purple) ($N = 1,463$), and Risk-Free choice with calculation required for Option A (pink) ($N = 1,487$), calculation required for Option B (green) ($N = 1,507$), or calculation required for both options (teal) ($N = 3,043$) for subjects in Experiment 7. Light shaded exes and triangles denote $+1$ and -1 tasks, while dark shaded circles indicate sample averages. Bars represent ± 1.96 standard errors (95% CIs).

Table A3: Experiment 7: Required Calculations for Risk-Free Tasks

Panel A: Option B Requires Calculation

	$p = 0.1$		$p = 0.2$		$p = 0.5$		$p = 0.8$		$p = 0.9$
Option A	Option B	Option A	Option B	Option A	Option B	Option A	Option B	Option A	Option B
2	$0.9*0+0.1*30$	5	$0.8*0+0.2*30$	14	$0.5*10+0.5*20$	23	$0.2*40+0.8*20$	26	$0.1*0+0.9*30$
4	$0.9*0+0.1*30$	7	$0.8*0+0.2*30$	16	$0.5*10+0.5*20$	25	$0.2*40+0.8*20$	28	$0.1*0+0.9*30$

Panel B: Option A Requires Calculation

	$p = 0.1$		$p = 0.2$		$p = 0.5$		$p = 0.8$		$p = 0.9$
Option A	Option B	Option A	Option B	Option A	Option B	Option A	Option B	Option A	Option B
$0.9*0+0.1*30$	2	$0.8*0+0.2*30$	5	$0.5*10+0.5*20$	14	$0.2*40+0.8*20$	23	$0.1*0+0.9*30$	26
$0.9*0+0.1*30$	4	$0.8*0+0.2*30$	7	$0.5*10+0.5*20$	16	$0.2*40+0.8*20$	25	$0.1*0+0.9*30$	28

Panel C: Both Options Require Calculation

	$p = 0.1$		$p = 0.2$		$p = 0.5$		$p = 0.8$		$p = 0.9$
Option A	Option B	Option A	Option B	Option A	Option B	Option A	Option B	Option A	Option B
$0.9*0+0.1*20$	$0.9*0+0.1*30$	$0.8*5+0.2*5$	$0.8*0+0.2*30$	$0.5*2+0.5*26$	$0.5*10+0.5*20$	$0.2*75+0.8*10$	$0.2*40+0.8*20$	$0.1*80+0.9*20$	$0.1*0+0.9*30$
$0.9*0+0.1*40$	$0.9*0+0.1*30$	$0.8*5+0.2*15$	$0.8*0+0.2*30$	$0.5*2+0.5*30$	$0.5*10+0.5*20$	$0.2*85+0.8*10$	$0.2*40+0.8*20$	$0.1*100+0.9*20$	$0.1*0+0.9*30$
$0.9*0+0.1*30$	$0.9*0+0.1*20$	$0.8*0+0.2*30$	$0.8*5+0.2*5$	$0.5*10+0.5*20$	$0.5*2+0.5*26$	$0.2*40+0.8*20$	$0.2*75+0.8*10$	$0.1*0+0.9*30$	$0.1*80+0.9*20$
$0.9*0+0.1*30$	$0.9*0+0.1*40$	$0.8*0+0.2*30$	$0.8*5+0.2*15$	$0.5*10+0.5*20$	$0.5*2+0.5*30$	$0.2*40+0.8*20$	$0.2*85+0.8*10$	$0.1*0+0.9*30$	$0.1*100+0.9*20$

B   Supplement: Experimental Instructions and Machine Learning Prompts

B.1   Experimental Instructions

B.1.1   Instructions: All Conditions in the Main Experiment and in Experiment 2

We report the screenshots that apply to all conditions in the main experiment. The instructions for Experiment 2 are identical, except that in the experimental screen titled "This Study," we inform subjects that they will complete 10 rather than 20 tasks.

Hello and Welcome

Welcome, and thanks for your participation!

We are researchers at California Institute of Technology inviting you to participate in a research study. The study should take approximately 37 minutes. Please click to review information about the study and to give your consent to participate.

<div>Next</div>

Prolific ID

You will be paid via Prolific for your participation in this study. In order to pay you, we need your Prolific ID.

Please enter your Prolific ID:

<div>Next</div>

Possible Rewards

You will receive $7.5 if you complete the entire study.

In addition to this payment, you have a 1 in 5 chance of being eligible to receive a bonus payment. The smallest possible bonus payment is $0 and the largest possible bonus payment is $30.

You will be informed of how your decisions will influence your bonus payment if you are eligible to receive one.

<div>Next</div>

This Study

In this study, you will complete 20 tasks, each of which involves your preferences over two options. The first option will always be called **Option A**. The second option will always be called **Option B**.

Option A and Option B refer to the possibility of winning monetary amounts ranging from $0 to $30 with some fixed chances. In each task you are asked to answer a few questions about Option A and Option B, and to decide which option you prefer.

Next

Example Options

We will now introduce you to the types of options that you will see in this study and the choices you will make.

Next

Example Options 1

Consider the following example:

Option A	Option B
Draw one ticket from this bag:	Draw one ticket from this bag:

100 blue tickets (❙) which pay **$4**

10 purple tickets (❙) which pay **$28**
90 gold tickets (❙) which pay **$2**

In this example:
- Option A allows you to draw one ticket from a bag containing 100 blue tickets (❙). Each blue ticket pays $4.
- Option B allows you to draw one ticket from a bag containing 10 purple tickets (❙) and 90 gold tickets (❙). Each purple ticket pays $28, and each gold ticket pays $2.

Therefore,
- Option A allows you to win $4 with a 100 in 100 chance.
- Option B allows you to win $28 with a 10 in 100 chance, or $2 with a 90 in 100 chance.

Next

Example Options 2

Consider the following example:

<div>

Option A
Draw one ticket from this bag:

100 ❙

100 blue tickets (❙) which pay **$13**

Option B
Draw one ticket from this bag:

50 ❙
50 ❙

50 purple tickets (❙) which pay **$y**

50 gold tickets (❙) which pay **$y**

$$y=(0.5\times2)+(0.5\times28)$$

</div>

In this example:
- Option A allows you to draw one ticket from a bag containing 100 blue tickets (❙). Each blue ticket pays $13.
- Option B allows you to draw one ticket from a bag containing 50 purple tickets (❙) and 50 gold tickets (❙). Each purple ticket pays $y and each gold ticket pays $y. $y is an amount equal to (0.5 × 2) + (0.5 × 28) that you can calculate. Therefore, in this example, y is equal to 15.

Therefore,
- Option A allows you to win $13 with a 100 in 100 chance.
- Option B allows you to win $y with a 100 in 100 chance.

Next

How will you make choices?

In each task, we will show you two options and will ask you to choose between the following two answer choices:

1. I prefer Option A
2. I prefer Option B

In each decision, after you choose which option you prefer, we will ask you how confident you are about your choice.
You will report your confidence on a scale from 0 to 100, where 0 indicates that you are *not confident at all* and 100 indicates that you are *completely confident*.

Next

Familiarize Yourself with the Study

- In each task, before you make your choices, we will also ask several questions to check your understanding of the task.
- We now ask you to complete two training tasks to familiarize yourself with these types of questions and the choices you will make in the study.
- As you will realize from the training tasks, in this study after you confirm an answer *you will not be able to modify it*.
- For this reason, we ask you to think carefully before confirming your answers. After the training, we will explain how your answers in the study will affect your bonus.

Next

Training Task 1 of 2
Familiarize Yourself with Option A

Option A	**Option B**
Draw one ticket from this bag:	Draw one ticket from this bag:

100 blue tickets (❘) which pay **$4**

10 purple tickets (❘) which pay **$28**
90 gold tickets (❘) which pay **$2**

Assign to each ticket the corresponding <u>chance</u> under **Option A**:

Option A

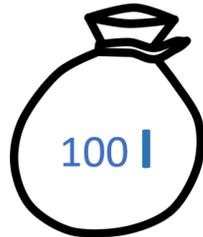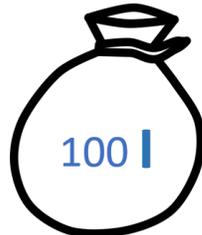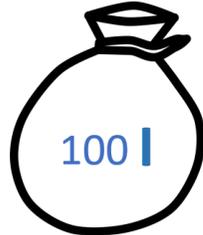| Ticket	❘	❘	❘
Chance	☐ in 100	☐ in 100	☐ in 100

Next

4

Training Task 1 of 2
Familiarize Yourself with Option A

Option A
Draw one ticket from this bag:

100 ▎

100 blue tickets (▎) which pay **$4**

Option B
Draw one ticket from this bag:

10 ▎
90 ▎

10 purple tickets (▎) which pay **$28**
90 gold tickets (▎) which pay **$2**

Assign to each ticket the corresponding <u>value</u> under **Option A**:

Remark: <u>we only ask you to report the values of the tickets that have a positive chance of being drawn *according to your answers*.</u>

Option A

Ticket	▎	▎	▎
Chance	100 in 100	0 in 100	0 in 100
Value	$	$	$

Next

5

Familiarize Yourself with Option B

Option A	**Option B**
Draw one ticket from this bag:	Draw one ticket from this bag:

100 |

10 |
90 |

100 blue tickets (|) which pay **$4**

10 purple tickets (|) which pay **$28**

90 gold tickets (|) which pay **$2**

Assign to each ticket the corresponding <u>chance</u> under **Option B**:

Option B

Ticket				
Chance	in 100	in 100	in 100

Next

6

Training Task 1 of 2
Familiarize Yourself with Option B

Option A

Draw one ticket from this bag:

100 ▌

100 blue tickets (▌) which pay **$4**

Option B

Draw one ticket from this bag:

10 ▌
90 ▌

10 purple tickets (▌) which pay **$28**

90 gold tickets (▌) which pay **$2**

Assign to each ticket the corresponding <u>value</u> under **Option B**:

Remark: <u>we only ask you to report the values of the tickets that have a positive chance of being drawn *according to your answers.*</u>

Option B

Ticket	▌	▌	▌
Chance	0 in 100	10 in 100	90 in 100
Value	$	$	$

Next

Make your Decision

Option A	Option B
Draw one ticket from this bag:	Draw one ticket from this bag:

100 blue tickets (|) which pay **$4**

10 purple tickets (|) which pay **$28**
90 gold tickets (|) which pay **$2**

Your previous answers about chances and values of tickets under Option A and Option B are:

Option A			
Ticket	**|**	**|**	**|**
Chance	100 in 100	0 in 100	0 in 100
Value	$ 4	$	$

Option B			
Ticket	**|**	**|**	**|**
Chance	0 in 100	10 in 100	90 in 100
Value	$	$ 28	$ 2

Which option do you **prefer**?

I prefer Option A ○   I prefer Option B ○

Next

Option A

Draw one ticket from this bag:

100 |

100 blue tickets (|) which pay **$4**

Option B

Draw one ticket from this bag:

10 |
90 |

10 purple tickets (|) which pay **$28**

90 gold tickets (|) which pay **$2**

Your previous answers about chances and values of tickets under Option A and Option B are:

Option A

Ticket	**|**	**|**	**|**
Chance	100 in 100	0 in 100	0 in 100
Value	$ 4	$	$

Option B

Ticket	**|**	**|**	**|**
Chance	0 in 100	10 in 100	90 in 100
Value	$	$ 28	$ 2

You chose **Option B**.

On a scale from 0 (not confident at all) to 100 (completely confident), how **confident** do you feel about this choice?

The higher the number, the more confident you are about this choice.

Not Confident
At all

Completely
Confident

Next

Familiarize Yourself with Option A

Option A
Draw one ticket from this bag:

100 ▌

100 blue tickets (▌) which pay **$13**

Option B
Draw one ticket from this bag:

50 ▌
50 ▌

50 purple tickets (▌) which pay **$y**

50 gold tickets (▌) which pay **$y**

y=(0.5×2)+(0.5×28)

Assign to each ticket the corresponding <u>chance</u> under **Option A**:

Option A

Ticket	▌	▌	▌
Chance	in 100	in 100	in 100

Next

Familiarize Yourself with Option A

Option A
Draw one ticket from this bag:

100 |

100 blue tickets (**|**) which pay **$13**

Option B
Draw one ticket from this bag:

50 |
50 |

50 purple tickets (**|**) which pay **$y**
50 gold tickets (**|**) which pay **$y**

y=(0.5×2)+(0.5×28)

Assign to each ticket the corresponding <u>value</u> under **Option A**:

Remark: <u>we only ask you to report the values of the tickets that have a positive chance of being drawn *according to your answers.*</u>

Option A

Ticket	**|**	**|**	**|**
Chance	100 in 100	0 in 100	0 in 100
Value	$	$	$

Next

Familiarize Yourself with Option B

Option A	**Option B**
Draw one ticket from this bag:	Draw one ticket from this bag:

100 blue tickets (**|**) which pay **$13**

50 purple tickets (**|**) which pay **$y**

50 gold tickets (**|**) which pay **$y**

$$y=(0.5 \times 2)+(0.5 \times 28)$$

Assign to each ticket the corresponding <u>chance</u> under **Option B**:

Option B

Ticket	**|**	**|**	**|**
Chance	in 100	in 100	in 100

Next

Familiarize Yourself with Option B

Option A	**Option B**
Draw one ticket from this bag:	Draw one ticket from this bag:

100 **|**

50 **|**
50 **|**

100 blue tickets (**|**) which pay **$13**

50 purple tickets (**|**) which pay **$y**

50 gold tickets (**|**) which pay **$y**

$$y=(0.5\times2)+(0.5\times28)$$

Assign to each ticket the corresponding <u>value</u> under **Option B**:

Remark: <u>we only ask you to report the values of the tickets that have a positive chance of being drawn *according to your answers.*</u>

Option B

Ticket	**|**	**|**	**|**
Chance	0 in 100	50 in 100	50 in 100
Value	$	$	$

Next

Make your Decision

Option A	Option B
Draw one ticket from this bag:	Draw one ticket from this bag:

100 blue tickets (❙) which pay **$13**

50 purple tickets (❙) which pay **$y**

50 gold tickets (❙) which pay **$y**

$$y=(0.5\times 2)+(0.5\times 28)$$

Your previous answers about chances and values of tickets under Option A and Option B are:

Option A

Ticket	❙	❙	❙
Chance	100 in 100	0 in 100	0 in 100
Value	$ 13	$	$

Option B

Ticket	❙	❙	❙
Chance	0 in 100	50 in 100	50 in 100
Value	$	$ 15	$ 15

Which option do you **prefer**?

I prefer Option A ○

I prefer Option B ○

Next

Express your Confidence

Option A	**Option B**
Draw one ticket from this bag:	Draw one ticket from this bag:

100 ▍	50 ▍
	50 ▍

100 blue tickets (▍) which pay **$13**

50 purple tickets (▍) which pay **$y**

50 gold tickets (▍) which pay **$y**

$$y=(0.5 \times 2)+(0.5 \times 28)$$

Your previous answers about chances and values of tickets under Option A and Option B are:

Option A			
Ticket	▍	▍	▍
Chance	100 in 100	0 in 100	0 in 100
Value	$ 13	$	$

Option B			
Ticket	▍	▍	▍
Chance	0 in 100	50 in 100	50 in 100
Value	$	$ 15	$ 15

You chose **Option B**.

On a scale from 0 (not confident at all) to 100 (completely confident), how **confident** do you feel about this choice?

The higher the number, the more confident you are about this choice.

Not Confident At all ————————————————— Completely Confident

Next

Familiarize Yourself with the Study

- We now ask you to complete two training tasks to familiarize yourself with the questions and the choices you will make in the study.
- As you will realize from the training task, in this study after you confirm an answer *you will not be able to modify it*.
- For this reason, we ask you to think carefully before confirming your answers. After the training, we will explain how your answers in the study will affect your bonus.

Next

Training Task 1 of 2
Make your Decision

Option A
Draw one ticket from this bag:

100 ▍

Option B
Draw one ticket from this bag:

10 ▍
90 ▍

100 blue tickets (▍) which pay **$4**

10 purple tickets (▍) which pay **$28**

90 gold tickets (▍) which pay **$2**

Which option do you **prefer**?

I prefer Option A   I prefer Option B
        ○   ○

Next

16

Express your Confidence

Option A
Draw one ticket from this bag:

100 |

100 blue tickets (|) which pay **$4**

Option B
Draw one ticket from this bag:

10 |
90 |

10 purple tickets (|) which pay **$28**

90 gold tickets (|) which pay **$2**

You chose **Option A**.
On a scale from 0 (not confident at all) to 100 (completely confident), how **confident** do you feel about this choice?

The higher the number, the more confident you are about this choice.

Not Confident
At all

Completely
Confident

Next

Make your Decision

Option A	**Option B**
Draw one ticket from this bag:	Draw one ticket from this bag:
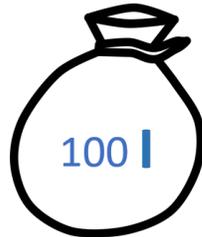
100 blue tickets (|) which pay **$13**

50 purple tickets (|) which pay **$y**

50 gold tickets (|) which pay **$y**

$$y=(0.5\times2)+(0.5\times28)$$

Which option do you **prefer**?

I prefer Option A   I prefer Option B
   ○   ○

Next

Training Task 2 of 2
Express your Confidence

Option A

Draw one ticket from this bag:

100 |

100 blue tickets (|) which pay **$13**

Option B

Draw one ticket from this bag:

50 |
50 |
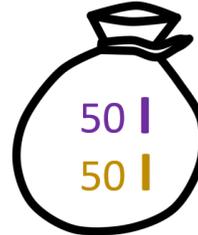
50 purple tickets (|) which pay **$y**

50 gold tickets (|) which pay **$y**

y=(0.5×2)+(0.5×28)

You chose **Option B**.
On a scale from 0 (not confident at all) to 100 (completely confident), how **confident** do you feel about this choice?

The higher the number, the more confident you are about this choice.

Not Confident
At all

Completely
Confident

Next

Screenshots of the choice tasks from the main experiment are omitted, as they are identical to the training tasks.

Bonus Payment

Congratulations, you filled out the tables correctly in the two training tasks! You are now ready to learn more about your potential bonus payment.

We call **potential bonus payment** an additional amount of money that you may receive at the end of the study on top of your participation fee. The potential bonus payment will be paid to you **only** if some conditions are met.

We first explain how the potential bonus payment is computed. Next, we explain the conditions under which you will receive the potential bonus payment.

Next

Determining Your Potential Bonus Payment

To determine your potential bonus payment, we will randomly select one of the tasks of the study to be the **task-that-counts**. For example, the task-that-counts could be the one illustrated below.

Option A	Option B
Draw one ticket from this bag:	Draw one ticket from this bag:

100 blue tickets (❙) which pay **$4**

10 purple tickets (❙) which pay **$28**

90 gold tickets (❙) which pay **$2**

If you preferred **Option A** in the task-that-counts, then Option A would be implemented and we would draw a single ticket from that bag.

- A blue ticket would definitely be chosen, so your potential bonus payment would be $4.

If you preferred **Option B** in the task-that-counts, then Option B would be implemented and we would draw a single ticket from that bag.

- If a purple ticket was drawn, then your potential bonus payment would be $28.
- If a gold ticket was drawn, then your potential bonus payment would be $2.

Because every task could be the task-that-counts, you should make each choice as if it is the one that could determine your potential bonus payment.

Next

Conditions for Your Bonus Payment

You will receive your potential bonus payment at the end of the study **only if** two conditions are satisfied:

1. After we compute your potential bonus payment, we will draw a number between 1 and 5. If the drawn number is 1, you will be eligible for the potential bonus payment. Otherwise, you will not receive a bonus payment.
2. If you are eligible for a bonus payment, then we will check your answers about the chances and values of the tickets under Option A and Option B that you filled out in the tables in the task-that-counts. If every one of your answers is correct then you will be paid your potential bonus payment. If any one of your answers is incorrect, then you will not receive a bonus payment.

Next

Conditions for Your Bonus Payment - Example

Suppose again that the task-that-counts is the one illustrated below.

Option A	Option B
Draw one ticket from this bag:	Draw one ticket from this bag:
100 ❙	10 ❙ 90 ❙
100 blue tickets (❙) which pay **$4**	**10** purple tickets (❙) which pay **$28** **90** gold tickets (❙) which pay **$2**
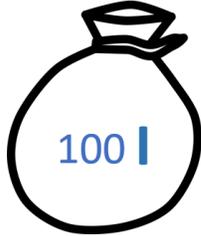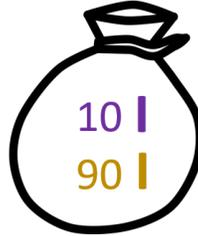
Imagine that you preferred Option B and a purple ticket is drawn so that your potential bonus payment is $28. Moreover, imagine that you are eligible to receive this potential bonus payment.

As described in the previous pages, you will have reported your answers about chances and values of tickets under Option A and Option B in every task by filling the following tables:

	Option A					Option B		
Ticket	❙	❙	❙		**Ticket**	❙	❙	❙
Chance	in 100	in 100	in 100		**Chance**	in 100	in 100	in 100
Value	$	$	$		**Value**	$	$	$

You will need to have filled out the tables correctly in order to receive your potential bonus payment. Please proceed to learn more about this.

Next

Conditions for Your Bonus Payment - Example

Suppose again that the task-that-counts is the one illustrated below.

Option A	**Option B**
Draw one ticket from this bag:	Draw one ticket from this bag:

100 ▌

10 ▌
90 ▌
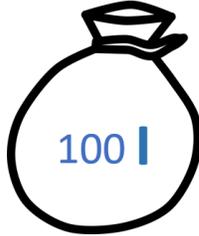
100 blue tickets (▌) which pay **$4**
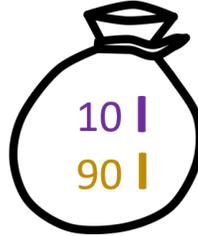
10 purple tickets (▌) which pay **$28**

90 gold tickets (▌) which pay **$2**

Imagine that you preferred Option B and a purple ticket is drawn so that your potential bonus payment is $28. Moreover, imagine being eligible to receive this potential bonus payment.

As described in the previous pages, you will report your answers about chances and values of tickets under Option A and Option B in every task by filling the following tables:

Option A

Ticket	▌	▌	▌
Chance	100 in 100	0 in 100	0 in 100
Value	$ 4	$	$

Option B

Ticket	▌	▌	▌
Chance	0 in 100	10 in 100	90 in 100
Value	$	$ 28	$ 2

If all your answers are correct as in the two tables above, you will receive your potential bonus payment of $28 at the end of the study. Otherwise, you will not receive a bonus payment.

Next

Test your Understanding
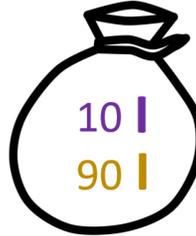
Suppose that this is the task-that-counts:

Option A	**Option B**
Draw one ticket from this bag:	Draw one ticket from this bag:

100 blue tickets (|) which pay **$4**

10 purple tickets (|) which pay **$28**

90 gold tickets (|) which pay **$2**

Here are a few questions to test your understanding:

How do we determine your potential bonus payment?
- ○ Draw a ticket from the bag corresponding to the option that you preferred
- ○ Always draw a ticket from the bag corresponding to Option A
- ○ Always draw a ticket from the bag corresponding to Option B
- ○ Draw a ticket from the bag corresponding to Option A with a 50 in 100 chance or from the bag corresponding to Option B with a 50 in 100 chance

What is your potential bonus payment if you preferred Option B and a purple ticket is drawn?
- ○ 2
- ○ 28
- ○ 15
- ○ 4

Under what condition will you receive the potential bonus payment at the end of the experiment?
- ○ I will receive the potential bonus payment only if I am eligible
- ○ I will receive the potential bonus payment only if all my answers in the tables of chances and values for the task-that-counts are correct
- ○ I will receive the potential bonus payment only if I am eligible and all my answers in the tables of chances and values for the task-that-counts are correct
- ○ I will always receive the potential bonus payment

What is the chance that you are eligible for a potential bonus payment?
- ○ 1 in 4
- ○ 1 in 2
- ○ 1 in 3
- ○ 1 in 5

B.1.5 Bonus Payment Instructions: Experiment 2, Experiment 3, Experiment 6, Experiment 7 and Pre-Choice Unincentivized and Post-Choice Conditions in the Main Experiment

Bonus Payment

Congratulations, you completed the two training tasks! You are now ready to learn more about your potential bonus payment.

We call **potential bonus payment** an additional amount of money that you may receive at the end of the study on top of your participation fee. The potential bonus payment will be paid to you **only** if some conditions are met.

We first explain how the potential bonus payment is computed. Next, we explain the conditions under which you will receive the potential bonus payment.

Next

Determining Your Potential Bonus Payment

To determine your potential bonus payment, we will randomly select one of the tasks of the study to be the **task-that-counts**. For example, the task-that-counts could be the one illustrated below.

Option A	**Option B**
Draw one ticket from this bag:	Draw one ticket from this bag:

100 blue tickets (❘) which pay **$4**

10 purple tickets (❘) which pay **$28**

90 gold tickets (❘) which pay **$2**

If you preferred **Option A** in the task-that-counts, then Option A would be implemented and we would draw a single ticket from that bag.

- A blue ticket would definitely be chosen, so your potential bonus payment would be $4.

If you preferred **Option B** in the task-that-counts, then Option B would be implemented and we would draw a single ticket from that bag.

- If a purple ticket was drawn, then your potential bonus payment would be $28.
- If a gold ticket was drawn, then your potential bonus payment would be $2.

Because every task could be the task-that-counts, you should make each choice as if it is the one that could determine your potential bonus payment.

Next

Conditions for Your Bonus Payment

- You will receive your potential bonus payment at the end of the study **only if** you are eligible to receive one.
- To determine whether you are eligible, we will draw a number between 1 and 5.
- If the drawn number is 1, you will be paid your potential bonus payment.
- Otherwise, you will not receive a bonus payment.

Next

Test your Understanding

Suppose that this is the task-that-counts:

Option A
Draw one ticket from this bag:

100 █

100 blue tickets (█) which pay **$4**

Option B
Draw one ticket from this bag:

10 █
90 █

10 purple tickets (█) which pay **$28**

90 gold tickets (█) which pay **$2**

Here are a few questions to test your understanding:

How do we determine your potential bonus payment?
- ◯ Draw a ticket from the bag corresponding to the option that you preferred
- ◯ Always draw a ticket from the bag corresponding to Option A
- ◯ Always draw a ticket from the bag corresponding to Option B
- ◯ Draw a ticket from the bag corresponding to Option A with a 50 in 100 chance or from the bag corresponding to Option B with a 50 in 100 chance

What is your potential bonus payment if you preferred Option B and a purple ticket is drawn?
- ◯ 2
- ◯ 28
- ◯ 15
- ◯ 4

Under what condition will you receive the potential bonus payment at the end of the experiment?
- ◯ I will receive the potential bonus payment only if I am eligible. I have a 1 in 4 chance of being eligible.
- ◯ I will receive the potential bonus payment only if I am eligible. I have a 1 in 2 chance of being eligible.
- ◯ I will receive the potential bonus payment only if I am eligible. I have a 1 in 5 chance of being eligible.
- ◯ I will receive the potential bonus payment only if I am eligible. I have a 1 in 3 chance of being eligible.

Next

B.1.6 Additional Instructions and Tasks for Experiment 2 (All Conditions)

Understanding your Overall Confidence

We asked you to report your overall confidence in each of the ten choices you made in this study so far. In some of your choices, you reported a confidence level of less than 100, indicating that you had less-than-complete confidence. Next, we will ask you to explain some of the possible reasons for your less-than-complete overall confidence level in 6 choices.

Next

Understanding your Overall Confidence

In the next pages, we list nine possible reasons for having less-than-complete confidence in your decisions. For each reason, we also provide an illustration. It's important to keep in mind that these are just illustrations. In a few minutes we will ask you to classify similar examples, so please review these reasons carefully.

All these examples concern an individual named Chris who must choose between attending two social events with two different groups of friends.

Next

Understanding your Overall Confidence

1. Not Sure What Each Option Means: You lack confidence in your decision because you're not sure you interpreted the available options correctly or understood their consequences.

Here's an example of why Chris might report less than complete confidence for this reason:

Chris was told weeks ago which friends were attending which event, but he found some of their messages confusing and isn't sure he understood them properly. He might express less-than-complete confidence in his decision because he isn't sure he understands the consequences of choosing to attend one event vs. the other.

Next

28

Understanding your Overall Confidence

2. Multiple Views About What You Want: You lack confidence in your decision because, in your view, there's probably more than one right way to think about what you want to achieve.

Here's an example of why Chris might report less than complete confidence for this reason:

Chris values experiences with different friends for different reasons and, after careful consideration, doesn't think one set of reasons is necessarily more or less important than another. He might express less-than-complete confidence in his decision because he doesn't think there's just one right way to think about what he wants from either social event.

<div align="center">

`Next`

</div>

Understanding your Overall Confidence

3. Don't Know How You'll Feel: You lack confidence in your decision because you aren't sure how you will feel about each of the possible outcomes.

Here's an example of why Chris might report less than complete confidence for this reason:

Chris enjoys different friends in different moods and is uncertain what his mood will be when the event arrives. He might express less-than-complete confidence in his decision because he isn't sure how he will feel about being with each group of friends.

<div align="center">

`Next`

</div>

Understanding your Overall Confidence

4. Didn't Apply Your Decision Criterion Correctly: You lack confidence in your decision because you think the criterion you decided to use and any simplifying assumptions you made may actually favor an alternative other than the alternative that you chose.

Here's an example of why Chris might report less than complete confidence for this reason:

To simplify his choice, Chris assumed he would have the most fun at the event with the largest number of friends, so he listed the friends planning to attend each event and counted them. Right after committing to the event with the larger count, he worried that he might have miscounted. He might express less-than-complete confidence in his decision because he now thinks the criterion he decided to use and the simplifying assumption he made might actually favor an alternative other than the one he chose.

<div align="center">

`Next`

</div>

Understanding your Overall Confidence

5. Didn't Think Carefully About What You Want: You lack confidence in your decision because you haven't thought through what you want to achieve as carefully as you could.

Here's an example of why Chris might report less than complete confidence for this reason:

Chris values experiences with different friends for different reasons but hasn't seriously considered whether one set of reasons is more important than another. He might express less-than-complete confidence in his decision because he hasn't carefully thought through what he wants from this social event.

Next

Understanding your Overall Confidence

6. It's Nearly a Toss-Up: You lack confidence in your decision because you think the choice is close to a toss-up.

Here's an example of why Chris might report less than complete confidence for this reason:

After assessing the desirability of each event, Chris concludes that he would enjoy them about the same. However, because he is slightly unsure about each assessment, either one might be slightly better. He might express less-than-complete confidence in his decision because he thinks it is nearly a toss-up.

Next

Understanding your Overall Confidence

7. Worried About Mental Shortcuts: You lack confidence in your decision because you're worried about having used a mental shortcut that might have given the wrong answer.

Here's an example of why Chris might report less than complete confidence for this reason:

Chris' enjoyment of social events depends on who else is there, but it's hard for him to think through the pluses and minuses of being with one large group of friends rather than another. Instead, he makes his decision based entirely on which event his friend Parker plans to attend. He might express less-than-complete confidence in his decision because he's worried that the mental shortcut he used — to only think about Parker — isn't a good one.

Next

Understanding your Overall Confidence

8. Can't Know For Sure What You'll Get: You lack confidence in your decision because, based on the information you've received, you can't be sure what you will receive given your chosen option.

Here's an example of why Chris might report less than complete confidence for this reason:

Some of Chris's friends told him they might attend his chosen event or they might just stay home. He might express less-than-complete confidence in his decision because he can't be sure what his chosen option will yield since he doesn't know which friends will actually turn up to the event.

Next

Understanding your Overall Confidence

9. Other Reasons: You lack confidence in your decision for other reasons.

Here's an example of why Chris might report less than complete confidence for this reason:

Chris may have other reasons for expressing less-than-complete confidence in his decision that we haven't mentioned.

Next

Explaining the Reasons for your Overall Confidence

When you reported your overall confidence in each choice, any answer below 100 indicated that you had less than complete confidence. We want to understand your reasons for reporting less-than-complete confidence.

For each reason on the previous screens, you will indicate the degree to which it accounted for your less-than-complete confidence. For example, if you rated your confidence as 60, we want to know the extent to which each reason was responsible for you answering 60 rather than 100. In each case, you will answer on a scale of **1 (Very Little)** to **7 (Very Much)**. So if you rated your overall confidence as 60, you would give a 1 for a particular reason if it had very little to do with saying your confidence was 60 rather than 100, and you would give a 7 if it had a great deal to do with saying your confidence was 60 rather than 100.

Please proceed to see an example of how the questions will look.

Next

Example

Imagine your overall confidence in a choice was 60 rather than 100. You would then face the following question:

Your overall confidence was 60 rather than 100. For each reason listed below, please indicate the degree to which it accounted for your less-than-complete confidence (in other words, the fact that you reported 60 rather than 100). In each case, select a number from **1 (Very Little)** to **7 (Very Much)**.

1. Not Sure What Each Option Means: You lack confidence in your decision because you're not sure you interpreted the available options correctly or understood their consequences.

Click **here** to reread a clarifying example.

Very Little   ○ ○ ○ ○ ○ ○ ○   Very Much
1 2 3 4 5 6 7

2. Multiple Views About What You Want: You lack confidence in your decision because, in your view, there's probably more than one right way to think about what you want to achieve.

Click **here** to reread a clarifying example.

Very Little   ○ ○ ○ ○ ○ ○ ○   Very Much
1 2 3 4 5 6 7

3. Don't Know How You'll Feel: You lack confidence in your decision because you aren't sure how you will feel about each of the possible outcomes.

Click **here** to reread a clarifying example.

Very Little   ○ ○ ○ ○ ○ ○ ○   Very Much
1 2 3 4 5 6 7

4. Didn't Apply Your Decision Criterion Correctly: You lack confidence in your decision because you think the criterion you decided to use and any simplifying assumptions you made may actually favor an alternative other than the alternative that you chose.

Click **here** to reread a clarifying example.

Very Little   ○ ○ ○ ○ ○ ○ ○   Very Much
1 2 3 4 5 6 7

5. Didn't Think Carefully About What You Want: You lack confidence in your decision because you haven't thought through what you want to achieve as carefully as you could.

Click **here** to reread a clarifying example.

Very Little   ○ ○ ○ ○ ○ ○ ○   Very Much
1 2 3 4 5 6 7

6. It's Nearly a Toss-Up: You lack confidence in your decision because you think the choice is close to a toss-up.

Click **here** to reread a clarifying example.

Very Little   ○ ○ ○ ○ ○ ○ ○   Very Much
           1   2   3   4   5   6   7

7. Worried About Mental Shortcuts: You lack confidence in your decision because you're worried about having used a mental shortcut that might have given the wrong answer.

Click **here** to reread a clarifying example.

Very Little   ○ ○ ○ ○ ○ ○ ○   Very Much
           1   2   3   4   5   6   7

8. Can't Know For Sure What You'll Get: You lack confidence in your decision because, based on the information you've received, you can't be sure what you will receive given your chosen option.

Click **here** to reread a clarifying example.

Very Little   ○ ○ ○ ○ ○ ○ ○   Very Much
           1   2   3   4   5   6   7

9. Other Reasons: You lack confidence in your decision for other reasons.

Click **here** to reread a clarifying example.

Very Little   ○ ○ ○ ○ ○ ○ ○   Very Much
           1   2   3   4   5   6   7

Next

Check your Understanding

We will now check your understanding of the different reasons for expressing less-than-complete confidence, and of how you should report the importance of different reasons. **If you answer at least 3 of 4 questions correctly, you will receive an extra $1 payment, so it's in your interests to take them seriously!**

Next

Test your Understanding

Question 1: Susan is deciding between two options. Option A is an apple. Option B is either a pear or an orange depending on whether Yellow comes before Blue on the color spectrum (i.e., the order of the rainbow). If Yellow comes before Blue then Option B is definitely a pear. If Yellow comes after Blue then Option B is definitely an orange. Susan doesn't recall the color spectrum precisely but remembers learning the acronym "ROY G BIV" for the colors of the rainbow. She thinks the "Y" stands for Yellow and the "B" stands for Blue, which would imply that Yellow comes before Blue, but she really isn't sure. She makes her decision as if Option B is a pear. She likes pears much more than apples so she chooses the pear. However, she reports overall confidence less than 100 because she is not that sure about the accuracy of the "ROY G BIV" acronym she used when making her choice.

Which of the following reasons should Susan mark as contributing "Very Much" to her less-than-complete confidence?

○ **Didn't Think Carefully About What You Want**: You lack confidence in your decision because you haven't thought through what you want to achieve as carefully as you could.

○ **It's Nearly a Toss-Up**: You lack confidence in your decision because you think the choice is close to a toss-up.

○ **Worried About Mental Shortcuts**: You lack confidence in your decision because you're worried about having used a mental shortcut that might have given the wrong answer.

<kbd>Next</kbd>

Test your Understanding

Question 2: Edward is deciding between two options. Option A is a box of oatmeal cookies, while Option B is a box of assorted chocolates. For Option B, he won't know which types of chocolates are in the box until he opens it. He knows he likes all types of chocolate more than oatmeal cookies, so he chooses Option B. He expresses overall confidence less than 100 because there's no way for him to know which chocolates are in the box.

Which of the following reasons should Edward mark as contributing "Very Much" to his less-than-complete confidence?

○ **Can't Know For Sure What You'll Get**: You lack confidence in your decision because, based on the information you've received, you can't be sure what you will receive given your chosen option.

○ **Didn't Apply Your Decision Criterion Correctly**: You lack confidence in your decision because you think the criterion you decided to use and any simplifying assumptions you made may actually favor an alternative other than the alternative that you chose.

○ **Worried About Mental Shortcuts**: You lack confidence in your decision because you're worried about having used a mental shortcut that might have given the wrong answer.

<kbd>Next</kbd>

Test your Understanding

Question 3: Samantha is deciding between two boxes. Box A contains either an apple or a pear, while Box B contains either one dollar or ten cents. The boxes are labeled with their exact contents, but the labels are written in a foreign language that Samantha has not studied. She does her best to guess what the words mean based on similarities to words in languages she knows. She concludes that Box A probably contains an apple while Box B probably contains one dollar. She chooses Box A because she prefers an apple to one dollar. She expresses overall confidence less than 100 because she's not sure she correctly translated the labels on the boxes.

Which of the following reasons should Samantha mark as contributing "Very Much" to her less-than-complete confidence?

○ **Didn't Think Carefully About What You Want**: You lack confidence in your decision because you haven't thought through what you want to achieve as carefully as you could.

○ **Not Sure What Each Option Means**: You lack confidence in your decision because you're not sure you interpreted the available options correctly or understood their consequences.

○ **Worried About Mental Shortcuts**: You lack confidence in your decision because you're worried about having used a mental shortcut that might have given the wrong answer.

Next

Test your Understanding

Question 4: Jude is deciding between the same two boxes as Samantha in the previous problem. Box A contains either an apple or a pear, while Box B contains either one dollar or ten cents. The boxes are labeled with their exact contents but the labels are written in a foreign language. Unlike Samantha, Jude can read the language and is fairly certain Box A contains an apple while Box B contains one dollar. An apple is worth more than a dollar to him, so he intends to select Box A. But after making his choice, he thinks he may have accidentally recorded his selection as Box B, containing the dollar. For that reason, he expresses overall confidence less than 100.

Which of the following reasons should Jude mark as contributing "Very Much" to his less-than-complete confidence?

○ **Not Sure What Each Option Means**: You lack confidence in your decision because you're not sure you interpreted the available options correctly or understood their consequences.

○ **Didn't Apply Your Decision Criterion Correctly**: You lack confidence in your decision because you think the criterion you decided to use and any simplifying assumptions you made may actually favor an alternative other than the alternative that you chose.

○ **It's Nearly a Toss-Up**: You lack confidence in your decision because you think the choice is close to a toss-up.

Next

Explaining the Reasons for your Overall Confidence

We will now ask you to provide the reasons for your overall confidence in 6 decisions that you made in this study.

Next

Explaining the Reasons for your Overall Confidence: Task 1

In a previous task you decided between the following two options:

Option A	Option B
Draw one ticket from this bag:	Draw one ticket from this bag:

Option A	Option B
20 blue tickets (❙) which pay **$25**	**16** purple tickets (❙) which pay **$30**
80 grey tickets (❙) which pay **$0**	**84** gold tickets (❙) which pay **$0**

> You selected **Option B** and reported a **51 out of 100** confidence in your choice.

For each reason listed below, please indicate the degree to which it accounted for your less-than-complete confidence (in other words, the fact that you reported 51 rather than 100). In each case, select a number from **1 (Very Little)** to **7 (Very Much)**.

Didn't Think Carefully About What You Want: You lack confidence in your decision because you haven't thought through what you want to achieve as carefully as you could.

Click **here** to reread a clarifying example.

Very Little ○ ○ ○ ○ ○ ○ ○ Very Much
          1 2 3 4 5 6 7

Don't Know How You'll Feel: You lack confidence in your decision because you aren't sure how you will feel about each of the possible outcomes.

Click **here** to reread a clarifying example.

Very Little ○ ○ ○ ○ ○ ○ ○ Very Much
          1 2 3 4 5 6 7

Worried About Mental Shortcuts: You lack confidence in your decision because you're worried about having used a mental shortcut that might have given the wrong answer.

Click **here** to reread a clarifying example.

Very Little ○ ○ ○ ○ ○ ○ ○ Very Much
          1 2 3 4 5 6 7

Didn't Apply Your Decision Criterion Correctly: You lack confidence in your decision because you think the criterion you decided to use and any simplifying assumptions you made may actually favor an alternative other than the alternative that you chose.

Click **here** to reread a clarifying example.

Very Little   ○ ○ ○ ○ ○ ○ ○   Very Much
         1   2   3   4   5   6   7

Multiple Views About What You Want: You lack confidence in your decision because, in your view, there's probably more than one right way to think about what you want to achieve.

Click **here** to reread a clarifying example.

Very Little   ○ ○ ○ ○ ○ ○ ○   Very Much
         1   2   3   4   5   6   7

Can't Know For Sure What You'll Get: You lack confidence in your decision because, based on the information you've received, you can't be sure what you will receive given your chosen option.

Click **here** to reread a clarifying example.

Very Little   ○ ○ ○ ○ ○ ○ ○   Very Much
         1   2   3   4   5   6   7

Not Sure What Each Option Means: You lack confidence in your decision because you're not sure you interpreted the available options correctly or understood their consequences.

Click **here** to reread a clarifying example.

Very Little   ○ ○ ○ ○ ○ ○ ○   Very Much
         1   2   3   4   5   6   7

It's Nearly a Toss-Up: You lack confidence in your decision because you think the choice is close to a toss-up.

Click **here** to reread a clarifying example.

Very Little   ○ ○ ○ ○ ○ ○ ○   Very Much
         1   2   3   4   5   6   7

Other Reasons: You lack confidence in your decision for other reasons.

Click **here** to reread a clarifying example.

Very Little   ○ ○ ○ ○ ○ ○ ○   Very Much
         1   2   3   4   5   6   7

Next

Hello and Welcome

Welcome, and thanks for your participation!

We are researchers from Caltech, Stanford, and The Ohio State University, inviting you to participate in a research study. The study should take roughly 30 to 35 minutes. Please click to review information about the study and to give your consent to participate.

Next

Prolific ID

You will be paid via Prolific for your participation in this study. In order to pay you, we need your Prolific ID.

Please enter your Prolific ID:

Next

Possible Rewards

You will receive $7.00 if you complete the entire study.

In addition to this payment, you have a 1 in 5 chance of being eligible to receive a bonus payment. The smallest possible bonus payment is $0 and the largest possible bonus payment is $30.

You will be informed of how your decisions will influence your bonus payment if you are eligible to receive one.

Next

This Study

In this study, you will complete 20 tasks, each of which involves your preferences over two options. The first option will always be called **Option A**. The second option will always be called **Option B**.

Option A and Option B refer to the possibility of winning monetary amounts ranging from $0 to $30 with some fixed chances. In each task you are asked to answer a few questions about Option A and Option B, and to choose between them.

Next

Example Options

We will now introduce you to the types of options that you will see in this study and the choices you will make.

Next

Example Options 1

Consider the following example:

Option A
Draw one ticket from this bag:

100 ❙

100 blue tickets (❙) which pay **$4**

Option B
Draw one ticket from this bag:

10 ❙
90 ❙

10 purple tickets (❙) which pay **$28**
90 gold tickets (❙) which pay **$2**

In this example:
- Option A allows you to draw one ticket from a bag containing 100 blue tickets (❙). Each blue ticket pays $4.
- Option B allows you to draw one ticket from a bag containing 10 purple tickets (❙) and 90 gold tickets (❙). Each purple ticket pays $28, and each gold ticket pays $2.

Therefore,
- Option A allows you to win $4 with a 100 in 100 chance.
- Option B allows you to win $28 with a 10 in 100 chance, or $2 with a 90 in 100 chance.

Next

Example Options 2

Consider the following example:

Option A
Draw one ticket from this bag:

100 blue tickets (▌) which pay **$13**

Option B
Draw one ticket from this bag:

50 ▌
50 ▌

50 purple tickets (▌) which pay **$y**

50 gold tickets (▌) which pay **$y**

$$y=(0.5\times 2)+(0.5\times 28)$$

In this example:
- Option A allows you to draw one ticket from a bag containing 100 blue tickets (▌). Each blue ticket pays $13.
- Option B allows you to draw one ticket from a bag containing 50 purple tickets (▌) and 50 gold tickets (▌). Each purple ticket pays $y and each gold ticket pays $y. $y is an amount equal to (0.5 × 2) + (0.5 × 28) that you can calculate. In this example, y is equal to 15.

Therefore,
- Option A allows you to win $13 with a 100 in 100 chance.
- Option B allows you to win $15 with a 100 in 100 chance.

Next

How will you make choices?

In each task, we will ask you two questions:

1. You will make a choice between Option A or Option B.
2. We will ask you **how certain** you are about your choice. Specifically, we are interested in how likely you think it is (in percentage terms) that the decision you made is actually your best decision, given your personal preferences and the available information.

Next

Comprehension check

Which one of the following statements is true?

○ When I am asked to indicate my certainty about my decision, the people running this study are interested in how certain I am that the decision I made is actually my best decision, given my personal preferences and the available information.

○ When I am asked to indicate my certainty about my decision, the people running this study are interested in how certain I am that I will actually receive the money from the options.

Next

Familiarize Yourself with the Study

- We now ask you to complete two training tasks to familiarize yourself with the questions and the choices you will make in the study.
- As you will realize from the training task, in this study after you confirm an answer *you will not be able to modify it*.
- For this reason, we ask you to think carefully before confirming your answers. After the training, we will explain how your answers in the study will affect your bonus.

Next

Training Task 1 of 2

Option A
Draw one ticket from this bag:
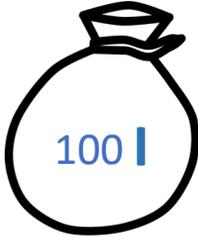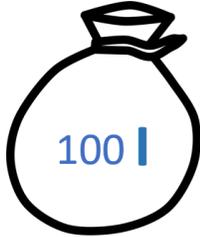
100 |

100 blue tickets (|) which pay **$4**

Option B
Draw one ticket from this bag:

10 |
90 |

10 purple tickets (|) which pay **$28**

90 gold tickets (|) which pay **$2**

Please make a choice.

Option A   Option B
○   ○

Next

Option A
Draw one ticket from this bag:

100 ▌

100 blue tickets (▌) which pay **$4**

Option B
Draw one ticket from this bag:

10 ▌
90 ▌

10 purple tickets (▌) which pay **$28**

90 gold tickets (▌) which pay **$2**

You chose **Option A**.
How certain are you that **choosing Option A** is actually your best decision, given your preferences and the available information?

| 50% | 60% | 70% | 80% | 90% | 100% |

Very
Uncertain

Completely
Certain

Next

Option A
Draw one ticket from this bag:

100 |

100 blue tickets (|) which pay **$13**

Option B
Draw one ticket from this bag:

50 |
50 |

50 purple tickets (|) which pay **$y**
50 gold tickets (|) which pay **$y**

y=(0.5×2)+(0.5×28)

Please make a choice.
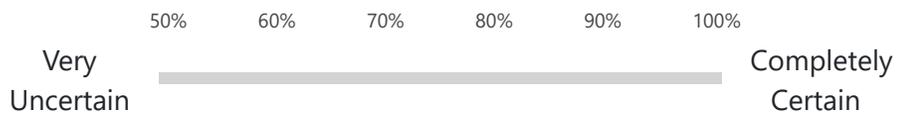
Option A   Option B
  ○   ○

Next

Training Task 2 of 2

Option A

Draw one ticket from this bag:
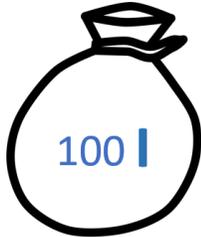
100 |

100 blue tickets (|) which pay **$13**

Option B

Draw one ticket from this bag:

50 |
50 |
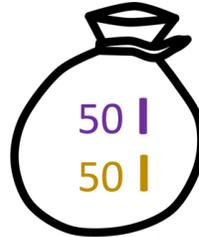
50 purple tickets (|) which pay **$y**

50 gold tickets (|) which pay **$y**

$$y=(0.5\times2)+(0.5\times28)$$

You chose **Option B**.
How certain are you that **choosing Option B** is actually your best decision, given your preferences and the available information?

| 50% | 60% | 70% | 80% | 90% | 100% |

Very
Uncertain

Completely
Certain
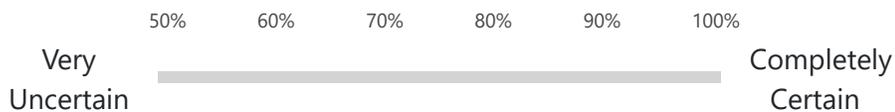
Next

Screenshots of the choice tasks from Experiment 3 are omitted, as they are identical to the training tasks.

Understanding your overall Certainty

We asked you to report your level of certainty in each of the twenty tasks you have completed so far in this study. In some tasks, you reported a certainty level of less than 100%, indicating less-than-complete certainty in your decisions. Next, we will ask you to explain some possible reasons for your less-than-complete certainty in 6 previous choices.

Next

Understanding your overall Certainty

In the next pages, we list nine possible reasons for having less-than-complete certainty in your decisions. For each reason, we also provide an illustration. It's important to keep in mind that these are just illustrations. In a few minutes we will ask you to classify similar examples, so please review these reasons carefully.

All these examples concern an individual named Chris who must choose between attending two social events with two different groups of friends.

Next

Understanding your Overall Certainty

1. Can't Know For Sure What You'll Get: You lack certainty in your decision because, based on the information you've received, you can't be sure what you will receive given your chosen option.

Here's an example of why Chris might report less-than-complete certainty for this reason:

Some of Chris's friends told him they might attend his chosen event or they might just stay home. He might express less-than-complete certainty in his decision because he can't be sure what his chosen option will yield since he doesn't know which friends will actually turn up to the event.

Next

47

Understanding your Overall Certainty

2. Didn't Think Carefully About What You Want: You lack certainty in your decision because you haven't thought through what you want to achieve as carefully as you could.

Here's an example of why Chris might report less-than-complete certainty for this reason:

Chris values experiences with different friends for different reasons but hasn't seriously considered whether one set of reasons is more important than another. He might express less-than-complete certainty in his decision because he hasn't carefully thought through what he wants from this social event.

Next

Understanding your Overall Certainty

3. Not Sure What Each Option Means: You lack certainty in your decision because you're not sure you interpreted the available options correctly or understood their consequences.

Here's an example of why Chris might report less-than-complete certainty for this reason:

Chris was told weeks ago which friends were attending which event, but he found some of their messages confusing and isn't sure he understood them properly. He might express less-than-complete certainty in his decision because he isn't sure he understands the consequences of choosing to attend one event vs. the other.

Next

Understanding your Overall Certainty

4. Worried About Mental Shortcuts: You lack certainty in your decision because you're worried about having used a mental shortcut that might have given the wrong answer.

Here's an example of why Chris might report less-than-complete certainty for this reason:

Chris' enjoyment of social events depends on who else is there, but it's hard for him to think through the pluses and minuses of being with one large group of friends rather than another. Instead, he makes his decision based entirely on which event his friend Parker plans to attend. He might express less-than-complete certainty in his decision because he's worried that the mental shortcut he used — to only think about Parker — isn't a good one.

Next

Understanding your Overall Certainty

5. Didn't Apply Your Decision Criterion Correctly: You lack certainty in your decision because you think the criterion you decided to use and any simplifying assumptions you made may actually favor an alternative other than the alternative that you chose.

Here's an example of why Chris might report less-than-complete certainty for this reason:

To simplify his choice, Chris assumed he would have the most fun at the event with the largest number of friends, so he listed the friends planning to attend each event and counted them. Right after committing to the event with the larger count, he worried that he might have miscounted. He might express less-than-complete certainty in his decision because he now thinks the criterion he decided to use and the simplifying assumption he made might actually favor an alternative other than the one he chose.

<div align="right">

Next

</div>

Understanding your Overall Certainty

6. Multiple Views About What You Want: You lack certainty in your decision because, in your view, there's probably more than one right way to think about what you want to achieve.

Here's an example of why Chris might report less-than-complete certainty for this reason:

Chris values experiences with different friends for different reasons and, after careful consideration, doesn't think one set of reasons is necessarily more or less important than another. He might express less-than-complete certainty in his decision because he doesn't think there's just one right way to think about what he wants from either social event.

<div align="right">

Next

</div>

Understanding your Overall Certainty

7. Don't Know How You'll Feel: You lack certainty in your decision because you aren't sure how you will feel about each of the possible outcomes.

Here's an example of why Chris might report less-than-complete certainty for this reason:

Chris enjoys different friends in different moods and is uncertain what his mood will be when the event arrives. He might express less-than-complete certainty in his decision because he isn't sure how he will feel about being with each group of friends.

<div align="right">

Next

</div>

Understanding your Overall Certainty

8. It's Nearly a Toss-Up: You lack certainty in your decision because you think the choice is close to a toss-up.

Here's an example of why Chris might report less-than-complete certainty for this reason:

After assessing the desirability of each event, Chris concludes that he would enjoy them about the same. However, because he is slightly unsure about each assessment, either one might be slightly better. He might express less-than-complete certainty in his decision because he thinks it is nearly a toss-up.

Next

Understanding your Overall Certainty

9. Other Reasons: You lack certainty in your decision for other reasons.

Here's an example of why Chris might report less-than-complete certainty for this reason:

Chris may have other reasons for expressing less-than-complete certainty in his decision that we haven't mentioned.

Next

Explaining the Reasons for your Overall Certainty

When you reported your overall certainty in each choice, any answer below 100% indicated that you had less-than-complete certainty. We want to understand your reasons for reporting less-than-complete certainty.

For each reason on the previous screens, you will indicate the degree to which it accounted for your less-than-complete certainty. For example, if you rated your certainty as 60%, we want to know the extent to which each reason was responsible for you answering 60% rather than 100%. In each case, you will answer on a scale of **1 (Very Little)** to **7 (Very Much)**. So if you rated your overall certainty as 60%, you would give a 1 for a particular reason if it had very little to do with saying your certainty was 60% rather than 100%, and you would give a 7 if it had a great deal to do with saying your certainty was 60% rather than 100%.

Please proceed to see an example of how the questions will look.

Next

Example

Imagine your overall certainty in a choice was 60% rather than 100%. You would then face the following question:

Your overall certainty was 60% rather than 100%. For each reason listed below, please indicate the degree to which it accounted for your less-than-complete certainty (in other words, the fact that you reported 60% rather than 100%). In each case, select a number from **1 (Very Little)** to **7 (Very Much)**.

1. Can't Know For Sure What You'll Get: You lack certainty in your decision because, based on the information you've received, you can't be sure what you will receive given your chosen option.

Click **here** to reread a clarifying example.

Very Little ○ ○ ○ ○ ○ ○ ○ Very Much
               1   2   3   4   5   6   7

2. Didn't Think Carefully About What You Want: You lack certainty in your decision because you haven't thought through what you want to achieve as carefully as you could.

Click **here** to reread a clarifying example.

Very Little ○ ○ ○ ○ ○ ○ ○ Very Much
               1   2   3   4   5   6   7

3. Not Sure What Each Option Means: You lack certainty in your decision because you're not sure you interpreted the available options correctly or understood their consequences.

Click **here** to reread a clarifying example.

Very Little ○ ○ ○ ○ ○ ○ ○ Very Much
               1   2   3   4   5   6   7

4. Worried About Mental Shortcuts: You lack certainty in your decision because you're worried about having used a mental shortcut that might have given the wrong answer.

Click **here** to reread a clarifying example.

Very Little ○ ○ ○ ○ ○ ○ ○ Very Much
               1   2   3   4   5   6   7

5. Didn't Apply Your Decision Criterion Correctly: You lack certainty in your decision because you think the criterion you decided to use and any simplifying assumptions you made may actually favor an alternative other than the alternative that you chose.

Click **here** to reread a clarifying example.

Very Little ○ ○ ○ ○ ○ ○ ○ Very Much
               1   2   3   4   5   6   7

6. Multiple Views About What You Want: You lack certainty in your decision because, in your view, there's probably more than one right way to think about what you want to achieve.

Click **here** to reread a clarifying example.

Very Little   ○ ○ ○ ○ ○ ○ ○   Very Much
              1 2 3 4 5 6 7

7. Don't Know How You'll Feel: <u>You lack certainty in your decision because you aren't sure how you will feel about each of the possible outcomes.</u>

Click **here** to reread a clarifying example.

Very Little   ○ ○ ○ ○ ○ ○ ○   Very Much
              1 2 3 4 5 6 7

8. It's Nearly a Toss-Up: <u>You lack certainty in your decision because you think the choice is close to a toss-up.</u>

Click **here** to reread a clarifying example.

Very Little   ○ ○ ○ ○ ○ ○ ○   Very Much
              1 2 3 4 5 6 7

9. Other Reasons: <u>You lack certainty in your decision for other reasons.</u>

Click **here** to reread a clarifying example.

Very Little   ○ ○ ○ ○ ○ ○ ○   Very Much
              1 2 3 4 5 6 7

Next

Check your Understanding

We will now check your understanding of the different reasons for expressing less-than-complete certainty, and of how you should report the importance of different reasons. **If you answer at least 3 of 4 questions correctly, you will receive an extra $1 payment, so it's in your interests to take them seriously!**

Next

Test your Understanding

Question 1: Susan is deciding between two options. Option A is an apple. Option B is either a pear or an orange depending on whether Yellow comes before Blue on the color spectrum (i.e., the order of the rainbow). If Yellow comes before Blue then Option B is definitely a pear. If Yellow comes after Blue then Option B is definitely an orange. Susan doesn't recall the color spectrum precisely but remembers learning the acronym "ROY G BIV" for the colors of the rainbow. She thinks the "Y" stands for Yellow and the "B" stands for Blue, which would imply that Yellow comes before Blue, but she really isn't sure. She makes her decision as if Option B is a pear. She likes pears much more than apples so she chooses the pear. However, she reports an overall certainty in her decision of less than 100% because she is not that sure about the accuracy of the "ROY G BIV" acronym she used when making her choice.

Which of the following reasons should Susan mark as contributing "Very Much" to her less-than-complete certainty?

○ **Didn't Think Carefully About What You Want**: You lack certainty in your decision because you haven't thought through what you want to achieve as carefully as you could.

○ **It's Nearly a Toss-Up**: You lack certainty in your decision because you think the choice is close to a toss-up.

○ **Worried About Mental Shortcuts**: You lack certainty in your decision because you're worried about having used a mental shortcut that might have given the wrong answer.

Next

Test your Understanding

Question 2: Edward is deciding between two options. Option A is a box of oatmeal cookies, while Option B is a box of assorted chocolates. For Option B, he won't know which types of chocolates are in the box until he opens it. He knows he likes all types of chocolate more than oatmeal cookies, so he chooses Option B. He expresses an overall certainty in his decision of less than 100% because there's no way for him to know which chocolates are in the box.

Which of the following reasons should Edward mark as contributing "Very Much" to his less-than-complete certainty?

○ **Can't Know For Sure What You'll Get**: You lack certainty in your decision because, based on the information you've received, you can't be sure what you will receive given your chosen option.

○ **Didn't Apply Your Decision Criterion Correctly**: You lack certainty in your decision because you think the criterion you decided to use and any simplifying assumptions you made may actually favor an alternative other than the alternative that you chose.

○ **Worried About Mental Shortcuts**: You lack certainty in your decision because you're worried about having used a mental shortcut that might have given the wrong answer.

Next

Test your Understanding

Question 3: Samantha is deciding between two boxes. Box A contains either an apple or a pear, while Box B contains either one dollar or ten cents. The boxes are labeled with their exact contents, but the labels are written in a foreign language that Samantha has not studied. She does her best to guess what the words mean based on similarities to words in languages she knows. She concludes that Box A probably contains an apple while Box B probably contains one dollar. She chooses Box A because she prefers an apple to one dollar. She expresses an overall certainty in her decision of less than 100% because she's not sure she correctly translated the labels on the boxes.

Which of the following reasons should Samantha mark as contributing "Very Much" to her less-than-complete certainty?

○ **Didn't Think Carefully About What You Want**: You lack certainty in your decision because you haven't thought through what you want to achieve as carefully as you could.

○ **Not Sure What Each Option Means**: You lack certainty in your decision because you're not sure you interpreted the available options correctly or understood their consequences.

○ **Worried About Mental Shortcuts**: You lack certainty in your decision because you're worried about having used a mental shortcut that might have given the wrong answer.

[Next]

Test your Understanding

Question 4: Jude is deciding between the same two boxes as Samantha in the previous problem. Box A contains either an apple or a pear, while Box B contains either one dollar or ten cents. The boxes are labeled with their exact contents but the labels are written in a foreign language. Unlike Samantha, Jude can read the language and is fairly certain Box A contains an apple while Box B contains one dollar. An apple is worth more than a dollar to him, so he intends to select Box A. But after making his choice, he thinks he may have accidentally recorded his selection as Box B, containing the dollar. For that reason, he expresses an overall certainty in his decision of less than 100%.

Which of the following reasons should Jude mark as contributing "Very Much" to his less-than-complete certainty?

○ **Not Sure What Each Option Means**: You lack certainty in your decision because you're not sure you interpreted the available options correctly or understood their consequences.

○ **Didn't Apply Your Decision Criterion Correctly**: You lack certainty in your decision because you think the criterion you decided to use and any simplifying assumptions you made may actually favor an alternative other than the alternative that you chose.

○ **It's Nearly a Toss-Up**: You lack certainty in your decision because you think the choice is close to a toss-up.

[Next]

Explaining the Reasons for your Overall Certainty

We will now ask you to provide the reasons for your overall certainty in 6 decisions that you made in this study.

[Next]

Explaining the Reasons for your Overall Certainty: Task 1

In a previous task you decided between the following two options:

Option A	**Option B**
Draw one ticket from this bag:	Draw one ticket from this bag:
20 |	18 |
80 |	82 |
20 blue tickets (|) which pay **$28**	**18** purple tickets (|) which pay **$30**
80 grey tickets (|) which pay **$0**	**82** gold tickets (|) which pay **$0**

You selected **Option B** and reported a **76%** certainty in your choice.

For each reason listed below, please indicate the degree to which it accounted for your less-than-complete certainty (in other words, the fact that you reported 76% rather than 100%). In each case, select a number from **1 (Very Little)** to **7 (Very Much)**.

Multiple Views About What You Want: You lack certainty in your decision because, in your view, there's probably more than one right way to think about what you want to achieve.

Click **here** to reread a clarifying example.

Very Little ○ ○ ○ ○ ○ ○ ○ Very Much
　　　　　　　1 2 3 4 5 6 7

Don't Know How You'll Feel: You lack certainty in your decision because you aren't sure how you will feel about each of the possible outcomes.

Click **here** to reread a clarifying example.

Very Little ○ ○ ○ ○ ○ ○ ○ Very Much
　　　　　　　1 2 3 4 5 6 7

Not Sure What Each Option Means: You lack certainty in your decision because you're not sure you interpreted the available options correctly or understood their consequences.

Click **here** to reread a clarifying example.

Very Little ○ ○ ○ ○ ○ ○ ○ Very Much
　　　　　　　1 2 3 4 5 6 7

Didn't Apply Your Decision Criterion Correctly: <u>You lack certainty in your decision because you think the criterion you decided to use and any simplifying assumptions you made may actually favor an alternative other than the alternative that you chose.</u>

Click **here** to reread a clarifying example.

Very Little ○ ○ ○ ○ ○ ○ ○ Very Much
　　　　　　1　2　3　4　5　6　7

It's Nearly a Toss-Up: <u>You lack certainty in your decision because you think the choice is close to a toss-up.</u>

Click **here** to reread a clarifying example.

Very Little ○ ○ ○ ○ ○ ○ ○ Very Much
　　　　　　1　2　3　4　5　6　7

Worried About Mental Shortcuts: <u>You lack certainty in your decision because you're worried about having used a mental shortcut that might have given the wrong answer.</u>

Click **here** to reread a clarifying example.

Very Little ○ ○ ○ ○ ○ ○ ○ Very Much
　　　　　　1　2　3　4　5　6　7

Didn't Think Carefully About What You Want: <u>You lack certainty in your decision because you haven't thought through what you want to achieve as carefully as you could.</u>

Click **here** to reread a clarifying example.

Very Little ○ ○ ○ ○ ○ ○ ○ Very Much
　　　　　　1　2　3　4　5　6　7

Can't Know For Sure What You'll Get: <u>You lack certainty in your decision because, based on the information you've received, you can't be sure what you will receive given your chosen option.</u>

Click **here** to reread a clarifying example.

Very Little ○ ○ ○ ○ ○ ○ ○ Very Much
　　　　　　1　2　3　4　5　6　7

Other Reasons: <u>You lack certainty in your decision for other reasons.</u>

Click **here** to reread a clarifying example.

Very Little ○ ○ ○ ○ ○ ○ ○ Very Much
　　　　　　1　2　3　4　5　6　7

Next

B.1.9 Instructions and Training: Risky Tasks in Experiment 4

Hello and Welcome

Welcome, and thanks for your participation!

We are researchers from Caltech, Stanford, and The Ohio State University, inviting you to participate in a research study. The study should take roughly 15 to 20 minutes. Please click to review information about the study and to give your consent to participate.

Next

Prolific ID

You will be paid via Prolific for your participation in this study. In order to pay you, we need your Prolific ID.

Please enter your Prolific ID:

Next

Possible Rewards

You will receive $4.00 if you complete the entire study.

In addition to this payment, you have a 1 in 5 chance of being eligible to receive a bonus payment. The smallest possible bonus payment is $0 and the largest possible bonus payment is $30.

You will be informed of how your decisions will influence your bonus payment if you are eligible to receive one.

Next

This Study

In this study, you will complete tasks in two separate blocks. If you are selected for a bonus payment, one of these blocks will be chosen at random. Within each block, we will explain how your bonus would be determined if that block is selected. Please click to learn about Block 1.

Next

Description of Block 1

In Block 1, you will complete 10 tasks, each of which involves your preferences over two options. The first option will always be called **Option A**. The second option will always be called **Option B**.

Option A and Option B refer to the possibility of winning monetary amounts ranging from $0 to $30 with some fixed chances. In each task, you will decide whether you prefer Option A or Option B.

Next

Example Options in Block 1

We will now introduce you to the types of options that you will encounter in Block 1 and the choices you will make.

Next

Example Options in Block 1

Consider the following example:

Option A	Option B
Draw one ticket from this bag:	Draw one ticket from this bag:

100 blue tickets (❘) which pay **$4**

10 purple tickets (❘) which pay **$28**
90 gold tickets (❘) which pay **$2**

In this example:
- Option A allows you to draw one ticket from a bag containing 100 blue tickets (❘). Each blue ticket pays $4.
- Option B allows you to draw one ticket from a bag containing 10 purple tickets (❘) and 90 gold tickets (❘). Each purple ticket pays $28, and each gold ticket pays $2.

Therefore:
- Option A allows you to win $4 with a 100 in 100 chance.
- Option B allows you to win $28 with a 10 in 100 chance, or $2 with a 90 in 100 chance.

Next

How will you make choices in Block 1?

In each task, we will show you two options and will ask you to choose between the following two answers:

1. I prefer Option A
2. I prefer Option B

Next

Familiarize Yourself with Block 1

- We now ask you to complete one training task to familiarize yourself with the choices you will make in Block 1.
- As you will notice from the training task, in this study after you confirm an answer *you will not be able to modify it*.
- For this reason, we ask you to think carefully before confirming your answers. After the training, we will explain how your answers in Block 1 may affect your bonus.

Next

Training Task - Block 1

Option A
Draw one ticket from this bag:
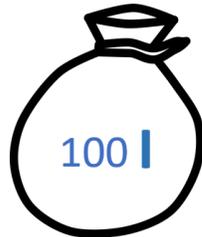
100 ▌

100 blue tickets (▌) which pay **$4**

Option B
Draw one ticket from this bag:

10 ▌
90 ▌

10 purple tickets (▌) which pay **$28**

90 gold tickets (▌) which pay **$2**

Which option do you **prefer**?

I prefer Option A   I prefer Option B
○   ○

Next

Bonus Payment in Block 1

Congratulations! You have completed your training task! You are now ready to learn more about how your potential bonus payment will be determined if Block 1 is selected.

We call **potential bonus payment in Block 1** an additional amount of money that you may receive at the end of the study, in addition to your participation fee. The potential bonus payment in Block 1 will **only** be paid if certain conditions are met.

First, we will explain how the potential bonus payment in Block 1 is determined. Then, we will describe the conditions under which you may receive it.

Next

Determining Your Potential Bonus Payment in Block 1

To determine your potential bonus payment in Block 1, we will randomly select one task from Block 1 as the **task-that-counts**. For example, the task-that-counts in Block 1 could be the one illustrated below.

Option A	**Option B**
Draw one ticket from this bag:	Draw one ticket from this bag:

100 blue tickets (▎) which pay **$4**

10 purple tickets (▎) which pay **$28**

90 gold tickets (▎) which pay **$2**

If you preferred **Option A** in the task-that-counts, then Option A will be implemented, and we will draw a single ticket from that bag.

- A blue ticket will definitely be chosen, so your potential bonus payment will be $4.

If you preferred **Option B** in the task-that-counts, then Option B will be implemented, and we will draw a single ticket from that bag.

- If a purple ticket is drawn, then your potential bonus payment will be $28.
- If a gold ticket is drawn, then your potential bonus payment will be $2.

Because any task in Block 1 could be selected as the task-that-counts, you should make each choice as if it could determine your potential bonus payment.

Next

Conditions for Your Bonus Payment in Block 1

You will receive your potential bonus payment in Block 1 at the end of the study **only if** two conditions are met:

1. Block 1 is randomly selected to determine your potential bonus payment.
2. After we compute your potential bonus payment from Block 1, we will draw a number between 1 and 5. If the drawn number is 1, you will be eligible to receive the potential bonus payment. Otherwise, you will not receive a bonus payment.

Next

Conditions for Your Bonus Payment in Block 1 - Example

Suppose again that the task-that-counts in Block 1 is the one illustrated below.

<div style="display:flex">

Option A
Draw one ticket from this bag:
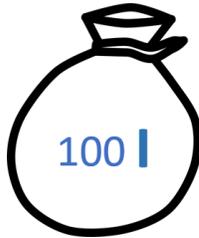
100 |

100 blue tickets (|) which pay **$4**

Option B
Draw one ticket from this bag:

10 |
90 |
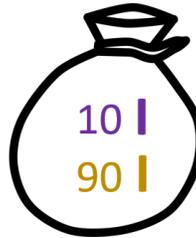
10 purple tickets (|) which pay **$28**

90 gold tickets (|) which pay **$2**

</div>

Imagine that you chose Option B, and a purple ticket was drawn, resulting in a potential bonus payment of $28.

You will receive $28 as your bonus payment at the end of the experiment if both of the following conditions are met:

1. Block 1 is selected to determine your potential bonus payment.
2. You are eligible to receive this potential bonus payment.

[Next]

Test your Understanding

Suppose that this is the task-that-counts in Block 1:

Option A
Draw one ticket from this bag:
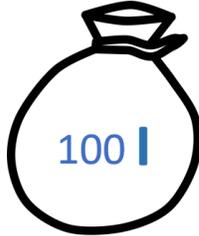
100 |

100 blue tickets (|) which pay **$4**

Option B
Draw one ticket from this bag:

10 |
90 |

10 purple tickets (|) which pay **$28**
90 gold tickets (|) which pay **$2**

Here are a few questions to test your understanding:

How is your potential bonus payment in Block 1 determined?
○ Draw a ticket from the bag corresponding to the option that you preferred.
○ Always draw a ticket from the bag corresponding to Option A.
○ Always draw a ticket from the bag corresponding to Option B.
○ Draw a ticket from the bag corresponding to Option A with a 50 in 100 chance or from the bag corresponding to Option B with a 50 in 100 chance.

What is your potential bonus payment in Block 1 if you preferred Option B and a purple ticket is drawn?
○ $2
○ $28
○ $15
○ $4

Under what condition will you receive the potential bonus payment in Block 1 at the end of the experiment?
○ I will receive the potential bonus payment in Block 1 only if I am eligible and Block 1 is selected to determine my potential bonus payment. I have a 1 in 4 chance of being eligible.
○ I will receive the potential bonus payment in Block 1 only if I am eligible and Block 1 is selected to determine my potential bonus payment. I have a 1 in 2 chance of being eligible.
○ I will receive the potential bonus payment in Block 1 only if I am eligible and Block 1 is selected to determine my potential bonus payment. I have a 1 in 5 chance of being eligible.
○ I will receive the potential bonus payment in Block 1 only if I am eligible and Block 1 is selected to determine my potential bonus payment. I have a 1 in 3 chance of being eligible.

Next

Screenshots of the choice tasks from Experiment 4 are omitted, as they are identical to the training tasks.

B.1.10 Final Questions: Risky Tasks in Experiment 4

Congratulations!

You've completed the 10 tasks of Block 1! Before proceeding to Block 2, we ask you to answer a few additional questions about how you made your choices in Block 1.

Your answers to these questions will not affect your potential bonus payment or your chances of receiving it, but they are important for our research. Therefore, we ask you to respond thoughtfully.

Next

The first question presented to all subjects is:

Additional Questions - Block 1

When you were making your decisions in Block 1, did you make any mathematical calculations to arrive at your choices? By "mathematical calculation" we mean any use of mathematical operations such as addition, subtraction, multiplication, or division.

○ Yes
○ No

Next

If subjects answer "Yes" to the first question, the second question is:

Additional Questions - Block 1

You indicated that you performed a mathematical calculation to arrive at your choices. Did you calculate a mean, an average, or an expected value, or did you calculate something else?

○ Mean/average/expected value
○ Something else

Next

If subjects answer "Yes" to the first question, they are asked to complete the following task regardless of their response to the second question:

Explain your Choices

One of the questions you encountered in Block 1 is shown below. Please describe in detail the mathematical calculation(s) you performed in Block 1 using this question as an example, and please explain how these calculations influenced your choice.

Option A
Draw one ticket from this bag:

100
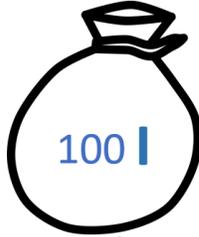
100 blue tickets (|) which pay **$26**

Option B
Draw one ticket from this bag:

90
10
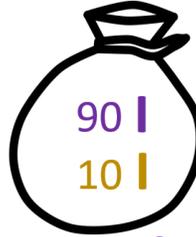
90 purple tickets (|) which pay **$30**
10 gold tickets (|) which pay **$0**

> Type your response here...

Next

If subjects answer "No" to the first question, the second question is:

Additional Questions - Block 1

You indicated that you did not perform mathematical calculations to arrive at your choices. Is this because mathematical calculations are not necessary for you to figure out which option you prefer? Or, would mathematical calculations have helped you figure out which option you prefer, but these calculations require too much time/effort/ energy etc.? Or, was it for some other reason?

○ Mathematical calculations are not necessary for me to figure out which option I prefer
○ Mathematical calculations would have helped me figure out which option I prefer, but these calculations require too much time/effort/ energy etc.
○ Other reason

Next

If subjects answer "No" to the first question and select "Mathematical calculations would have helped me figure out which option I prefer, but these calculations require too much time/effort/ energy etc." in the second question, they are asked to complete the following task:

Describe your Mathematical Calculation(s)

One of the questions you encountered in Block 1 is shown below. Please describe the mathematical calculation(s) you would need to perform using this question as an example, and please explain how these calculations would influence your choice. You do not need to actually perform the calculation, but please just describe what it would be.

	Option A Draw one ticket from this bag:		**Option B** Draw one ticket from this bag:

100 **|**

20 **|**
80 **|**

100 blue tickets (**|**) which pay **$7**

20 purple tickets (**|**) which pay **$30**
80 gold tickets (**|**) which pay **$0**

Type your response here...

Next

If subjects answer "No" to the first question and select "Other reason" in the second question, they are asked to complete the following task:

Explain your Other Reason(s)

Please describe in detail any other reasons why you did not perform mathematical calculations to make your choices in Block 1.

Type your response here...

Next

Description of Block 2

In Block 2, you will complete 10 tasks involving your preferences between two options, each of which **guarantees** a fixed monetary prize **with certainty**. The first option will always be called **Option A**. The second option will always be called **Option B**.

In each task, you will decide whether you prefer Option A or Option B.

Next

Example Options in Block 2

We will now introduce you to the types of options that you will encounter in Block 2 and the choices you will make.

Next

Example Options in Block 2

Consider the following example:
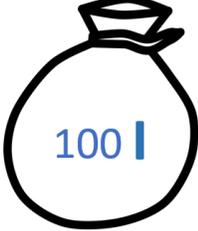
Option A

Draw one ticket from this bag:

100 blue tickets (|) which pay **$13**

Option B

Draw one ticket from this bag:

50 |
50 |

50 purple tickets (|) which pay **$y**

50 gold tickets (|) which pay **$y**

$$y=(0.5 \times 2)+(0.5 \times 28)$$

In this example:

- Option A allows you to draw one ticket from a bag containing 100 blue tickets (|). Each blue ticket pays $13.
- Option B allows you to draw one ticket from a bag containing 50 purple tickets (|) and 50 gold tickets (|). Each purple ticket pays $y and each gold ticket pays $y. $y is an amount equal to (0.5 × 2) + (0.5 × 28) that you can calculate. In this example, y is equal to 15.

Therefore, both Option A and Option B **do not involve any risk**:

- Option A allows you to win $13 with a 100 in 100 chance.
- Option B allows you to win $15 with a 100 in 100 chance.

Next

How will you make choices in Block 2?

In each task, we will show you two options and will ask you to choose between the following two answers:

1. I prefer Option A
2. I prefer Option B

Next

Familiarize Yourself with Block 2

- We now ask you to complete one training task to familiarize yourself with the choices you will make in Block 2.
- As you will notice from the training task, in this study after you confirm an answer *you will not be able to modify it*.
- For this reason, we ask you to think carefully before confirming your answers. After the training, we will explain how your answers in Block 2 may affect your bonus.

Next

Training Task - Block 2

Option A
Draw one ticket from this bag:
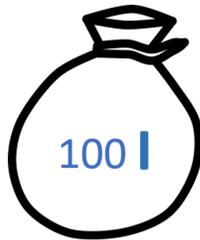
100 ▎

100 blue tickets (▎) which pay **$13**

Option B
Draw one ticket from this bag:

50 ▎
50 ▎

50 purple tickets (▎) which pay **$y**

50 gold tickets (▎) which pay **$y**

y=(0.5×2)+(0.5×28)

Remember, this task **DOES NOT INVOLVE ANY RISK.** When a bag contains tickets of more than one color, all of them give **the same payment**.

Which option do you **prefer**?

I prefer Option A
○

I prefer Option B
○

Next

70

Bonus Payment in Block 2

Congratulations! You have completed your training task! You are now ready to learn more about your potential bonus payment in Block 2.

We call **potential bonus payment in Block 2** an additional amount of money that you may receive at the end of the study, in addition to your participation fee. The potential bonus payment in Block 2 will **only** be paid if certain conditions are met.

First, we will explain how the potential bonus payment in Block 2 is determined. Then, we will describe the conditions under which you may receive it.

Next

Determining Your Potential Bonus Payment in Block 2

To determine your potential bonus payment in Block 2, we will randomly select one task from Block 2 as the **task-that-counts**. For example, the task-that-counts in Block 2 could be the one illustrated below.

Option A	**Option B**
Draw one ticket from this bag:	Draw one ticket from this bag:
100 ∣	50 ∣
	50 ∣
100 blue tickets (∣) which pay **$13**	50 purple tickets (∣) which pay **$y**
	50 gold tickets (∣) which pay **$y**
	y=(0.5×2)+(0.5×28)

If you preferred **Option A** in the task-that-counts, then Option A will be implemented, and we will draw a single ticket from that bag.

- A blue ticket will definitely be chosen, so your potential bonus payment will be $13.

If you preferred **Option B** in the task-that-counts, then Option B will be implemented, and we will draw a single ticket from that bag.

- A purple or gold ticket will be drawn. Regardless of the color of the drawn ticket, your potential bonus payment will be $y. In this example, y is equal to 15.

Because any task in Block 2 could be selected as the task-that-counts, you should make each choice as if it could determine your potential bonus payment.

Next

Conditions for Your Bonus Payment in Block 2

You will receive your potential bonus payment in Block 2 at the end of the study **only if** two conditions are met:

1. Block 2 is randomly selected to determine your potential bonus payment.
2. After we compute your potential bonus payment from Block 2, we will draw a number between 1 and 5. <u>If the drawn number is 1</u>, you will be eligible to receive the potential bonus payment. Otherwise, you will not receive a bonus payment.

Next

Conditions for Your Bonus Payment in Block 2 - Example

Suppose again that the task-that-counts in Block 2 is the one illustrated below.

Option A	Option B
Draw one ticket from this bag:	Draw one ticket from this bag:

100 blue tickets (❘) which pay **$13**

50 purple tickets (❘) which pay **$y**

50 gold tickets (❘) which pay **$y**

$$y=(0.5 \times 2)+(0.5 \times 28)$$

Imagine that you chose Option B. Regardless of which ticket will be drawn, your potential bonus payment will be $15.

You will receive $15 as your bonus payment at the end of the experiment if both of the following conditions are met:

1. Block 2 is selected to determine your potential bonus payment.
2. You are eligible to receive this potential bonus payment.

Next

Test your Understanding

Suppose that this is the task-that-counts in Block 2:

Option A
Draw one ticket from this bag:
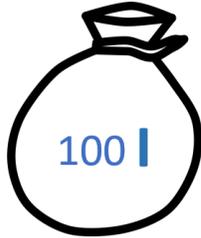
100 |

100 blue tickets (|) which pay **$13**

Option B
Draw one ticket from this bag:

50 |
50 |
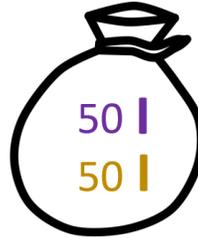
50 purple tickets (|) which pay **$y**

50 gold tickets (|) which pay **$y**

$$y=(0.5\times2)+(0.5\times28)$$

Here are a few questions to test your understanding:

How is your potential bonus payment in Block 2 determined?

○ Draw a ticket from the bag corresponding to the option that you preferred.

○ Always draw a ticket from the bag corresponding to Option A.

○ Always draw a ticket from the bag corresponding to Option B.

○ Draw a ticket from the bag corresponding to Option A with a 50 in 100 chance or from the bag corresponding to Option B with a 50 in 100 chance.

What is your potential bonus payment in Block 2 if you preferred Option B and a purple ticket is drawn?

○ $2 with a 50 in 100 chance, or $28 with a 50 in 100 chance

○ $28

○ $2

○ $15

Under what condition will you receive the potential bonus payment in Block 2 at the end of the experiment?

○ I will receive the potential bonus payment in Block 2 only if I am eligible and Block 2 is selected to determine my potential bonus payment. I have a 1 in 5 chance of being eligible.

○ I will receive the potential bonus payment in Block 2 only if I am eligible and Block 2 is selected to determine my potential bonus payment. I have a 1 in 2 chance of being eligible.

○ I will receive the potential bonus payment in Block 2 only if I am eligible and Block 2 is selected to determine my potential bonus payment. I have a 1 in 4 chance of being eligible.

○ I will receive the potential bonus payment in Block 2 only if I am eligible and Block 2 is selected to determine my potential bonus payment. I have a 1 in 3 chance of being eligible.

Next

Screenshots of the choice tasks from Experiment 4 are omitted, as they are identical to the training tasks.

Congratulations!

You've completed the 10 tasks of Block 2! Before learning about your potential bonus payment and if you are eligible to receive it, we ask you to answer a few additional questions about how you made your choices in Block 2.

Your answers to these questions will not affect your potential bonus payment or your chances of receiving it, but they are important for our research. Therefore, we ask you to respond thoughtfully.

Next

The first question presented to all subjects is:

Additional Questions - Block 2

When you were making your decisions in Block 2, did you make any mathematical calculations to arrive at your choices? By "mathematical calculation" we mean any use of mathematical operations such as addition, subtraction, multiplication, or division.

○ Yes
○ No

Next

If subjects answer "Yes" to the first question, they are asked to complete the following task:

Explain your Choices

One of the questions you encountered in Block 2 is shown below. Please describe in detail the mathematical calculation(s) you performed in Block 2 using this question as an example, and please explain how these calculations influenced your choice.

Option A	**Option B**
Draw one ticket from this bag:	Draw one ticket from this bag:

20 blue tickets (❘) which pay **$z**
80 grey tickets (❘) which pay **$z**

$$z=(0.8\times0)+(0.2\times23)$$

16 purple tickets (❘) which pay **$y**
84 gold tickets (❘) which pay **$y**

$$y=(0.84\times0)+(0.16\times30)$$

Type your response here...

Next

If subjects answer "No" to the first question, they are asked to complete the following task:

Additional Questions - Block 2

You indicated that you did not perform mathematical calculations to arrive at your choices. Is this because mathematical calculations are not necessary for you to figure out which option you prefer? Or, would mathematical calculations have helped you figure out which option you prefer, but these calculations require too much time/effort/ energy etc.? Or, was it for some other reason?

○ Mathematical calculations are not necessary for me to figure out which option I prefer
○ Mathematical calculations would have helped me figure out which option I prefer, but these calculations require too much time/effort/ energy etc.
○ Other reason

Next

If subjects answer "No" to the first question and select "Mathematical calculations would have helped me figure out which option I prefer, but these calculations require too much time/effort/energy etc." in the second question, they are asked to complete the following task:

Describe your Mathematical Calculation(s)

One of the questions you encountered in Block 2 is shown below. Please describe the mathematical calculation(s) you would need to perform using this question as an example, and please explain how these calculations would influence your choice. You do not need to actually perform the calculation, but please just describe what it would be.

<div align="center">

Option A
Draw one ticket from this bag:

Option B
Draw one ticket from this bag:

</div>

20 |
80 |

18 |
82 |

20 blue tickets (|) which pay **$z**
80 grey tickets (|) which pay **$z**

$z = (0.8 \times 0) + (0.2 \times 28)$

18 purple tickets (|) which pay **$y**
82 gold tickets (|) which pay **$y**

$y = (0.82 \times 0) + (0.18 \times 30)$

Type your response here...

Next

If subjects answer "No" to the first question and select "Other reason" in the second question, they are asked to complete the following task:

Explain your Other Reason(s)

Please describe in detail any other reasons why you did not perform mathematical calculations to make your choices in Block 2.

Type your response here...

Next

Hello and Welcome

Welcome, and thanks for your participation!

We are researchers from Caltech, Stanford, and The Ohio State University inviting you to participate in a research study. The study should take roughly 20 to 25 minutes. Please click to review information about the study and to give your consent to participate.

Next

Prolific ID

You will be paid via Prolific for your participation in this study. In order to pay you, we need your Prolific ID.

Please enter your Prolific ID:

Next

Possible Rewards

You will receive $5.00 if you complete the entire study.

In addition to this payment, you have a 1 in 5 chance of being eligible to receive a bonus payment. The smallest possible bonus payment is $0 and the largest possible bonus payment is $30.

You will be informed of how your decisions will influence your bonus payment if you are eligible to receive one.

Next

This Study

In this study, you will choose between options consisting of sets of TICKETS like the ones below:

Option A
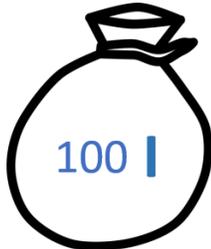
100 blue tickets (❙) which are labeled **$7**

Option B

25 purple tickets (❙) which are labeled **$30**

75 gold tickets (❙) which are labeled **$0**

Each ticket is labeled with an amount of money ranging from $0 to $30. In some cases the computer will select one ticket from your chosen option, and in other cases it will select more than one ticket. The label on the ticket or tickets it selects will be used to determine the payment you will receive according to rules we will explain.

In the example above:

- **Option A** consists of 100 tickets labeled with **$7** each.
- **Option B** consists of 25 tickets labeled with **$30** each, and 75 tickets labeled with **$0** each.

Your job will be to decide which set of tickets (Option A or Option B) you would like the computer to pay you based on.

Next

This Study

In this study, you will complete tasks in two separate blocks. If you are selected for a bonus payment, one of these blocks will be chosen at random. Within each block, we will explain how your bonus would be determined if that block is selected. Please click to learn about Block 1.

Next

The order of Risky and Risk-Free tasks is randomized in Experiment 5. In the screenshots below, Block 1 contains Risk-Free tasks. The instructions and comprehension questions remain the same when the Risk-Free tasks are presented in Block 2.

Block 1: Select-All-Tickets

For each task in Block 1, the computer will pay you by calculating the **AVERAGE** amount of money across all 100 ticket labels for **whichever option you've chosen**. That is, it will add up the monetary value of each of the 100 ticket labels and divide that sum by 100.

Note that, because the computer will pay you the **AVERAGE** of all the 100 ticket labels, this means that there is **no risk involved** in any task. There is only one possible payment amount you can receive from a given option, and this amount is determined by the average amount of money across all 100 ticket labels.

Next

Bonus Payment in Block 1

We will refer to the additional amount of money that you may receive at the end of the study, in addition to your participation fee, as the **potential bonus payment in Block 1**. The potential bonus payment in Block 1 will **only** be paid if certain conditions are met.

First, we will explain how the potential bonus payment in Block 1 is determined. Then, we will describe the conditions under which you may receive it.

Next

Potential Bonus Payment in Block 1

To determine your potential bonus payment in Block 1, we will randomly select one task from Block 1 as the **task-that-counts**. For example, the task-that-counts in Block 1 could be the one illustrated below.

Option A

Option B

100

60
40

100 blue tickets (|) which are labeled **$17**

60 purple tickets (|) which are labeled **$30**

40 gold tickets (|) which are labeled **$0**

If you preferred **Option A** in the task-that-counts, your potential bonus payment will be **(100 × $17) / 100 = $17**.

If you preferred **Option B** in the task-that-counts, your potential bonus payment will be **(60 × $30 + 40 × $0) / 100 = $18**.

Because any task in Block 1 could be selected as the task-that-counts, you should make each choice as if it will determine your potential bonus payment.

Again, note that **there is no risk** in this example task. There is only one amount that you could receive from Option A (you receive $17 with 100% chance), and one amount that you could receive from Option B (you receive $18 with 100% chance).

Next

Conditions for Your Bonus Payment in Block 1

You will receive your potential bonus payment in Block 1 at the end of the study **only if** two conditions are met:

1. Block 1 is randomly selected to determine your potential bonus payment.

2. After we compute your potential bonus payment from Block 1, we will draw a number between 1 and 5. If the drawn number is 1, you will be eligible to receive the potential bonus payment. Otherwise, you will not receive a bonus payment.

Next

Comprehension Questions

Option A

100 blue tickets (❘) which are labeled **$11**

Option B

40 purple tickets (❘) which are labeled **$30**

60 gold tickets (❘) which are labeled **$0**

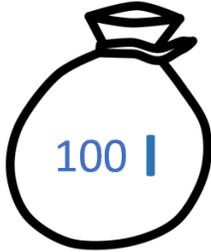**Suppose that the task in the example above is the task-that-counts in Block 1, and you chose <u>Option B</u>.**

What is the chance that $30 is your potential bonus payment?

○ 0 in 100 (0%)

○ 40 in 100 (40%)

○ 60 in 100 (60%)

○ 100 in 100 (100%)

Submit Answer

Comprehension Questions

Option A

100 blue tickets (|) which are labeled **$11**

Option B

40 purple tickets (|) which are labeled **$30**

60 gold tickets (|) which are labeled **$0**

Suppose that the task in the example above is the task-that-counts in Block 1, and you chose <u>Option B</u>.

What is the chance that $12 is your potential bonus payment?

○ 0 in 100 (0%)

○ 40 in 100 (40%)

○ 60 in 100 (60%)

○ 100 in 100 (100%)

Submit Answer

Comprehension Questions

Option A

100 blue tickets (❘) which are labeled **$11**

Option B

40 purple tickets (❘) which are labeled **$30**

60 gold tickets (❘) which are labeled **$0**

Suppose that the task in the example above is the task-that-counts in Block 1, and you chose <u>Option B</u>.

What is the chance that $11 is your potential bonus payment?

○ 0 in 100 (0%)

○ 40 in 100 (40%)

○ 60 in 100 (60%)

○ 100 in 100 (100%)

[Submit Answer]

Comprehension Questions

Option A

100 blue tickets (|) which are labeled **$11**

Option B

40 purple tickets (|) which are labeled **$30**

60 gold tickets (|) which are labeled **$0**

Suppose that the task in the example above is the task-that-counts in Block 1, and you chose <u>Option B</u>.

What is the chance that $0 is your potential bonus payment?

○ 0 in 100 (0%)

○ 40 in 100 (40%)

○ 60 in 100 (60%)

○ 100 in 100 (100%)

Submit Answer

Begin Block 1

- Congratulations! You answered all the comprehension questions correctly! You are now ready to begin Block 1.

- Remember that your choices in Block 1 may affect your bonus payment if you are selected to receive one.

- Therefore, **it is in your best interest to answer the next questions thoughtfully.**

- If anything is unclear, please let us know through the Prolific anonymized internal messaging service. Otherwise, proceed to begin Block 1.

Next

(Reminder of payment procedures)

Option A

Option B

100 |

50 |
50 |

100 blue tickets (|) which are labeled **$14**

50 purple tickets (|) which are labeled **$30**

50 gold tickets (|) which are labeled **$0**

Remember, this task **DOES NOT INVOLVE ANY RISK**. When you choose an option, there is only one possible amount you can receive.

Which option do you **prefer**?

I prefer **Option A**   I prefer **Option B**

○   ○

Next

The screenshot below shows the pop-up that appears when subjects click the "Reminder of payment procedures" button in the choice task.

The order of Risky and Risk-Free tasks is randomized in Experiment 5. In the screenshots below, Block 1 contains Risky tasks. The instructions and comprehension questions remain the same when the Risky tasks are presented in Block 2.

Block 1: Select-One-Ticket

For each task in Block 1, the computer will **RANDOMLY** select one of the 100 tickets from whichever option you've chosen (each ticket in the option you chose is **EQUALLY** likely to be selected by the computer).

Note that, because the computer will **RANDOMLY** select one of the 100 tickets and pay you the value of that ticket label, there **is risk involved** in each task. There could be multiple possible amounts you can receive, and the amount that you receive would be determined by the ticket label on one randomly-selected ticket.

Next

Bonus Payment in Block 1

We will refer to the additional amount of money that you may receive at the end of the study, in addition to your participation fee, as the **potential bonus payment in Block 1**. The potential bonus payment in Block 1 will **only** be paid if certain conditions are met.

First, we will explain how the potential bonus payment in Block 1 is determined. Then, we will describe the conditions under which you may receive it.

Next

Potential Bonus Payment in Block 1

To determine your potential bonus payment in Block 1, we will randomly select one task from Block 1 as the **task-that-counts**. For example, the task-that-counts in Block 1 could be the one illustrated below.

Option A

Option B

100 |

60 |
40 |

100 blue tickets (|) which are labeled **$17**

60 purple tickets (|) which are labeled **$30**

40 gold tickets (|) which are labeled **$0**

If you preferred **Option A** in the task-that-counts, your potential bonus payment will be **$17**.

If you preferred **Option B** in the task-that-counts, your potential bonus payment will be **$30 with a 60 in 100 chance**, or **$0 with a 40 in 100 chance**.

Because any task in Block 1 could be selected as the task-that-counts, you should make each choice as if it will determine your potential bonus payment.

Again, note that **there is risk** involved in this example task. There is only one amount that you could receive from Option A ($17) but there are two possible amounts that you could receive from Option B ($30 with 60% chance or $0 with 40% chance).

Next

Conditions for Your Bonus Payment in Block 1

You will receive your potential bonus payment in Block 1 at the end of the study **only if** two conditions are met:

1. Block 1 is randomly selected to determine your potential bonus payment.
2. After we compute your potential bonus payment from Block 1, we will draw a number between 1 and 5. If the drawn number is 1, you will be eligible to receive the potential bonus payment. Otherwise, you will not receive a bonus payment.

Next

Comprehension Questions

Option A

100 blue tickets (❘) which are labeled **$11**

Option B

40 purple tickets (❘) which are labeled **$30**

60 gold tickets (❘) which are labeled **$0**

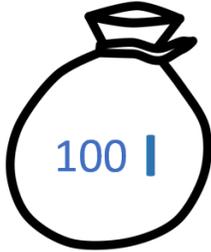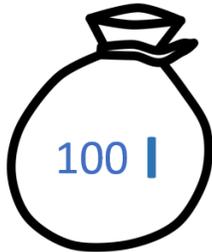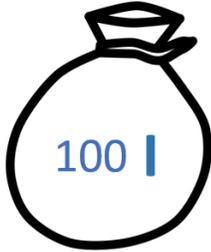**Suppose that the task in the example above is the task-that-counts in Block 1, and you chose <u>Option B</u>.**

What is the chance that $30 is your potential bonus payment?

○ 0 in 100 (0%)

○ 40 in 100 (40%)

○ 60 in 100 (60%)

○ 100 in 100 (100%)

[Submit Answer]

90

Comprehension Questions

Option A

100 blue tickets (**|**) which are labeled **$11**

Option B

40 purple tickets (**|**) which are labeled **$30**

60 gold tickets (**|**) which are labeled **$0**

Suppose that the task in the example above is the task-that-counts in Block 1, and you chose <u>Option B</u>.

What is the chance that $12 is your potential bonus payment?

○ 0 in 100 (0%)

○ 40 in 100 (40%)

○ 60 in 100 (60%)

○ 100 in 100 (100%)

Submit Answer

Comprehension Questions

Option A

100 blue tickets (▎) which are labeled **$11**

Option B

40 purple tickets (▎) which are labeled **$30**

60 gold tickets (▎) which are labeled **$0**

Suppose that the task in the example above is the task-that-counts in Block 1, and you chose <u>Option B</u>.

What is the chance that $11 is your potential bonus payment?

○ 0 in 100 (0%)

○ 40 in 100 (40%)

○ 60 in 100 (60%)

○ 100 in 100 (100%)

Submit Answer

Comprehension Questions

Option A

100 blue tickets (|) which are labeled **$11**

Option B

40 purple tickets (|) which are labeled **$30**

60 gold tickets (|) which are labeled **$0**

Suppose that the task in the example above is the task-that-counts in Block 1, and you chose <u>Option B</u>.

What is the chance that $0 is your potential bonus payment?

○ 0 in 100 (0%)

○ 40 in 100 (40%)

○ 60 in 100 (60%)

○ 100 in 100 (100%)

Submit Answer

Begin Block 1

- Congratulations! You answered all the comprehension questions correctly! You are now ready to begin Block 1.

- Remember that your choices in Block 1 may affect your bonus payment if you are selected to receive one.

- Therefore, **it is in your best interest to answer the next questions thoughtfully.**

- If anything is unclear, please let us know through the Prolific anonymized internal messaging service. Otherwise, proceed to begin Block 1.

Next

Reminder of payment procedures

Option A	Option B
20 ▏	2 ▏
80 ▏	98 ▏

20 purple tickets (▏) which are labeled **$4**

80 gold tickets (▏) which are labeled **$0**

2 purple tickets (▏) which are labeled **$30**

98 gold tickets (▏) which are labeled **$0**

Which option do you **prefer**?

I prefer **Option A**   I prefer **Option B**

○   ○

Next

The screenshot below shows the pop-up that appears when subjects click the "Reminder of payment procedures" button in the choice task.

The order of Risky and Risk-Free tasks is randomized in Experiment 5. In the screenshots below, Block 1 contains Risk-Free tasks. We omit the screenshots for the comprehension questions, since they are identical to those reported for the Control condition of Experiment 5 in Appendix B.1.14. The instructions and comprehension questions remain unchanged when the Risk-Free tasks are presented in Block 2.

Block 1: Select-All-Tickets

For each task in Block 1, the computer will pay you by calculating the **AVERAGE** amount of money across all 100 ticket labels for **whichever option you've chosen**. That is, it will add up the monetary value of each of the 100 ticket labels and divide that sum by 100.

Next

Bonus Payment in Block 1

We will refer to the additional amount of money that you may receive at the end of the study, in addition to your participation fee, as the **potential bonus payment in Block 1**. The potential bonus payment in Block 1 will **only** be paid if certain conditions are met.

First, we will explain how the potential bonus payment in Block 1 is determined. Then, we will describe the conditions under which you may receive it.

Next

Potential Bonus Payment in Block 1

To determine your potential bonus payment in Block 1, we will randomly select one task from Block 1 as the **task-that-counts**. For example, the task-that-counts in Block 1 could be the one illustrated below.

Option A

Option B

100 |

60 |
40 |

100 blue tickets (|) which are labeled **$17**

60 purple tickets (|) which are labeled **$30**

40 gold tickets (|) which are labeled **$0**

If you preferred **Option A** in the task-that-counts, your potential bonus payment will be **(100 × $17) / 100 = $17**.

If you preferred **Option B** in the task-that-counts, your potential bonus payment will be **(60 × $30 + 40 × $0) / 100 = $18**.

Because any task in Block 1 could be selected as the task-that-counts, you should make each choice as if it will determine your potential bonus payment.

Next

Conditions for Your Bonus Payment in Block 1

You will receive your potential bonus payment in Block 1 at the end of the study **only if** two conditions are met:

1. Block 1 is randomly selected to determine your potential bonus payment.

2. After we compute your potential bonus payment from Block 1, we will draw a number between 1 and 5. If the drawn number is 1, you will be eligible to receive the potential bonus payment. Otherwise, you will not receive a bonus payment.

Next

Begin Block 1

- Congratulations! You answered all the comprehension questions correctly! You are now ready to begin Block 1.

- Remember that your choices in Block 1 may affect your bonus payment if you are selected to receive one.

- Therefore, **it is in your best interest to answer the next questions thoughtfully.**

- If anything is unclear, please let us know through the Prolific anonymized internal messaging service. Otherwise, proceed to begin Block 1.

Next

Task 1 of 10 – Block 1

Option A	Option B

20
80

16
84

20 purple tickets (▮) which are labeled **$23**

80 gold tickets (▮) which are labeled **$0**

16 purple tickets (▮) which are labeled **$30**

84 gold tickets (▮) which are labeled **$0**

Which option do you **prefer**?

I prefer **Option A**   I prefer **Option B**
○   ○

Next

B.1.17 Instructions and Comprehension Questions for Risky Tasks: Treatment Condition in Experiment 5

The order of Risky and Risk-Free tasks is randomized in Experiment 5. In the screenshots below, Block 1 contains Risky tasks. We omit the screenshots for the comprehension questions, since they are identical to those reported for the Control condition of Experiment 5 in Appendix B.1.15. The instructions and comprehension questions remain unchanged when the Risk-Free tasks are presented in Block 2.

Block 1: Select-One-Ticket

For each task in Block 1, the computer will **RANDOMLY** select one of the 100 tickets from whichever option you've chosen (each ticket in the option you chose is **EQUALLY** likely to be selected by the computer).

Next

Bonus Payment in Block 1

We will refer to the additional amount of money that you may receive at the end of the study, in addition to your participation fee, as the **potential bonus payment in Block 1**. The potential bonus payment in Block 1 will **only** be paid if certain conditions are met.

First, we will explain how the potential bonus payment in Block 1 is determined. Then, we will describe the conditions under which you may receive it.

Next

Potential Bonus Payment in Block 1

To determine your potential bonus payment in Block 1, we will randomly select one task from Block 1 as the **task-that-counts**. For example, the task-that-counts in Block 1 could be the one illustrated below.

Option A

100 |

100 blue tickets (|) which are labeled **$17**

Option B

60 |
40 |

60 purple tickets (|) which are labeled **$30**

40 gold tickets (|) which are labeled **$0**

If you preferred **Option A** in the task-that-counts, your potential bonus payment will be **$17**.

If you preferred **Option B** in the task-that-counts, your potential bonus payment will be **$30 with a 60 in 100 chance**, or **$0 with a 40 in 100 chance**.

Because any task in Block 1 could be selected as the task-that-counts, you should make each choice as if it will determine your potential bonus payment.

Next

Conditions for Your Bonus Payment in Block 1

You will receive your potential bonus payment in Block 1 at the end of the study **only if** two conditions are met:

1. Block 1 is randomly selected to determine your potential bonus payment.
2. After we compute your potential bonus payment from Block 1, we will draw a number between 1 and 5. If the drawn number is 1, you will be eligible to receive the potential bonus payment. Otherwise, you will not receive a bonus payment.

Next

Begin Block 1

- Congratulations! You answered all the comprehension questions correctly! You are now ready to begin Block 1.

- Remember that your choices in Block 1 may affect your bonus payment if you are selected to receive one.

- Therefore, **it is in your best interest to answer the next questions thoughtfully.**

- If anything is unclear, please let us know through the Prolific anonymized internal messaging service. Otherwise, proceed to begin Block 1.

Next

Task 1 of 10 – Block 1

Option A

20 |
80 |

20 purple tickets (|) which are labeled **$14**
80 gold tickets (|) which are labeled **$0**

Option B

10 |
90 |

10 purple tickets (|) which are labeled **$30**
90 gold tickets (|) which are labeled **$0**

Which option do you **prefer**?

I prefer **Option A**   I prefer **Option B**
○   ○

Next

B.1.18 Instructions and Follow-up Questions after Block 1 and Block 2: All Conditions in Experiment 5

We omit the follow-up questions on mathematical calculations after each block, as they are identical to those reported in Appendix B.1.10 for the Risky tasks in Experiment 4 and in Appendix B.1.12 for the Risk-Free tasks in Experiment 4. Unlike Experiment 4, however, Experiment 5 includes an example to clarify what it means for subjects to use mathematical calculations, together with two comprehension questions. Both the new example and the comprehension questions are shown in the screenshots below.

We first present the screenshots from Experiment 5 following the completion of the 10 choice tasks in Block 1.

Congratulations!

You've completed the 10 tasks of Block 1! Before proceeding to Block 2, we ask you to answer a few additional questions about how you made your choices in Block 1.

Your answers to these questions will not affect your potential bonus payment or your chances of receiving it, but they are important for our research. Therefore, we ask you to respond thoughtfully.

Next

Mathematical Calculations in Block 1

We are interested in knowing whether you used any **mathematical calculations** when making your decisions in **Block 1**. First, we will explain what we mean by using "mathematical calculations" with an example, and then we will ask whether you used any while making your decisions.

Next

Mathematical Calculations: Example

Example: Jamie is trying to decide between two phone plans. Plan A allows him to keep his old phone number and costs $0.20 per minute. Plan B doesn't allow him to keep his old phone number but costs only $0.10 per minute.

Instead of simply picking the plan that feels better or sticking with the brand he usually chooses, Jamie performs some **mathematical calculations** to guide his decision. He computes his total monthly cost for phone calls under each plan and concludes that Plan B would save him $7 per month. Jamie decides that saving $7 each month is more important than keeping his old phone number, so he chooses Plan B.

Since Jamie used mathematical operations such as multiplication to calculate each plan's monthly cost and subtraction to find the monthly savings with Plan B, we would say that Jamie used mathematical calculations when making his choice.

Next

Check your Understanding about the use of Mathematical Calculations

Before asking whether you used any mathematical calculations in making your decisions, we will ask two comprehension questions to check your understanding of when someone uses mathematical calculations to make a decision.

Next

Comprehension Questions

Question 1: Jordan is choosing between two apartment listings that cost the same. Apartment A is 600 square feet and close to work. Apartment B is 800 square feet but farther from work. Jordan feels that living in a larger apartment is more important than being close to work, so he chooses Apartment B.

Was Jordan using a mathematical calculation to make his decision?

○ Yes

○ No

Submit Answer

Comprehension Questions

Question 2: Emma is choosing between two grocery stores. One store sells a 12-pack of sparkling water for $5.40, and the other sells an 18-pack for $7.20. To decide which store offers the better deal, Emma divides the total price by the number of cans at each store and compares the cost per can.

Was Emma using a mathematical calculation to make her decision?

○ Yes

○ No

Submit Answer

Mental Shortcuts in Block 1

We are interested in knowing whether you used any **mental shortcuts** when making your decisions in **Block 1**. First, we will explain what we mean by "mental shortcuts" with an example, and then we will ask whether you used any while making your decisions.

Next

Mental Shortcuts: Example

Example: Chris must choose between attending two social events with two different groups of friends. Chris' enjoyment of social events depends on who else is there, but it's hard for him to think through the pluses and minuses of being with one large group of friends rather than another.

Instead of thinking through the pluses and minuses of being with one large group of friends rather than another, Chris uses the following **mental shortcut**: he makes his decision based entirely on which event his friend Parker plans to attend.

Since Chris bases his decision entirely on which event his friend Parker plans to attend even though he cares about who else will attend, we would say that Chris used a mental shortcut in making his choice.

[Next]

Check your Understanding about Mental Shortcuts

Before asking whether you used any mental shortcuts in making your decisions, we will ask two comprehension questions to check your understanding of when someone uses mental shortcuts to make a decision.

[Next]

Comprehension Questions

Question 1: Susan is deciding between two options. Option A is an apple. Option B is either a pear or an orange depending on whether Yellow comes before Blue on the color spectrum (i.e., the order of the rainbow). If Yellow comes before Blue then Option B is definitely a pear. If Yellow comes after Blue then Option B is definitely an orange. Susan doesn't recall the color spectrum precisely but remembers learning the acronym "ROY G BIV" for the colors of the rainbow. She thinks the "Y" stands for Yellow and the "B" stands for Blue, which would imply that Yellow comes before Blue, but she really isn't sure. She makes her decision as if Option B is a pear. She likes pears much more than apples so she chooses the pear.

Was Susan using a mental shortcut to make her decision?

○ Yes

○ No

[Submit Answer]

Comprehension Questions

Question 2: Jude is deciding between two boxes: Box A contains either an apple or a pear, while Box B contains either one dollar or ten cents. The boxes are labeled with their exact contents but the labels are written in a foreign language. Jude can read the language and is fairly certain Box A contains an apple while Box B contains one dollar. An apple is worth more than a dollar to him, so he chooses Box A.

Was Jude using a mental shortcut to make his decision?

○ Yes

○ No

Submit Answer

Mental Shortcuts in Block 1: Question

When making your decisions in Block 1, did you rely on any mental shortcuts?

○ Yes
○ No

Next

If subjects answer "Yes" to the question about shortcuts, they are asked to complete the following task.

Describe your Shortcut

One of the tasks from Block 1 is shown below. Using this task as an example, describe in detail any shortcuts you used in that block.

Option A

100 |

100 blue tickets (|) which are labeled **$4**

Option B

10 |
90 |
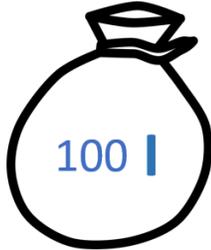
10 purple tickets (|) which are labeled **$30**

90 gold tickets (|) which are labeled **$0**

Please describe the shortcut you used.

Type your response here...

Next

Block 1 Completed!

You've completed Block 1! Click to learn about Block 2.

Next

Next, we present the screenshots from Experiment 5 following the completion of the 10 choice tasks in Block 2.

Congratulations!

You've completed the 10 tasks of Block 2! Before learning about your potential bonus payment and if you are eligible to receive it, we ask you to answer a few additional questions about how you made your choices in Block 2.

Your answers to these questions will not affect your potential bonus payment or your chances of receiving it, but they are important for our research. Therefore, we ask you to respond thoughtfully.

Next

Mathematical Calculations in Block 2

We are interested in knowing whether you used any **mathematical calculations** when making your decisions in **Block 2**. You will first review the example to remind yourself what we mean by using "mathematical calculations." After the example, we will ask whether you used any mathematical calculations while making your decisions.

Next

Mathematical Calculations: Example

Example: Jamie is trying to decide between two phone plans. Plan A allows him to keep his old phone number and costs $0.20 per minute. Plan B doesn't allow him to keep his old phone number but costs only $0.10 per minute.

Instead of simply picking the plan that feels better or sticking with the brand he usually chooses, Jamie performs some **mathematical calculations** to guide his decision. He computes his total monthly cost for phone calls under each plan and concludes that Plan B would save him $7 per month. Jamie decides that saving $7 each month is more important than keeping his old phone number, so he chooses Plan B.

Since Jamie used mathematical operations such as multiplication to calculate each plan's monthly cost and subtraction to find the monthly savings with Plan B, we would say that Jamie used mathematical calculations when making his choice.

Next

Mental Shortcuts in Block 2

We are interested in knowing whether you used any **mental shortcuts** when making your decisions in **Block 2**. You will first review the example to remind yourself what we mean by "mental shortcuts." After the example, we will ask whether you used any mental shortcuts while making your decisions.

Next

Mental Shortcuts: Example

Example: Chris must choose between attending two social events with two different groups of friends. Chris' enjoyment of social events depends on who else is there, but it's hard for him to think through the pluses and minuses of being with one large group of friends rather than another.

Instead of thinking through the pluses and minuses of being with one large group of friends rather than another, Chris uses the following **mental shortcut**: he makes his decision based entirely on which event his friend Parker plans to attend.

Since Chris bases his decision entirely on which event his friend Parker plans to attend even though he cares about who else will attend, we would say that Chris used a mental shortcut in making his choice.

Next

Mental Shortcuts in Block 2: Question

When making your decisions in Block 2, did you rely on any mental shortcuts?

○ Yes
○ No

Next

If subjects answer "Yes" to the question about shortcuts, they are asked to complete the following task.

108

Describe your Shortcut

One of the tasks from Block 2 is shown below. Using this task as an example, describe in detail any shortcuts you used in that block.

Option A

100 |

100 blue tickets (|) which are labeled **$4**

Option B

10 |
90 |
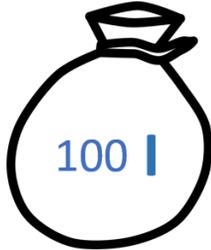
10 purple tickets (|) which are labeled **$30**

90 gold tickets (|) which are labeled **$0**

Type your response here...

Next

Hello and Welcome

Welcome, and thanks for your participation!

We are researchers from Caltech, Stanford, and The Ohio State University, inviting you to participate in a research study. The study should take roughly 15 to 20 minutes. Please click to review information about the study and to give your consent to participate.

Next

Prolific ID

You will be paid via Prolific for your participation in this study. In order to pay you, we need your Prolific ID.

Please enter your Prolific ID:

Next

Possible Rewards

You will receive $4.00 if you complete the entire study.

In addition to this payment, you have a 1 in 5 chance of being eligible to receive a bonus payment. The smallest possible bonus payment is $0 and the largest possible bonus payment is $30.

You will be informed of how your decisions will influence your bonus payment if you are eligible to receive one.

Next

This Study

In this study, you will complete 20 tasks, each of which involves your preferences over two options. The first option will always be called **Option A**. The second option will always be called **Option B**.

Option A and Option B refer to the possibility of winning monetary amounts ranging from $0 to $30 with some fixed chances. In each task you are asked to answer a few questions about Option A and Option B, and to decide which option you prefer.
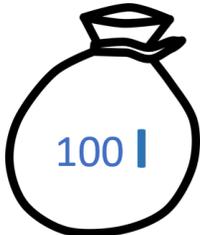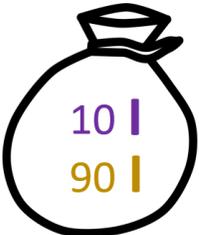
Next

Example Options

We will now introduce you to the types of options that you will see in this study and the choices you will make.

Next

Example Options 1

Consider the following example:

Option A	Option B
Draw one ticket from this bag:	Draw one ticket from this bag:

100 blue tickets (|) which pay **$4**

10 purple tickets (|) which pay **$28**
90 gold tickets (|) which pay **$2**

In this example:
- Option A allows you to draw one ticket from a bag containing 100 blue tickets (|). Each blue ticket pays $4.
- Option B allows you to draw one ticket from a bag containing 10 purple tickets (|) and 90 gold tickets (|). Each purple ticket pays $28, and each gold ticket pays $2.
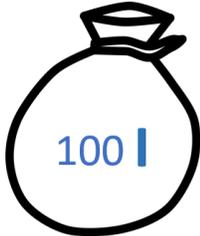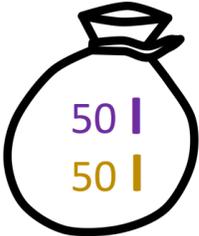
Therefore,
- Option A allows you to win $4 with a 100 in 100 chance.
- Option B allows you to win $28 with a 10 in 100 chance, or $2 with a 90 in 100 chance.

Next

111

Example Options 2

Consider the following example:

Option A	Option B
Draw one ticket from this bag:	Draw one ticket from this bag:

100 blue tickets (❘) which pay **$13**

50 purple tickets (❘) which pay **$y**

50 gold tickets (❘) which pay **$y**

$$y=(0.5\times2)+(0.5\times28)$$

In this example:

- Option A allows you to draw one ticket from a bag containing 100 blue tickets (❘). Each blue ticket pays $13.
- Option B allows you to draw one ticket from a bag containing 50 purple tickets (❘) and 50 gold tickets (❘). Each purple ticket pays $y and each gold ticket pays $y. In this example, $y is an amount equal to (0.5 × 2) + (0.5 × 28) that you can calculate.

We will first ask you to compute the value of $y, which is $15 in this example. Then, we will ask you to choose between Option A and Option B.

Therefore,

- Option A allows you to win $13 with a 100 in 100 chance.
- Option B allows you to win $15 with a 100 in 100 chance.

Next

How will you make choices?

In each task, we will present you with two options and ask you to choose between the following two answers:

1. I prefer Option A
2. I prefer Option B

Before making your choice, we may ask you to perform some calculations to determine the possible monetary rewards associated with each option.
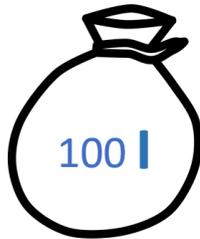
Next

Familiarize Yourself with the Study

- We now ask you to complete two training tasks to familiarize yourself with the questions and the choices you will make in the study.
- As you will realize from the training taks, in this study after you confirm an answer *you will not be able to modify it*.
- For this reason, we ask you to think carefully before confirming your answers. After the training, we will explain how your answers in the study will affect your bonus.

Next

Training Task 1 of 2
Make your Decision

Option A
Draw one ticket from this bag:

100 blue tickets (❙) which pay **$4**

Option B
Draw one ticket from this bag:

10 ❙
90 ❙

10 purple tickets (❙) which pay **$28**
90 gold tickets (❙) which pay **$2**

Which option do you **prefer**?
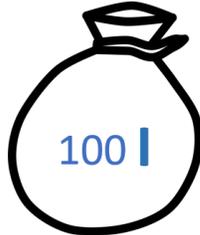
I prefer Option A   I prefer Option B
○   ○
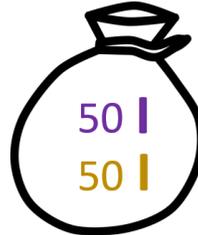
Next

Training Task 2 of 2
Make your Calculations

Option A

Draw one ticket from this bag:

100 |

100 blue tickets (|) which pay **$13**

Option B

Draw one ticket from this bag:

50 |
50 |

50 purple tickets (|) which pay **$y**

50 gold tickets (|) which pay **$y**

y=(0.5×2)+(0.5×28)

The value of y is: [＿＿＿]

Next

Option A	Option B
Draw one ticket from this bag:	Draw one ticket from this bag:

100 blue tickets (|) which pay **$13**
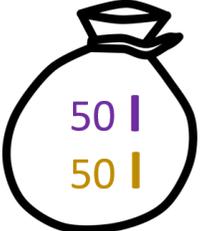
50 purple tickets (|) which pay **$y**

50 gold tickets (|) which pay **$y**

$$y=(0.5\times2)+(0.5\times28)$$

Based on your calculations:

- The value of y is: 15

Which option do you **prefer**?

I prefer Option A   I prefer Option B
○   ○

Next

B.2 Machine Learning Prompts

B.2.1 Experiment 4: Prompt for ChatGPT-5 Pro

Below we reproduce verbatim the instructions provided to ChatGPT-5 Pro to code subjects' open-ended explanations of their calculations in Experiment 4.

> Use this exact rubric and do not modify it after seeing any items. Treat each item independently. Do not adapt, tune, or calibrate based on examples shown during the session.

> You will classify short, free-text descriptions of how subjects made choices between two options (Option A and Option B). Each option is a bag with 100 tickets labeled

with at most two monetary amounts.

Question to answer for each description: Do you think that the subject made any mathematical calculations to arrive at their choice?

Answer format:

- Output exactly one word: Yes, No, or Uncertain

- No punctuation or extra words. One line per item.

- Classify each description independently. Do not use any previous items, frequencies, or patterns to influence your decision.

Decision rules (frozen; do not adapt)

1. Answer Yes only if the subject's description contains any explicit mathematical calculation such as addition, subtraction, multiplication, or division.

2. Answer Uncertain if subjects claim to have made a mathematical calculation without actually describing it.

3. Answer No if the subjects are neither describing nor referencing a mathematical calculation.

Important notes (frozen; do not adapt)

- Treat explicit math words/symbols ("add", "sum", "difference", "times", "×", "/", "*", "=", "divide", "average/mean", "expected value/EV", "computed 20%", "0.2×\$50 + 0.8×\$0") as mathematical calculations.

- Do not infer calculations from qualitative-only reasoning ("gut", "risk-averse", "felt safer", "bigger prize", "higher chance").

- Do not infer calculations from listing or recalling the options or probabilities.

- If the subject says "did some math", "calculated", or "used EV" without showing any operation, numbers, formula, or computed result, output Uncertain.

- Option A is always expressed as a certain prize. Therefore, the fact that a subject knows the value of Option A is not useful for establishing whether a subject made a mathematical calculation.

- If a subject refers to Option B as a certain prize, assume they had to make a calculation to compute it. If they did not explain their calculation explicitly, report Uncertain. Example: If Option B is a bag with 20 tickets labeled $120 and 80 tickets labeled $0, then stating "Option B is $24 for sure" implies computing 0.2 * $120 + 0.8 * $0 = $24; if they do not show this computation, output Uncertain.

- Do not update or recalibrate these rules based on any items you are asked to classify.

B.2.2   Experiment 5: Prompt for ChatGPT-5 Pro

Below we reproduce verbatim the instructions provided to ChatGPT-5 Pro to code subjects' open-ended explanations of their shortcuts in Block 1 of Experiment 5.

"Use this exact rubric and do not modify it after seeing any items. Treat each item independently. Do not adapt, tune, or calibrate based on examples shown during the session.

You will classify short, free-text descriptions of how subjects made choices between two options (Option A and Option B). Each option is a bag with 100 tickets labeled with at most two monetary amounts.

Task types (you will not be told which one the description came from):

- Risky Tasks: After choosing, one ticket is drawn at random and the subject gets that amount.

- Risk-Free Tasks: After choosing, the subject receives the average value of the 100 tickets.

Question to answer for each description: Do you think the subject's preferences over risk and uncertainty impacted their decision-making process?

Answer format:

- Output exactly one word: Yes, No, or Uncertain.

- No punctuation or extra words. One line per item.

- Classify each description independently. Do not use any previous items, frequencies, or patterns to influence your decision.

Decision rules (frozen; do not adapt)

1. Answer Yes only if the subject's reasoning shows that perceived risk/uncertainty affected the choice. This requires at least one of:

   - They explicitly compare a guaranteed/sure outcome to a chance-based outcome and say that this impacted their choice.

   - They describe post-choice randomness (drawing, chance of getting some amount) and say that this riskiness influenced their choice.

   - They explicitly mention meta-uncertainty about their own calculation that changed their choice (e.g., "I chose Option A because I am not sure that I computed the average of Option B correctly").

2. Answer No if the reasoning is clearly independent of risk preferences, such as:

   - Pure arithmetic/expected-value calculations or tallying (e.g., multiply counts by amounts, add and divide by 100), including "ignoring zeros because 0 adds nothing," with no stated preference about risk.

   - Always choosing the higher average/amount regardless of chances or lotteries (e.g., "I prefer 10 to 5").

   - Deterministic comparisons only, with no mention that riskiness affected the choice.

3. Answer Uncertain if it's ambiguous whether risk preferences played a role, including:

   - Mentions of percentages/odds/chances used as part of a calculation without saying that riskiness influenced the choice (unless it's clear that it was Risk-Free, according to the criteria above).

   - Descriptions compatible with either Risky or Risk-Free tasks without a clear statement that randomness (or perceived randomness) changed their preference.

   - Vague statements like "better odds" or "more likely" without stating that the risk aspect drove the choice.

Important notes (frozen; do not adapt)

- Do not infer risk preference from the mere presence of zeros, percentages, or counts; those may be used to compute averages in Risk-Free tasks.

- Only classify Yes when their preference is shaped by perceived uncertainty (e.g., preferring guaranteed or known outcomes). Otherwise use No (for clearly purely arithmetic) or Uncertain (if ambiguous).

- Do not update or recalibrate these rules based on any items you are asked to classify."