Working Paper in Economics

202402

A Systematic Test of the Independence Axiom Near Certainty

Ritesh Jain*, Kirby Nielsen†

*University of Liverpool Management School, Liverpool, email: ritesh.jain@liverpool.ac.uk
†Division of the Humanities and Social Sciences, California Institute of Technology; kirby@caltech.edu.

A Systematic Test of the Independence Axiom Near Certainty*

Ritesh Jain† Kirby Nielsen‡

February 22, 2024

Abstract

A large literature has documented violations of expected utility consistent with a preference for certainty (the "certainty effect"), but recent studies question the prominence of this phenomenon. We design an experiment using lotteries spanning over the entire probability simplex to establish the prevalence of the certainty effect relative to its opposite. We find that violations of independence consistent with the *reverse* certainty effect are much more common than violations consistent with the certainty effect. Results hold as we test robustness along three dimensions: varying the mixing lottery, moving slightly away from certainty, and having a zero outcome.

Keywords: independence axiom; expected utility theory; certainty effect; Allais Paradox; common ratio effect

JEL classification: C79, D82

*We want to thank Yaron Azrieli, Doug Bernheim, Anujit Chakraborty, Paul J. Healy, Collin Raymond, John Rehbeck, Charles D. Sprenger, Colin Sullivan, and the participants of the experimental reading group at the Ohio State University for providing helpful feedback on this project. All errors remain our own. Research support was provided by the Decision Science Collaborative of the Ohio State University. This study was approved by the Institutional Review Boards at Academia Sinica and Stanford University. Ritesh Jain gratefully acknowledges financial support from Ministry of Science and Technology Grant, Taiwan, #108-2410-H-001-038. Jung Hsin, Cheng and Cheng-Hsiao Cheng provided valuable research assistance.

†University of Liverpool Management School, Liverpool, *email: ritesh.jain@liverpool.ac.uk*

‡Division of the Humanities and Social Sciences, California Institute of Technology; *kirby@caltech.edu*

> Consequently, I viewed the principle of independence as incompatible with the preference for security in the neighborhood of certainty shown by every subject.
>
> Maurice Allais (2008)

I. Introduction

Experimental evidence has shown that individuals reliably violate the independence axiom, the central tenet of expected utility theory (EU).[1] In 1953, Maurice Allais proposed one of the earliest, and still to-date most famous, counter-examples, now known as the "Allais Paradox."[2] Ask a decision maker the following two binary choices:

	Option A:		Option B:
Q1:	100% chance of \$100 million	vs.	98% chance of \$500 million 2% chance of \$0
Q2:	1% chance of \$100 million 99% chance of \$0	vs.	0.98% chance of \$500 million 99.02% chance of \$0

Allais hypothesized that many individuals would choose Option A in the first decision and would choose Option B in the second decision. This choice pattern violates the independence axiom, since the lotteries in Question 2 are the same as the lotteries in Question 1, just multiplied by a common chance of the low outcome.[3] This choice pattern is now known as the "common ratio effect," and decision problems of this form have been studied extensively in the literature with many formulations confirming Allais's intuition (see Kahneman and Tversky, 1979 for an early example).

Allais attributed these violations to a *preference for security*, quoted above, now referred to as the "certainty effect" (Kahneman and Tversky, 1979). Kahneman and

[1] Independence states that for any three lotteries p, q, and r, and any number λ in [0,1], if p is preferred to q, then $\lambda p+(1-\lambda)r$ is preferred to $\lambda q+(1-\lambda)r$. That is, mixing both lotteries p and q with a common lottery r, and in common proportions, should not change the relative preference between p and q.

[2] Both the "common ratio" and "common consequence" violations of Independence are often referred to as the Allais Paradox. We focus on the common ratio effect in this paper.

[3] To see how the choices violate independence, let $\lambda = 0.01$ and r be 100% chance of \$0.

Tversky describe the certainty effect as the phenomenon in which "people overweight outcomes that are considered certain, relative to outcomes which are merely probable" (Kahneman and Tversky, 1979, p. 265). The intuition behind the certainty effect in the Allais Paradox is that the preference of Option A over Option B in Q1 is driven, in part, by the fact that Option A offers a sure payoff. When both options are risky, as in Q2, neither offers the appeal of certainty, so preferences can reverse. Allais's original intuition, shared by many and confirmed by experimental evidence, has led to large theoretical and experimental literatures in search of a descriptive non-expected utility model. We review these papers in Section II.[4]

While previous papers have shown violations of independence consistent with the certainty effect, we aim to test independence *systematically* to see the relative proportion of EU violations that are in the direction predicted by the certainty effect. Given the prominence of the Allais Paradox, a persistent thread in the literature is that the certainty effect is the *main* contributor to independence violations. For example, Schmidt (1998) says that "the bulk of observed violations of the independence axiom is due to the certainty effect." Allais himself conjectured that, far from certainty, individuals would act as expected utility maximizers (Allais, 1953; Andreoni and Sprenger, 2010). This suggests that common ratio violations of independence would be relatively uncommon absent a certain option, and that violations would be relatively uncommon in situations where individuals prefer risk over certainty.[5]

On the other hand, more recent work has shown examples of the *reverse* certainty effect (Starmer, 1992; Humphrey and Verschoor, 2004; Blavatskyy, 2013), and a recent meta-analysis by Blavatskyy et al. (2022) shows that the common ratio effect is not universal and can be affected by various choices in the experimental design and parameters. Despite these examples, most common ratio tests involving certainty focus on a small region of the probability simplex. Indeed, among all of the papers surveyed in Blavatskyy et al. (2022), about one-third use the exact parameter configuration that was used in Kahneman and Tversky (1979) (as noted by McGranaghan et al., 2023). This suggests need for a a broader exploration of tests of the indepen-

[4]The certainty effect also has been invoked to explain behaviors outside the domain of simple lotteries, such as present bias (Halevy, 2008) and aversion to gradual pieces of information (Dillenberger, 2010).

[5]For example, if we were to reduce the prize in Option A from 100 million to 10 million in the example above, we would expect most people to prefer Option B in Question 1. Since the appeal of certainty does not drive the preference in Question 1, intuition relying on certainty effect suggests that we would not see violations of expected utility anymore.

dence axiom. We aim to document the prevalence of the certainty effect—and its opposite—in a unified and systematic test of independence throughout the simplex. We view our paper as complementary to Blavatskyy et al. (2022) and results therein, providing supportive evidence from a unified experimental paradigm.

We fix a probability simplex, which, in our experiment, is the set of possible lotteries over {\$10, \$20, \$30}. We pick 45 lotteries uniformly across this simplex. Subjects face binary choices between \$20 for sure and a lottery selected at random from these uniformly-distributed risky lotteries. We then mix both alternatives according to three different mixture weights to see if preferences reverse, constituting a violation of independence.

Given the wide range of lotteries we sample, subjects will prefer \$20 to the risky lottery in some questions, while in other questions they will prefer the risky lottery to \$20. Questions of the latter type are relatively uncommon in the literature; as noted by Blavatskyy (2010), in most experimental studies, "the sure monetary payoff is deliberately selected... so that the majority of people are likely to choose the sure alternative over the risky lottery. In a sense, the common ratio effect is already pre-programmed in this setup."[6] Instead of deliberately selecting questions this way, our systematic test allows us to detect independence violations when certainty is preferred to risk—consistent with the certainty effect—and compare them to those when risk is preferred to certainty, consistent with the reverse certainty effect.[7] The standard parameterizations in the literature tend to deliver situations where subjects prefer \$20 to the risky lottery, while our design allows us to compare these situations to parallel situations where subjects prefer the risky lottery.

We find that *reverse* certainty effect violations are far more common than certainty effect violations in our data. Conditional on preferring certainty to risk, individuals violate independence 15% of the time. In stark contrast, individuals violate independence almost 40% of the time conditional on preferring risk to certainty. This result holds when we control for "strength-of-preference" effects, indicating that we cannot

[6] For example, in one of the most well-known examples of the common ratio effect Kahneman and Tversky, 1979, 80% of subjects prefer the sure payment in Q1. In their paper, as in most papers in the literature, there is no equivalent comparison question where most subjects prefer the risky lottery in Q1.

[7] Following this intuition, Blavatskyy (2010) also includes binary comparisons where risk is likely preferred to certainty, and finds evidence of the reverse common ratio effect. While both common ratio and reverse common ratio examples have been documented, our contribution is to test the relative frequencies of these violations and their interaction with certainty.

attribute these different patterns to pure noise. Our design also allows us to look at a more rigorous test of the independence axiom, comparing choices across all four mixing probabilities. This analysis confirms our main results, with most violations coming from instances where individuals choose the risky lottery over certainty and then switch to choosing the safer lottery as the alternatives are mixed away from certainty.

We test the robustness of these results along three dimensions. First, we move slightly away from certainty by comparing with a lottery that gives \$20 with 90% chance (otherwise a 5% chance of \$30 and a 5% chance of \$10) rather than \$20 with certainty. Second, we vary the "mixing lottery" from one in the spirit of the Allais Paradox—mixing with the lowest possible outcome—to one less commonly studied—mixing with a lottery that puts equal weight on all three outcomes. Finally, we run a version of the experiment with \$0 as the lowest payoff—in contrast with \$10 in our main experiments. Overall, our results are robust to these perturbations.

Our results contribute to a large experimental literature testing the independence axiom in common ratio questions. Ours benefits from being a systematic test around certainty, providing evidence on the frequency and location of independence violations. We believe our results will be particularly useful to incorporate into theoretical models of choice under risk. Recent theories seek to characterize and axiomatize the certainty effect in building descriptive models of choice (Cerreia-Vioglio et al., 2015). Our results suggest that these theories may miss an important pattern of behavior: In our data, we could explain significantly more choices by modeling the exact opposite phenomenon.

II. Literature Review

Our paper contributes to the literature on the certainty effect as well as the literature testing and relaxing expected utility theory (EU). We quickly review the theoretical literature before turning to the most closely related experimental papers.

The prominence of the certainty effect in the experimental literature has led to theoretical work attempting to capture the documented choice patterns. Some of the popular alternative theories to EU that are able to accommodate the certainty effect in common ratio choices include disappointment aversion (Gul, 1991), cumulative prospect theory (Tversky and Kahneman, 1992), rank dependent utility theory

(Quiggin, 1982), and cautious expected utility theory (Cerreia-Vioglio et al., 2015). While many of these theories can accommodate the reverse certainty effect in the "opposite" way, most, if not all, were designed to capture the certainty effect.

Each non-EU theory listed above can accommodate the certainty effect, but we highlight here Cautious Expected Utility (Cerreia-Vioglio et al., 2015) because it was designed explicitly to characterize certainty-effect preferences. Cerreia-Vioglio et al. weaken independence by requiring it to hold only when risk is already preferred to certainty, allowing for independence violations when certainty is preferred to risk. Formally, they replace independence with an axiom, Negative Certainty Independence (NCI), first introduced in Dillenberger (2010). NCI states that for all lotteries $p, q \in \Delta(X)$, prizes $x \in X$, degenerate lotteries δ_x, and probabilities $\lambda \in [0,1]$,

$$p \succeq \delta_x \Rightarrow \lambda p + (1-\lambda)q \succeq \lambda \delta_x + (1-\lambda)q.$$

This requires that independence holds when a lottery, p, is preferred to \$$x$ for sure, but does not require independence to hold in the opposite case where certainty is preferred to risk. As such, this theory exactly characterizes the preference for certainty underlying the common ratio effect and other well-documented patterns of behavior. Cautious expected utility theory is a very appealing model that can accommodate other behavioral phenomena (e.g., the endowment effect), but it is important to note that its central axiom characterizes the certainty effect.

The authors cite the large body of evidence on the certainty effect, but point out that "no comprehensive tests of NCI have been conducted thus far" (Cerreia-Vioglio et al., 2015, p. 713). We see our paper as a natural step in this dialogue between theory and experiments. Our results suggest that the certainty effect is not always the main obstacle for the independence axiom—in our data, the *reverse* certainty effect is the main obstacle for independence. Indeed, in a more recent paper, Cerreia-Vioglio et al. (2020) characterize preferences with the opposite axiom, Positive Certainty Independence (PCI), which requires independence hold instead when a sure payment is preferred to a risky lottery but allows for violations otherwise. One could interpret our experiment as a test of the relative prevalence of NCI and PCI violations, finding more violations of NCI than PCI.

Hand-in-hand with theoretical advancements is a large experimental program aimed at testing these theories. The experimental literature is vast (see, for example,

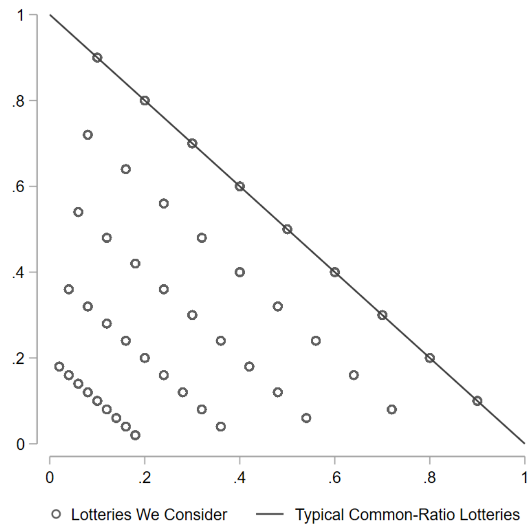


Figure I: Lotteries Used in Our Experiment
Notes: These points represent the main lotteries used in our experiment. Each of these lotteries is compared against receiving $20 for sure. In contrast, in most of the literature, common ratio questions focus on comparing certainty with lotteries along the hypotenuse.

Conlisk, 1989; Harless, 1992; Starmer, 1992; Neilson, 1992a,b; Sopher and Gigliotti, 1993; Camerer and Ho, 1994; Loomes and Sugden, 1998; Schmidt, 1998; Humphrey and Verschoor, 2004; Huck and Müller, 2012; Incekara-Hafalir et al., 2020), so we cannot summarize every paper here. Camerer (1995) and Starmer (2000) review the older literature, and we refer the interested reader to those surveys. A paper that is closely related in spirit to ours, but very different in methodology, is a recent meta-analysis by Blavatskyy et al. (2022). They survey 143 common ratio experiments involving a certain outcome. They find that the prevalence of the common ratio effect (synonymous with the certainty effect in these questions) varies predictably with features of the experimental design, including the value of the common ratio itself, real vs. hypothetical stakes, etc. Given that the papers in this meta-analysis vary on many dimensions in addition to these identified, we conduct our experiment to validate these findings in a simple and controlled environment. Importantly, we focus on the location of common ratio and reverse common ratio effects in the simplex; therefore, we hold fixed many of the factors that the meta-analysis identifies as relevant (e.g., real vs. hypothetical stakes, presenting lotteries as simple probabilities vs. compound lotteries or frequencies, and the distance between the middle and high outcome).

Perhaps more importantly, we aim to span the space of lotteries in a more systematic way. Most common ratio tests that involve certainty compare a certain outcome to one lottery or a few lotteries over a low and high prize (as in the papers surveyed in Blavatskyy et al., 2022). In contrast, we compare a certain outcome to 45 points uniformly selected across the simplex; see Figure I for a visualization of our lotteries. Some of these are lotteries over the low and high outcome (i.e., lotteries on the hypotenuse), but most also put probability on the middle outcome. Interior lotteries have certainly been used in expected utility tests—and in common ratio tests—but rarely in questions involving certainty (Camerer, 1989; Harless, 1992; Sopher and Gigliotti, 1993). Furthermore, many papers in the literature (including a large majority of the papers surveyed in Blavatskyy et al., 2022) use the same or similar parameterizations when studying the common ratio effect, e.g., putting 80% probability on the high outcome of the risky lottery in the unmixed question following Kahneman and Tversky (1979). Taken together, this highlights that, despite the abundance of common ratio tests in the literature, we know very little about the certainty effect in a large part of the simplex; most papers in the literature include only a single common ratio effect question with certainty, and there is no uniform variation in the probabilities of the lotteries. This gap emphasizes the need for a systematic test.

By selecting our lotteries uniformly, we cover much more area than the previous literature has done which allows us to identify potential regions of the simplex that exhibit different patterns. We can also compare the "amount" of certainty effect to the amount of reverse certainty effect in a meaningful way. Furthermore, a few recent papers in various domains suggest that the certainty effect requires true certainty (probability one), and differs predictably from "near certainty" (Halevy, 2008; Andreoni and Harbaugh, 2010; Andreoni and Sprenger, 2010, 2011, 2012). To explore this, we also systematically compare these same 45 lotteries to a lottery that is "close" to certainty. Also importantly, we cover (within-subject) areas where individuals are likely to prefer the riskier option and areas where individuals are likely to prefer the safer option, which allows us to compare the prevalence of common ratio and reverse common ratio violations. We discuss these design details more below.

Thus, while a large number of common ratio tests have been conducted in the literature, and while there is already evidence of the reverse common ratio effect and the reverse certainty effect (particularly from Blavatskyy, 2010 and Blavatskyy, 2013), our goal in this paper is to replicate and establish these findings in a systematic manner so that we can interpret the prominence of these various choice patterns. We fix the lottery prizes, framing, and other aspects that vary across existing studies. We ask a large number of questions involving certainty and near certainty to obtain a rich picture of the certainty effect, and we include questions where both risk is preferred to certainty and certainty is preferred to risk in order to obtain a deeper understanding of adherence to independence in the simplex. Furthermore, we include all of this within-subject, so that we can conduct individual-level tests as well as aggregate tests. To the best of our knowledge, our paper is the first to conduct such a widespread test of the certainty effect in a single simplex. We conclude that the certainty effect is not the main contributor to common ratio violations of independence, and that the reverse certainty effect is a nontrivial phenomenon in observed choices.

III. Theoretical Framework

We describe the theoretical framework in the context of our experimental design. All questions involve lotteries over US dollars. The set of possible prizes in our experiment is $X = \{10, 20, 30\}$. We represent the set of lotteries with prizes in X by $\Delta(X)$, with weak preferences $\succeq$ defined over $\Delta(X)$. We denote generic prizes in X by x, y, z, and denote generic lotteries in $\Delta(X)$ by p, q, r, s. The probability of receiving prize x under lottery p is denoted $p(x)$. We represent the three-outcome lottery, p, giving \$10 with probability $p(10)$, \$20 with probability $p(20)$, and \$30 with probability $p(30)$ by $(\$30, p(30); \$20, p(20); \$10, p(10))$. We represent the degenerate lottery giving \$$x$ for sure as δ_x.

The independence axiom states that for all $p, q, r \in \Delta(X)$ and for all $\lambda \in [0,1]$,

$$p \succeq q \Leftrightarrow \lambda p + (1-\lambda) r \succeq \lambda q + (1-\lambda) r.$$

We consider only "one-stage" lottery mixtures, rather than two-stage compound lotteries.[8] In our experiment, we will test the independence axiom by presenting subjects with binary choices over these one-stage lotteries.

There are two ways individuals can violate independence when one option is certain. The *certainty effect* (CE) captures the idea that individuals place disproportion-

[8]In other words, we study *mixture independence*, rather than *compound independence*, as defined in Segal (1990).

ate weight on an outcome when it is certain (Kahneman and Tversky, 1979). Individuals with a preference for certainty will be more likely to violate independence when certainty is preferred to a risky lottery before mixing. The intuition is that the preference of δ_x over p may be driven, in part, by the certainty appeal of receiving \$$x$ for sure. When these lotteries are mixed as in independence, $\lambda\delta_x + (1-\lambda)r$ does not carry the same certainty appeal, which might result in a preference for $\lambda p + (1-\lambda)r$ over $\lambda\delta_x + (1-\lambda)r$. When individuals violate independence in this way, we call it a "CE" violation.

The *reverse certainty effect* (RCE) is the exact opposite pattern. This refers to an individual who chooses p over δ_x and then chooses $\lambda\delta_x + (1-\lambda)r$ over $\lambda p + (1-\lambda)r$. We refer to this as an "RCE" violation. Our main research question is documenting the prevalence of independence violations and comparing the frequency of these two patterns of violations in a systematic way.

IV. Experimental Design

We chose three payments—\$10, \$20, and \$30—and all questions involve lotteries over these three payments.[9,10] In order to compare CE and RCE violations, we needed to ask questions where a risky lottery is likely to be preferred to certainty, as well as questions where certainty is likely to be preferred to the risky lottery. To ensure this, we selected 45 points uniformly across the simplex. These 45 questions are denoted with circles in the top left graph of Figure II, and we refer to these as the "unmixed lotteries." We asked binary questions comparing these lotteries against a sure payment of \$20: a choice of p vs. δ_{20}.

[9]There is evidence that independence violations under certainty are more prevalent with large stakes than small stakes, reviewed in Cerreia-Vioglio et al. (2015). Therefore, we wanted to pick payments that were fairly high. Our payments averaged to around \$20 per person, and sessions took only 30 minutes. Subjects knew this ahead of time. We felt this \$40/hr average payment would be reasonably high stakes based on the literature.

[10]One limitation of our study is that we do not consider systematic variation in the payment outcomes, aside from the Zero Treatment noted below. We believe systematic variation in payments—analogous to our systematic variation in probabilities—is an important avenue for future work.

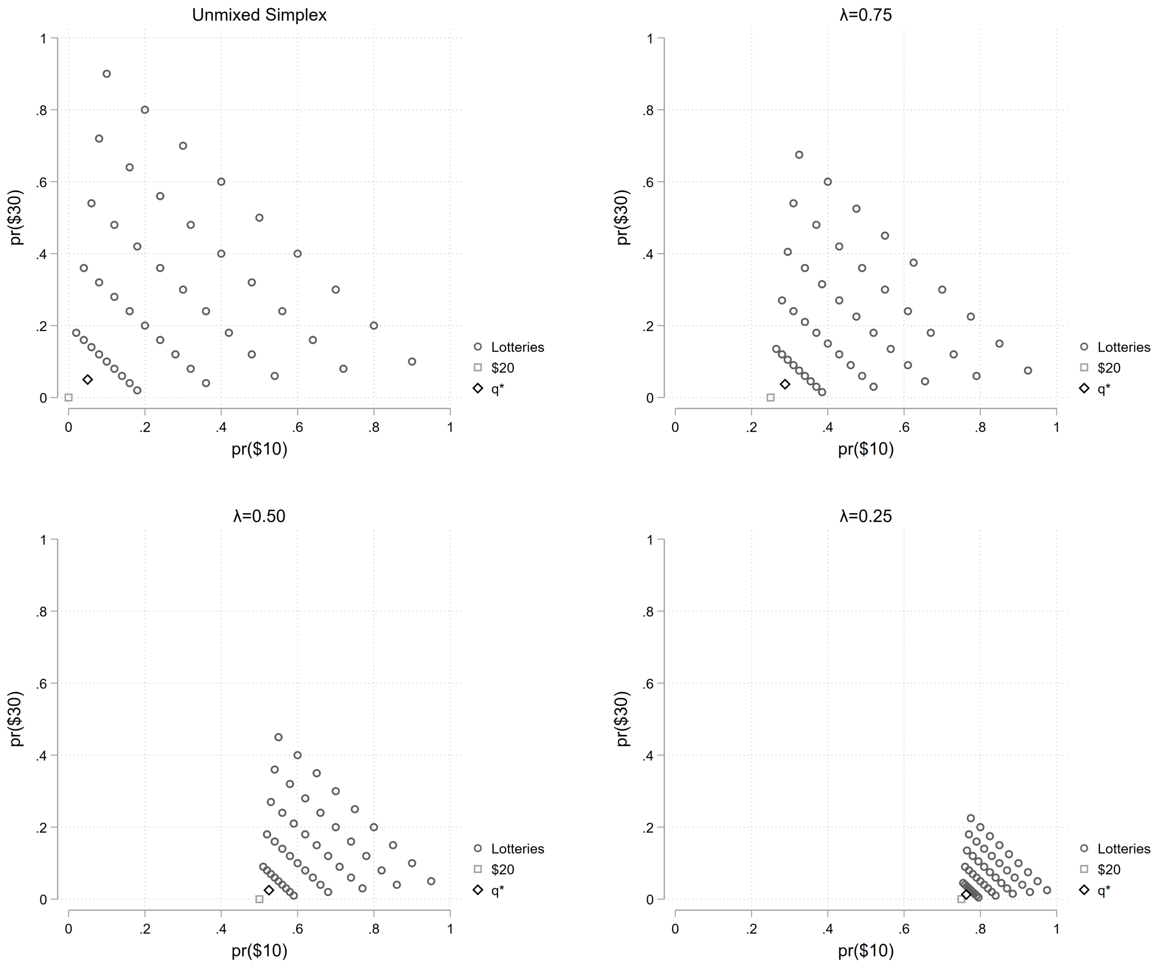


Figure II: Questions

Notes: The top left panel shows the "unmixed" questions, and the other three panels show the lotteries after mixing. Subjects report binary preferences between δ_{20} or q^* and the lotteries shown. Each subject makes 68 binary choices.

To test independence, we mixed these lotteries with $r = (\$10, 1)$. We used three different mixing probabilities, $\lambda = \{0.25, 0.50, 0.75\}$. This results in 45 new binary choices for each value of λ: λp vs. $\lambda\delta_{20}$. The lotteries after mixing are shown in the remaining three panels of Figure II.[11]

We test the robustness of the certainty effect by moving δ_{20} slightly away from certainty, comparing these unmixed lotteries against $(\$30, 0.05; \$20, 0.90; \$10, 0.05)$, denoted by a diamond in Figure II. This lottery is "close" to a sure payment of $20,

[11]Though we sometimes refer to lotteries "before" or "after" mixing for ease of exposition, there is no temporal component to the experiment. As we explain, questions were presented to subjects in random order.

but does not offer the same security. For simplicity, we'll call this lottery q^*, and we'll refer to these questions as "near-certain." Subjects face both certain and near-certain questions, as we explain below. When talking about alternatives in these binary comparisons, we refer to p and λp as the "risky lotteries" and refer to either δ_{20} and $\lambda\delta_{20}$, or q^* and λq^*, as the "safer lotteries." We reserve "certainty" only for δ_{20}.

In total, we have 360 possible questions–the 45 unmixed lotteries compared with \$20 in the *certain* condition (45 questions) and compared with q^* in the *near-certain* condition (45 questions). These 90 questions comprise the "unmixed" comparisons, and each is mixed by $\lambda = \{0.75, 0.50, 0.25\}$ $(90 \times 4 = 360)$. Since it might be unreasonable for individuals to answer all 360 questions, each subject instead answered 68 binary questions from the set of 360 possible questions.[12] To perform the random selection, we created a bank of 90 questions—the 45 unmixed lotteries compared against \$20 and the same 45 unmixed lotteries compared against q^*. We randomly and independently selected 17 of these 90 questions for each subject. For those 17 questions, we asked subjects the unmixed question and all three $\lambda = \{0.25, 0.50, 0.75\}$ mixtures. This gives a total of $17 \times 4 = 68$ binary choices per subject.

This random selection process helps ensure that, on average for each subject, we will have observations where the risky lottery is preferred to the safer lottery and vice versa, and we will also have observations for both certain and near-certain comparisons. It also allows us to test independence more rigorously than in single binary choices, as independence requires an individual to choose either the risky or safer option in all four λ comparisons. This design also rules out the possibility that independence violations result from indifference, which is a common critique of experiments that observe preference reversals (Blavatskyy, 2010). Given the number and diversity of questions we ask, systematic and persistent violations of independence cannot be explained through indifference.[13]

Finally, we conducted two between-subject treatments. The first, which we have explained above, mixes lotteries with the bottom right of the simplex, $r = (\$10, 1)$. This is closest in spirit to the original Allais Paradox where the lotteries were mixed with the lowest possible payoff. We refer to this as the "Allais Mix" treatment. To

[12] 68 was calibrated based on duration of the experiment.

[13] Given the structure of our lotteries, subjects could be exactly indifferent to \$20 on one "ray" from the origin, which is at most 5 questions.

further test the robustness of independence violations, we ran a separate treatment that mixes lotteries instead with the midpoint of the simplex, $r = (30, \frac{1}{3}; 20, \frac{1}{3}; 10, \frac{1}{3})$, which we refer to as the "Middle Mix" treatment. We include this mixture lottery since this region is relatively under-explored in the literature.[14] Furthermore, cautious expected utility, which is characterized by the certainty effect, predicts that the indifference curve through the origin is steepest and linear, so the middle mix treatment allows us to focus on this region of the simplex (Cerreia-Vioglio et al., 2015). Each subject participated in either the Allais Mix or Middle Mix treatment, but not both. We defer explanation of the Middle Mix treatment to Section V.

IV.A. Procedures

We present results from 14 experimental sessions with a total of 265 subjects, 118 in the Allais Mix treatment and 147 in the Middle Mix treatment. Subjects were mainly undergraduates from Ohio State University, recruited using ORSEE (Greiner, 2004). The experiment was programmed using z-Tree (Fischbacher, 2007). Sessions lasted approximately 30 minutes and subject payments averaged $20.

The experimenter read instructions out loud to all subjects. Instructions explained the binary choices and how the probabilities would translate into payoffs. Computer screens displayed the written probabilities and payoffs, as well as color-coded pie charts. Figure VIII in the Appendix shows a screenshot, and we also include instructions in the Appendix. All 68 questions were displayed in random order, randomized separately across subjects. In particular, it was not necessarily the case that subjects first saw the unmixed question, then the $\lambda = 0.75, 0.5, 0.25$ questions, and subjects were not aware that questions were related to each other in any way. Furthermore, each question was displayed on a separate screen, and we randomized the left-right screen position of the risky and safer lottery.

Subjects were paid after everyone in the session completed the experiment. We used physical randomization devices to determine payments, and subjects knew this ahead of time. The experimenter rolled two 10-sided dice at the front of the room to generate a number 1–68.[15] This determined the random question that would be

[14] There are other regions of the simplex that remain under-explored in the literature, including mixing with the best possible outcome in the lottery. We believe this is an interesting open avenue for future work.

[15] If the number came up larger than 68, she rolled again.

paid. Then, the experimenter rolled the dice again to generate a number 1–100 to resolve any risk in the randomly-selected lottery. Subjects were paid the realization from whichever lottery they had chosen in the randomly-selected decision. Therefore, subjects were paid based on exactly one decision they made in the entire experiment.

This payment method, denoted the "random payment selection" (RPS) mechanism, has been used in many binary choice experiments. As discussed in Azrieli et al. (2019), by using the RPS mechanism, we are assuming that *compound independence* holds (Segal, 1990).[16] Brown and Healy (2018) give evidence that compound independence holds when presenting choices on separate screens, as we do in our experiment. Segal (1990) shows that compound independence and reduction of compound lotteries together imply *mixture independence*, which is the form of independence we study. Therefore, by using this payment mechanism, we assume that individuals do not always satisfy reduction of compound lotteries, since we observe violations of mixture independence.

Given that we study the certainty effect, there might be a worry that individuals don't view certainty here as truly certain given that they are paid for one random decision. We acknowledge this weakness, and view our paper as a comparison to the rest of the literature which uses a similar payment structure. We rely on prior work that finds no difference in independence violations when comparing across different incentive structures (Starmer and Sugden, 1991; Cubitt et al., 19989). In addition, Nielsen (2020) studies preferences over the timing of uncertainty resolution—also motivated by the NCI axiom (Dillenberger, 2010)—and finds no differences in preferences for one-shot resolution when using the random payment selection mechanism versus implementing a single decision.

V. Results

We focus our main results on the *certain* comparisons in the Allais Mix treatment. These are questions where subjects chose between a risky lottery and \$20 for sure in the unmixed question, and separately made the same binary comparison when both were mixed with 100% chance of \$10, for three different mixing probabilities $\lambda =$

[16]Let A and B be two-stage lotteries over the simple lotteries in our experiment. That is, A=$(\alpha_p, p; \alpha_q, q; ...; \alpha_r, r; ...; \alpha_s, s)$ is a two-stage lottery that gives simple lottery p with probability α_p, lottery q with probability α_q, etc. Let B=$(\alpha_p, p; \alpha_q, q; ...; \alpha_r, t; ...; \alpha_s, s)$, meaning that lottery B differs from lottery A only in that B gives lottery r with probability α_r while A gives lottery t with that same probability. Compound independence says that A is preferred to B if and only if r is preferred to t.

$\{0.75, 0.50, 0.25\}$. In aggregate, 25% of all of such paired choices revealed a violation of independence.[17]

Figure III shows the violations of independence in the simplex, separated by mixing probability. We find higher violations of independence as λ decreases.[18] We find individuals violate independence in 22% of decisions when $\lambda = 0.75$, 26% when $\lambda = 0.50$, and 27% when $\lambda = 0.25$ (Wilcoxon ranksum p-values, 0.75 vs. 0.25 $p = 0.007$, 0.75 vs. 0.50 $p = 0.074$, 0.50 vs. 0.25 $p = 0.361$). We also see that violations appear more common for risky lotteries with higher expected value (lotteries to the northwest). These are lotteries where individuals are more likely to have chosen the risky option in the unmixed question. We formalize this observation below, showing that individuals indeed violate independence more when the risky lottery is preferred to certainty.

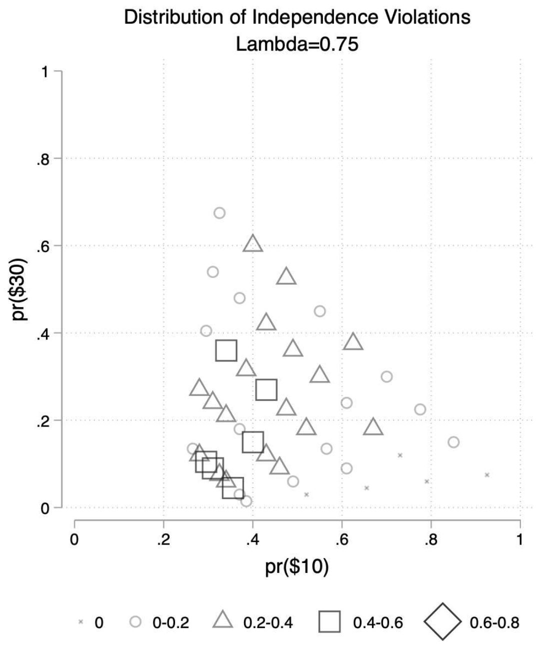
V.A. Certainty Effect vs. Reverse Certainty Effect

We denote a violation of independence as a "reverse certainty effect" (RCE) violation when individuals prefer the risky lottery to δ_{20} in the unmixed question but reverse their preference in the mixed question. We refer to the opposite as a "certainty effect" (CE) violation, when individuals prefer δ_{20} to the risky lottery in the unmixed question but reverse their preference in the mixed question. The first set of bars in Figure IV presents our main results. We find that, conditional on choosing the safe option in the unmixed question, individuals violate independence in 15% of possible opportunities. In contrast, conditional on choosing the risky option in the unmixed question, individuals violate independence in 39% of possible opportunities (15% vs. 39%, Fisher-Pitman permutation test $p < 0.001$). Thus, RCE violations are significantly more common than CE violations.

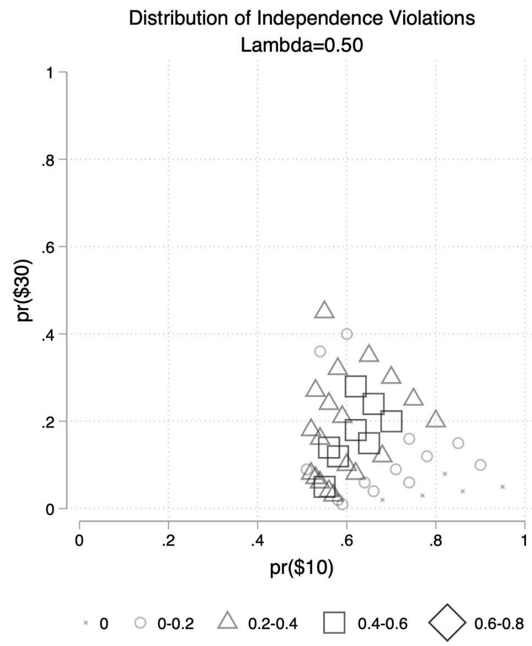
Analyzing the data differently, we can look at all observed violations of independence and ask whether these violations come from situations where individuals chose the risky option in the unmixed question or from situations where individuals chose the safe option in the unmixed question. In other words, given an observed inde-

[17] This percentage likely would change as we change payoffs, unmixed lotteries, etc. Therefore, we do not emphasize the raw percentage of violations, and leave it to the reader to decide whether this is a large number or not.

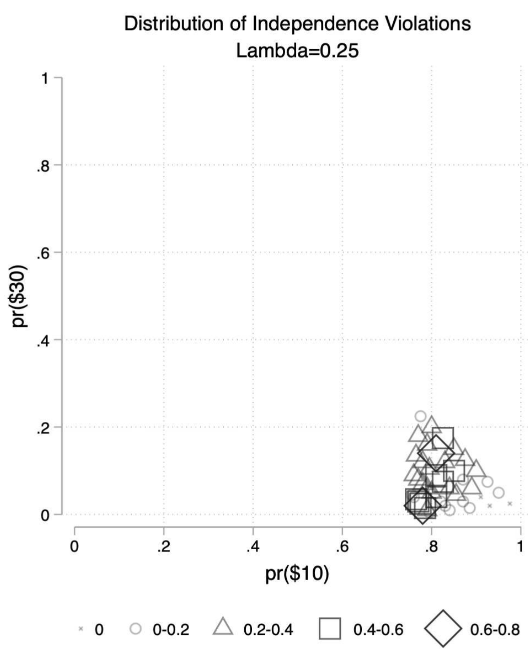
[18] This could be because the lotteries converge as λ decreases, so they become closer to one another in expected value. Decision error could lead to more violations of independence as alternatives become closer together McGranaghan et al., 2023.



(1) $\lambda = 0.75$



(2) $\lambda = 0.50$



(3) $\lambda = 0.25$

Figure III: Independence Violations in the Simplex

Notes: Figures show percentage of independence violations in the Allais Mix questions, compared with δ_{20}. Size and shape of markers denote frequency of violations, with percentages as indicated in the legend.

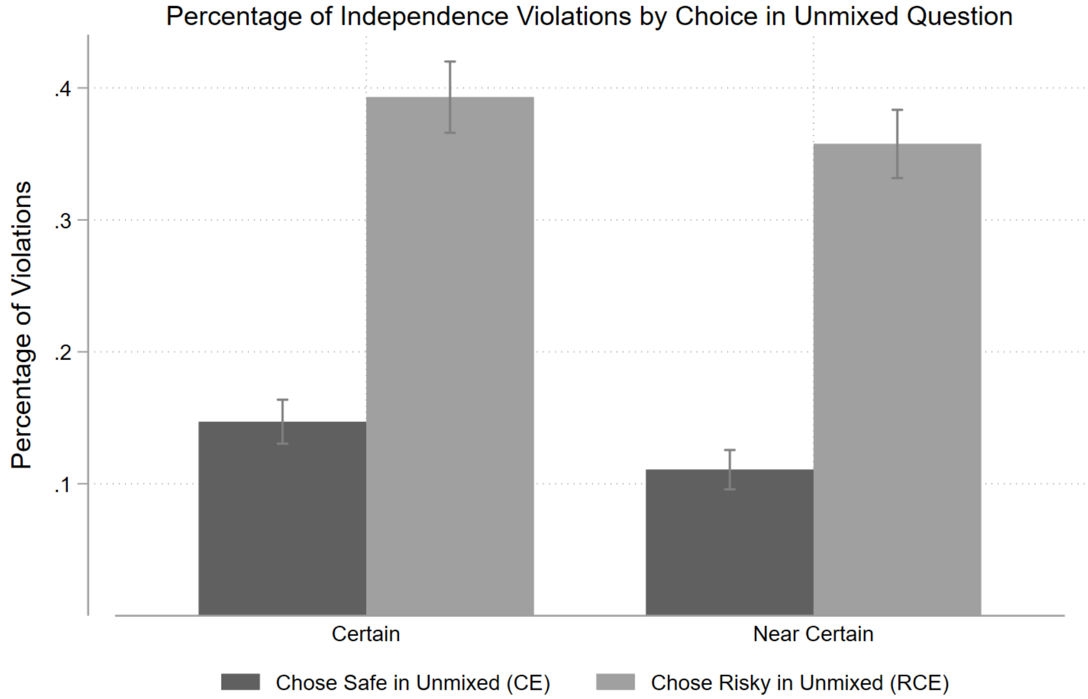


Figure IV: Violations of Independence Consistent with the Certainty Effect and Reverse Certainty Effect

Notes: The left two columns show independence violations in the Certain condition, where subjects choose between δ_{20} and a risky lottery. The right two columns show violations in the Near Certain condition, where subjects choose between $q^* = (\$30, 0.05; \$20, 0.90; \$10, 0.05)$ and a risky lottery. The darker bars show, among all questions where individuals chose the safer option ($\delta_2 0$ or q^*) in the unmixed question, the percentage of comparisons constituting an independence violation. The lighter bars show the same thing among questions where individuals chose the riskier option in the unmixed question.

pendence violation, we can characterize whether it is more likely to be a CE or RCE violation. We find that two thirds of all independence violations in the data are RCE violations, where individuals chose the risky option in the unmixed question (66% vs. 34%, Fisher-Pitman permutation test $p < 0.001$).

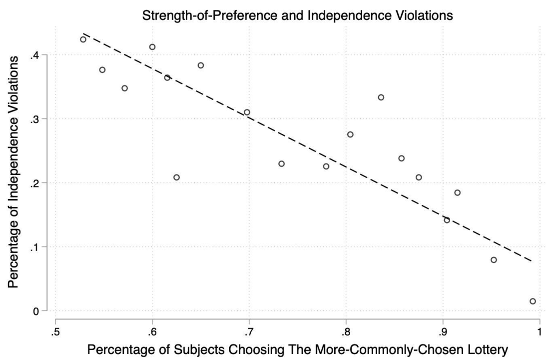
Our main result is surprising in light of a large literature on the certainty effect. We find that independence violations are much more common when individuals prefer risk to certainty in the absence of mixing. This is exactly the opposite of Allais's intuition, which hypothesized that independence violations would be driven by a "preference for security." Instead, two-thirds of our violations result when a risky lottery is preferred to certainty. These are also violations of negative certainty independence, the core axiom in cautious expected utility (Cerreia-Vioglio et al., 2015).

V.B. The Role of Noise

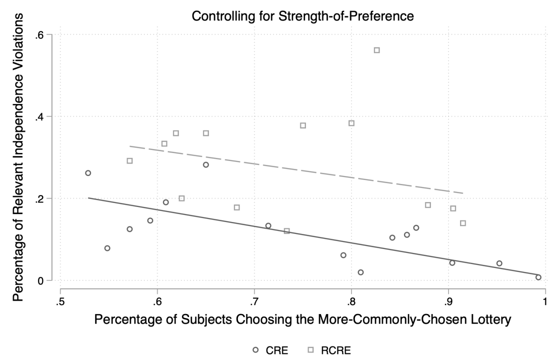
Prior work has demonstrated how noise can generate a common ratio effect, even if underlying noise-free preferences are consistent with expected utility (Ballinger and Wilcox, 1997; Loomes, 2005; Hey, 2005; Wilcox, 2008; Blavatskyy, 2007, 2010; Bhatia and Loomes, 2017; McGranaghan et al., 2023). The argument is that scaling down lotteries by multiplying probabilities by a common ratio brings the lotteries in the decision closer together in expected utility, thus making the impact of noise larger in the scaled down decisions. This can bias toward finding a CRE or RCRE depending on whether individuals are more likely to choose the safe or risky option in the unmixed question. Intuitively, if individuals always choose the safe option in the unmixed question and noise only has an impact in the mixed question, then we can only observe a CRE; the opposite is true if individuals choose the risky option in the unmixed question, leading to observed RCRE. Thus, it is important for us to account for the impact of noise in comparing the prevalence of CRE vs. RCRE. In our design, this is particularly true given that the simplex we consider is "balanced" only for risk-neutral participants. For risk-averse participants, there will be fewer questions where the individual strongly prefers the risky lottery—and independence violations are relatively unlikely—compared to questions where the individual strongly prefers the safe lottery.[19]

We do not have individual-level measures of strength of preference in a given decision. Instead, we proxy for this using the aggregate choice probabilities in the population. Specifically, we create a "strength-of-preference" measure that takes values from 0.5 to 1, defined as the percentage of subjects in the study who chose the more-commonly-chosen alternative in a given unmixed question. As an example, this measure would equal 0.7 both when 70% of individuals chose the safe option in a given unmixed comparison and when 70% of subjects chose the risky option in a given unmixed comparison. We interpret values close to 0.5 as reflecting that participants were, on average, close to indifferent in that decision. When values are close

[19]As noted by Rabin, such small-stakes risk aversion presents challenges to extrapolating risk preferences to higher stakes decisions. Our results are consistent with other experiments, and prior work has proposed explanations such as reference-dependence Bleichrodt et al., 2020 and narrow bracketing Zhang, 2021.



(1) Total Independence Violations



(2) CRE vs. RCRE

Figure V: Strength-of-Preference Effects

Notes: The left panel shows the percentage of total independence violations as a function of our strength-of-preference measure, where strength-of-preference is defined as the percentage of subjects in the study who chose the more-commonly-chosen alternative in a given unmixed question. The right panel shows the percentage of CRE violations for questions where a majority of individuals chose the safe option in the unmixed question and shows the percentage of RCRE violations for questions where a majority of individuals chose the risky option in the unmixed question, both as a function of strength-of-preference in the unmixed question. Data include only comparisons involving certainty in the unmixed question.

to 1, this indicates that participants all preferred either the safe or the risky lottery.

We find that overall violations strongly correlate with strength-of-preference. Individuals are much more likely to violate independence when this value is close to 0.5; we document this in panel (1) of Figure V. To ensure that unbalanced strength-of-preference between risky and safe lotteries does not drive our results, we conduct the following exercise. For unmixed questions where a majority of individuals prefer the safe lottery, we calculate the percentage of CRE violations; for unmixed questions where a majority of individuals prefer the risky lottery, we calculate the percentage of RCRE violations. Noisy expected utility would predict these values to be the same controlling for strength-of-preference.

Panel (2) of Figure V shows that this is not the case. Controlling for strength-of-preference, individuals are more likely to exhibit a RCRE than a CRE ($p < 0.001$ from a probit regressing violations on choice in the unmixed question and the strength-of-preference measure).

V.C. *The Role of Certainty*

The original intuition of the certainty effect claimed that "certainty" held a fundamentally different appeal from "near certainty." We test this by analyzing differences in independence violations when comparisons involve true certainty (δ_{20}) versus near-certain ($q^* = (\$30, 0.05; \$20, 0.90; \$10, 0.05)$) options. We find the overall percentage of independence violations is slightly but significantly higher under certainty (25% vs. 22%, Fisher-Pitman Permutation Test, $p = .00415$).

Nevertheless, the second set of bars in Figure IV shows that our main result holds equally under certainty as near certainty: Individuals are over three times more likely to violate independence in questions where risk is preferred to certainty or near-certainty.

Table I confirms these results in a probit regression. The dependent variable is a dummy taking the value of 1 for a violation of independence, 0 otherwise. Independent variables include an *Unmixed Risky* dummy, taking the value of 1 if the individual chose the risky lottery in the unmixed question, 0 otherwise, a *Certain* dummy taking the value of 1 when the unmixed question compared against δ_{20}, 0 for questions compared against q^*, the interaction between these two variables, and the strength-of-preference measure described above. We cluster standard errors at the subject level.

The results from the regression confirm the conclusions above: Violations of independence are significantly more common when individuals choose the riskier option in the unmixed question.

V.D. *Individual-Level Results*

Our main result holds on an individual level, as well. Figure VI shows the percentage of independence violations per subject, broken down by CE and RCE types. That is, for each individual, we separate the 17 unmixed questions they answered—including both certain and near-certain comparisons—according to whether they preferred the risky or safer lottery. Then, we compute the average independence violations, for each individual, within these two sets. We see that the modal subject never violates independence in questions where they preferred the safer lottery to the risky lottery. When they preferred the risky lottery, however, the modal subject violates indepen-

Dependent Variable: Violation of Independence	
Unmixed Risky	0.838***
	(0.105)
Certain	0.0638
	(0.0799)
Unmixed Risky × Certain	-0.175
	(0.108)
Strength-of-Preference	-2.620
	(0.194)
Constant	878***
	(0.165)
No. Observations	6,018
No. Clusters	118

Table I: Probit regression predicting violations of Independence

Notes: The dependent variable is a dummy taking the value of 1 for a violation of independence, 0 otherwise. *Unmixed Risky* is a dummy taking the value of 1 if the individual chose the risky lottery in the unmixed question, 0 otherwise. The *Certain* dummy takes the value of 1 when the unmixed question compared against δ_{20}, 0 for questions compared against q^*. Strength-of-preference is defined as the percentage of subjects in the study who chose the more-commonly-chosen alternative in a given unmixed question. We cluster standard errors at the subject level.

dence one-third of the time.

In addition, for each subject, we take the difference between their likelihood of RCE and CE violations (RCE - CE). For example, a subject who violated independence in 25% of questions where they preferred the risky lottery over δ_{20} and violated independence in 15% of questions where they preferred δ_{20} over the risky lottery would give a difference of 0.10. Figure VII shows the results. A large majority of subjects (81%) demonstrate a positive difference in violations, meaning that a majority of our subjects express more RCE than CE violations. Therefore, we conclude that RCE violations are more common both in aggregate and on an individual-level.

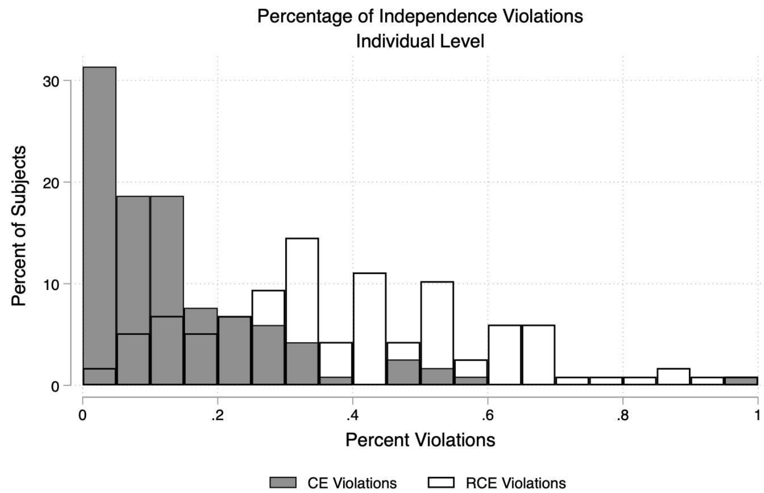


Figure VI: Individual-Level Violations by Unmixed Risky vs. Unmixed Safe
Notes: Histogram shows the percentage of questions in which each individual violates the independence axiom, separated out by CE violations (where the individual preferred δ_{20} to the risky lottery) and RCE violations (where they preferred the risky lottery to δ_{20}).

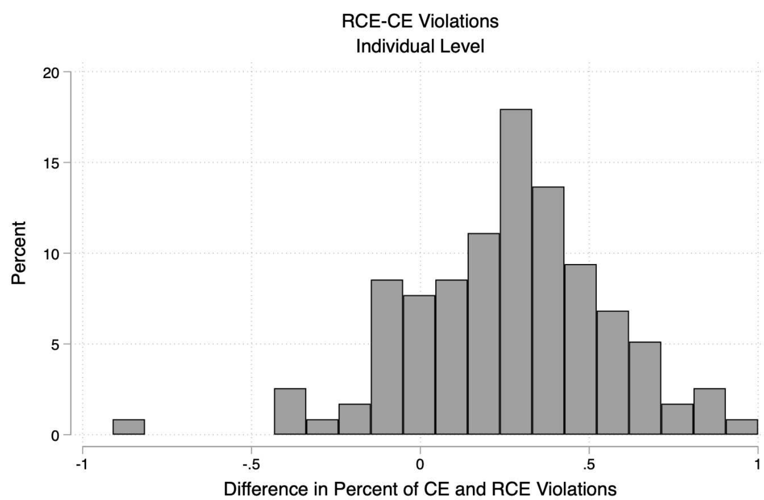


Figure VII: Individual-Level Difference Between Percentage of RCE and CE Violations
Notes: Histogram shows individual-level difference in the percentage of questions in which the individual violates the independence axiom in preferring the risky lottery over δ_{20} minus the percentage of questions in which the individual violates the independence axiom in preferring δ_{20} over the risky lottery.

V.E. A More Rigorous Test

Given that each individual answers four questions all linked by independence—$\lambda = \{1, 0.75, 0.50, 0.75\}$—we can analyze violations of independence using all four choices as a unit of observation. We define an individual's four choices by a string of four letters, each R (riskier) or S (safer). The first letter represents their choice in the unmixed question ($\lambda = 1$), the second in the $\lambda = 0.75$ mixed question, the third in the $\lambda = 0.50$ question, and the last in the $\lambda = 0.25$ question. Thus, the string represents choices as we move towards the bottom right of the simplex.

Pattern	Near-Certain	Certain
SSSS	449	425
RRRR	191	159
RSSS	71	72
RRRS	41	36
RRSS	38	45
RRSR	32	38
SSSR	29	43
RSRR	24	24
RSRS	22	19
SSRS	22	18
RSSR	21	26
SRRR	18	29
SRSS	18	19
SSRR	15	12
SRSR	12	17
SRRS	6	15

Table II: Pattern of Choices Per Question
Notes: We define an individual's four choices by a string of four letters, each R (riskier) or S (safer). The first letter represents their choice in the unmixed question ($\lambda = 1$), the second in the $\lambda = 0.75$ mixed question, the third in the $\lambda = 0.50$ question, and the last in the $\lambda = 0.25$ question.

Independence requires the individual choose either R or S in all four questions. This is a more stringent requirement than in our main analysis, as it requires individuals to be consistent in all three binary comparisons. Nevertheless, these are the two most common patterns we see, first SSSS followed by RRRR—about 60% of our data falls into one of those two patterns.[20] Consistent with the analysis above,

[20]Recall that just looking at binary comparisons, about 75% of choices were consistent with independence. Furthermore, we note that the strength-of-preference argument above again holds here;

however, the next-most-common patterns involve choosing the riskier option in the unmixed question and choosing the safer alternative as λ decreases. In the most common independence violation, RSSS, individuals choose the riskier option only in the unmixed question but then choose the safer option in all mixtures. In the second most common violation, RRSS, individuals choose the riskier option in the unmixed question and $\lambda = 0.75$, but choose the safer option in the $\lambda = 0.50, 0.25$ mixtures. All patterns can be found in Table II. We find no significant difference across certain and uncertain questions (Chi-square $p = 0.351$).

Analyzing the data from all four questions also allows us to test whether violations of independence are less common in the interior of the simplex. Allais hypothesized that "'far from certainty,' individuals act as expected utility maximizers" (Allais, 1953, translated by Andreoni and Sprenger, 2010). If this were the case, we would only see independence violations of the SRRR and RSSS types, where individuals act consistently with independence in the $\lambda = 0.75, 0.50, 0.25$ mixtures but might violate it near certainty. Instead, we find that these patterns make up only 24% of independence violations in the certain comparisons in our sample. Over three quarters of choice patterns violating independence involve violations in questions that lie strictly on the interior of the simplex.[21]

Furthermore, the SRRR pattern most closely associated with the certainty effect is very rare. These are instances in which individuals choose the safe option when it is certain or near-certain, but then reverse their preferences when both alternatives move away from certainty. Conditional on choosing S in the $\lambda = 1$ question, only 17% of violations follow this pattern.

V.F. Robustness to the Mixing Lottery

We test the robustness of our results to the choice of r, the mixing lottery in the definition of the independence axiom: $p \succeq q \Leftrightarrow \lambda p + (1-\lambda)r \succeq \lambda q + (1-\lambda)r$. In our main treatment, we chose r to be (\$10, 1), or 100% chance of the lowest payoff. This is closest in spirit to the original Allais paradox, where the lotteries were mixed with a large chance of receiving \$0. In our robustness sessions, we conducted exactly the same experiment, except we mixed all lotteries instead with the midpoint of the

there is a very strong relationship between strength-of-preference in the unmixed question and being consistent with EU in all four questions (probit $p < 0.001$).

[21] All near-certain violations are also strictly interior.

simplex, ($\$30, \frac{1}{3}; \$20, \frac{1}{3}; \$10, \frac{1}{3}$). Figure IX in the Appendix shows a visualization of these lotteries in the simplex. We chose this point so that mixing would converge to a different area of the simplex, one which has not been studied in detail based on our review of the literature. Other than the choice of r, all procedures in these sessions followed identically to those in the main session.

Overall, individuals violate independence in 21% of questions, which is slightly but significantly lower than in our original treatment (21% vs. 23%, Chi-square $p < 0.001$). Figures XI and XII in the Appendix show the distribution of violations across the simplex.

Again we find that RCE violations are more common than CE violations. Table III confirms that, for both certain and near-certain comparisons, individuals are more likely to violate independence after choosing the risky lottery in the unmixed question ($p < 0.001$). Figure XIII in the Appendix shows results controlling for strength-of-preference, analogous to Figure V above; results again persist.

	Percentage of Independence Violations	
	Certain	Near-Certain
Chose Safer in Unmixed	14%	14%
Chose Risky in Unmixed	37%	21%
p-value	< 0.001	< 0.001

Table III: Percentage of Independence Violations Conditional on Choice in Unmixed Question, by Near-Certain and Certain, in the Middle Mix Treatment

Notes: The first column shows independence violations in the Certain condition, where subjects choose between δ_{20} and a risky lottery. The second column shows violations in the Near Certain condition, where subjects choose between $q^* = (\$30, 0.05; \$20, 0.90; \$10, 0.05)$ and a risky lottery. The first row reports, among all questions where individuals chose the safer option (δ_{20} or q^*) in the unmixed question, the percentage of comparisons constituting an independence violation. The second row the same thing among questions where individuals chose the riskier option in the unmixed question.

V.G. Robustness to Payoffs

Recent evidence has shown that the Allais Paradox may be driven, in part, by an aversion to receiving a \$0 payoff rather than a preference for certainty (Incekara-Hafalir et al., 2020). In our main treatments, subjects did not receive a show-up fee and the lotteries were expressed in terms of {\$10, \$20, \$30} payoffs. To test the sensitivity of our results to this "zero effect," we run the same experiment with a \$10 show-up fee, where lotteries were expressed in terms of {\$0, \$10, \$20} payoffs.[22] We call this new treatment the "Zero Treatment."

Due to the COVID-19 pandemic, we ran these new experiments online. Specifically, we re-programmed the experiment using oTree (Chen et al., 2016). We recruited participants again from the OSU subject pool, but instead of holding a live session, subjects were sent a link to the experiment and could participate anytime during a pre-specified 24 hour window. Subjects were told to complete the experiment all in one sitting. Uncertainty was resolved independently for each subject by computerized randomization, and subjects were paid within 3 days via PayPal.

We recruited a total of 126 new participants to participate in the Zero Treatment. In addition, we ran a small sample to replicate our original treatment using these new protocols. These results can be found in the Appendix. The results from this Baseline replication treatment are weaker than in our original sample, so the Zero Treatment results, reported in Table IV, are stark relative to this. Once again, we find that individuals are more likely to violate independence in questions where they choose the *risky* option in the unmixed question.

	Percentage of Independence Violations	
	Certain	Near-Certain
Chose Safer in Unmixed	19%	15%
Chose Risky in Unmixed	33%	29%
p-value	< 0.001	< 0.001

Table IV: Percentage of Independence Violations Conditional on Choice in Unmixed Question, by Near-Certain and Certain, in the Zero Treatment

Notes: The first column shows independence violations in the Certain condition, where subjects choose between δ_{20} and a risky lottery. The second column shows violations in the Near Certain condition, where subjects choose between $q^* = (\$30, 0.05; \$20, 0.90; \$10, 0.05)$ and a risky lottery. The first row reports, among all questions where individuals chose the safer option (δ_{20} or q^*) in the unmixed question, the percentage of comparisons constituting an independence violation. The second row the same thing among questions where individuals chose the riskier option in the unmixed question.

[22] If subjects integrate the show-up fee into their payoffs, these lotteries are exactly the same as in the main treatments. However, Incekara-Hafalir et al. (2020) also include a show-up fee, so we believe the framing will prevent integration of payments in a way that would bury any zero effect.

VI. DISCUSSION

We study the independence axiom and how violations of independence interact with a preference for certainty. Contrary to a prominent thread in the literature, we find that violations of independence in our data are not predominantly driven by this preference for certainty. Instead, violations are more common when individuals prefer risk to certainty. We find this is also true when we move slightly away from certainty.

Our results are surprising in light of the large literature following up on the original Allais Paradox counter-examples to independence. The certainty effect is well-documented and is one of the primary pieces of evidence motivating new theoretical models. Our paper aims to provide a more structured analysis to document violations of the independence axiom near certainty. Our results suggest caution in attributing violations of independence to the certainty effect primarily, but more evidence is required before making general statements on where and when to expect violations of expected utility.

Related to this, it is still an open question as to which models in the literature best approximate risk preferences. In addition to cautious expected utility theory (Cerreia-Vioglio et al., 2015), models such as prospect theory (Kahneman and Tversky, 1979) and disappointment aversion (Bell, 1985; Gul, 1991) are founded on evidence of the common ratio effect. Thus, our data are incompatible with these classes of models. Given that we observe both common ratio and reverse common ratio violations in different areas of the simplex—and that we observe both types of violations within-subject—suggests need for more flexible models of decision-making.

Our results also highlight the importance of conducting such "systematic tests" of axioms. As noted by Blavatskyy (2010), previous common ratio tests focused on questions with a specific structure—where the sure payment was likely to be selected over the risky lottery—leading to mechanical confirmation of the certainty effect. The fact that these are the minority of independence violations in our data, but comprise the majority of the past literature, shows the value of testing axioms uniformly and agnostically.

We document consistent patterns of behavior, but leave open the question of what drives these preferences. In particular, correlation of independence violations with measures such as IQ, cognitive reflection test (CRT) scores, etc. remain an interesting open question for future research. We also leave open the questions of how these patterns of violations change with payment amounts and other parameters of the decision environment.

References

Maurice Allais. Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'ecole americaine,. *Econometrica*, 21(4):503–546, 1953.

James Andreoni and William T. Harbaugh. Unexpected utility: Experimental tests of five key questions about preferences over risk. *Working Paper*, 2010.

James Andreoni and Charles Sprenger. Certain and uncertain utility: The allais paradox and five decision theory phenomena. *Working Paper*, 2010.

James Andreoni and Charles Sprenger. Uncertainty equivalents: Testing the limits of the independence axiom. *NBER Working Paper*, 2011.

James Andreoni and Charles Sprenger. Risk preferences are not time preferences. *American Economic Review*, 102(7):3357–3376, 2012.

Yaron Azrieli, Christopher P Chambers, and Paul J Healy. Incentives in experiments with objective lotteries. *Experimental Economics*, pages 1–29, 2019.

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

David E. Bell. Disappointment in decision making under uncertainty. *Operations Research*, 33(1):1–27, 1985.

Sudeep Bhatia and Graham Loomes. Noisy preferences in risky choice: A cautionary note. *Psychological Review*, 124(5):678–687, 2017.

Pavlo Blavatskyy, Valentyn Panchenko, and Andreas Ortmann. How common is the common-ratio effect? *Experimental Economics*, 2022.

Pavlo R. Blavatskyy. Stochastic expected utility theory. *Journal of Risk and Uncertainty*, 34(3):259–286, 2007.

Pavlo R Blavatskyy. Reverse common ratio effect. *Journal of Risk and Uncertainty*, 40(3):219–241, 2010.

Pavlo R Blavatskyy. The reverse allais paradox. *Economics Letters*, 119(1):60–64, 2013.

Han Bleichrodt, Jasono N. Doctor, Yu Gao, Chen Li, Daniella Meeker, and Peter P. Wakker. Resolving rabin's paradox. *Journal of Risk and Uncertainty*, 59:239—-260, 2020.

Alexander L Brown and Paul J Healy. Separated decisions. *European Economic Review*, 101:20–34, 2018.

Colin Camerer. Individual decision making. *Handbook of experimental economics*, 1995.

Colin F. Camerer. An experimental test of several generalized utility theories. *Journal of Risk and uncertainty*, 2(1):61–104, 1989.

Colin F. Camerer and Teck-Hua Ho. Violations of the betweenness axiom and nonlinearity in probability. *Journal of Risk and uncertainty*, 8:167–196, 1994.

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

Simone Cerreia-Vioglio, David Dillenberger, and Pietro Ortoleva. An explicit representation for disappointment aversion and other betweenness preferences. *Forthcoming, Theoretical Economics*, 2020.

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

John Conlisk. Three variants on the allais example. *The American Economic Review*, pages 392–407, 1989.

Robin P. Cubitt, Chris Starmer, and Robert Sugden. On the validity of random lottery incentive systems. *Expeirmental Economics*, 1(2):115–131, 19989.

David Dillenberger. Preferences for one-shot resolution of uncertainty and allais-type behavior. *Econometrica*, 78(6):1973–2004, 2010.

Urs Fischbacher. z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2), 2007.

B Greiner. An online recruitment system for economics experiments. 2004.

Faruk Gul. A theory of disappointment aversion. *Econometrica*, 59(3), 1991.

Yoram Halevy. Strotz meets allais: Diminishing impatience and the certainty effect. *American Economic Review*, 98(3):1145–1162, 2008.

David W Harless. Predictions about indifference curves inside the unit triangle: A test of variants of expected utility theory. *Journal of Economic Behavior & Organization*, 18(3):391–414, 1992.

John D. Hey. Why we should not be silent about noise. *Experimental Economics*, 8 (4):325–345, 2005.

Steffen Huck and Wieland Müller. Allais for all: Revisiting the paradox in a large representative sample. *Journal of Risk and Uncertainty*, 44(3):261–293, 2012.

Steven Humphrey and Arjan Verschoor. The probability weighting function: Experimental evidence from uganda, india, and ethiopia. *Economics Letters*, 84:419–425, 2004.

Elif Incekara-Hafalir, Eungsik Kim, and Jack D. Stecher. Is the allais paradox due to appeal of certainty or aversion to zero? *Experimental Economics*, 2020.

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

Graham Loomes. Modelling the stochastic component of behaviour in experiments: Some issues for the interpretation of data. *Experimental Economics*, 8(4):301–323, 2005.

Graham Loomes and Robert Sugden. Testing different stochastic specifications of risky choice. *Economica*, 65:581–598, 1998.

Christina McGranaghan, Kirby Nielsen, Ted O'Donoghue, Jason Somerville, and Charles D. Sprenger. Distinguishing common ratio preferences from common ratio effects using paired valuation tasks. *American Economic Review*, 2023.

William S Neilson. A mixed fan hypothesis and its implications for behavior toward risk. *Journal of Economic Behavior & Organization*, 19(2):197–211, 1992a.

William S Neilson. Some mixed results on boundary effects. *Economics Letters*, 39 (3):275–278, 1992b.

Kirby Nielsen. Preferences for the resolution of uncertainty and the timing of information. *Journal of Economic Theory*, 189, 2020.

John Quiggin. A theory of anticipated utility. *Journal of Economic Behavior & Organization*, 3(4):323–343, 1982.

Matthew Rabin. Risk aversion and expected-utility theory: A calibration theorem. *Econometrica*, 68:1281–1292, year=2000.

Ulrich Schmidt. A measurement of the certainty effect. *Journal of Mathematical Psychology*, 42:32–47, 1998.

Uzi Segal. Two-stage lotteries without the reduction axiom. *Econometrica*, 58:349–377, 1990.

Barry Sopher and Gary Gigliotti. A test of generalized expected utility theory. *Theory and Decision*, 35(1):75–106, 1993.

Chris Starmer. Testing new theories of choice under uncertainty using the common consequence effect. *The Review of Economic Studies*, 59(4):813–830, 1992.

Chris Starmer. Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of economic literature*, 38(2):332–382, 2000.

Chris Starmer and Robert Sugden. Does the random-lottery incentive system elicit true preferences? an experimental investigation. *American Economic Review*, 81 (4):971–978, 1991.

Amos Tversky and Daniel Kahneman. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, 5(4):297–323, 1992.

Nathaniel T. Wilcox. Stochastic models for binary discrete choice under risk: A critical primer and econometric comparison. In James C. Cox and Glenn W. Harrison, editors, *Risk Aversion in Experiments (Research in Experimental Economics, Vol. 12)*, pages 197–292. Emerald Group Publishing Limited, 2008.

Mu Zhang. A theory of choice bracketing under risk. *Working Paper*, 2021.

A. Additional Results

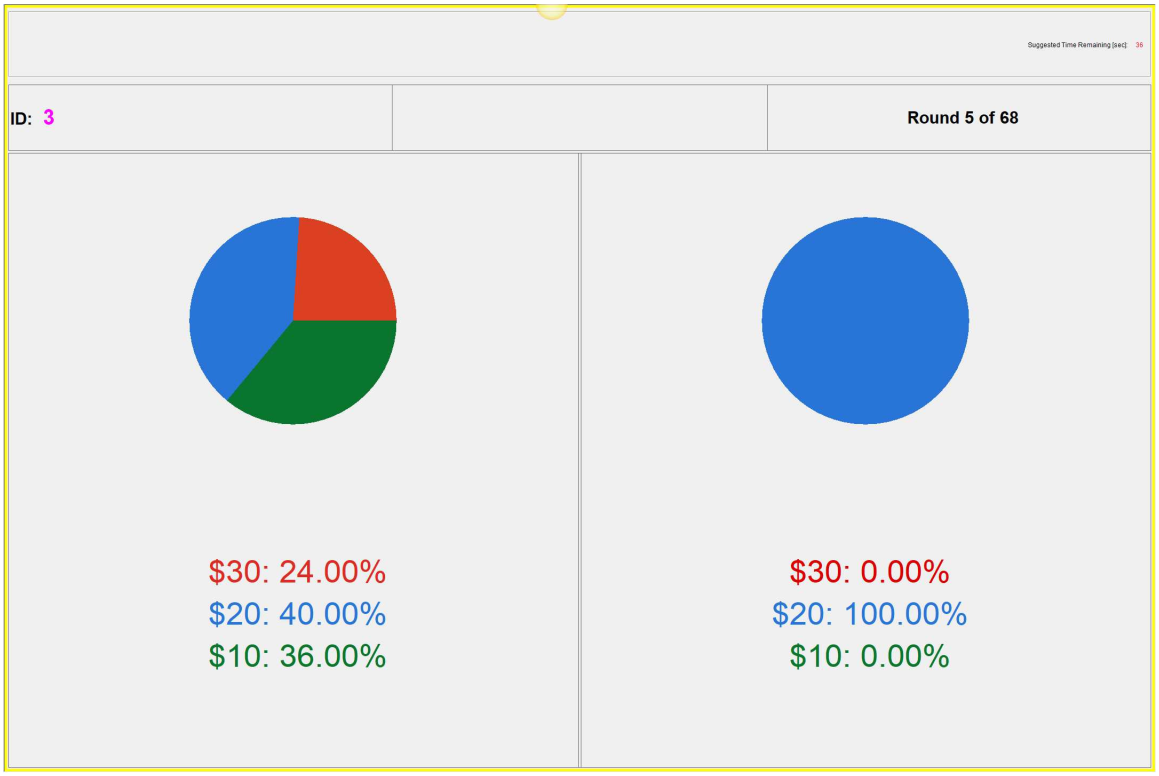


Figure VIII: Screenshot of subjects' display during the experiment

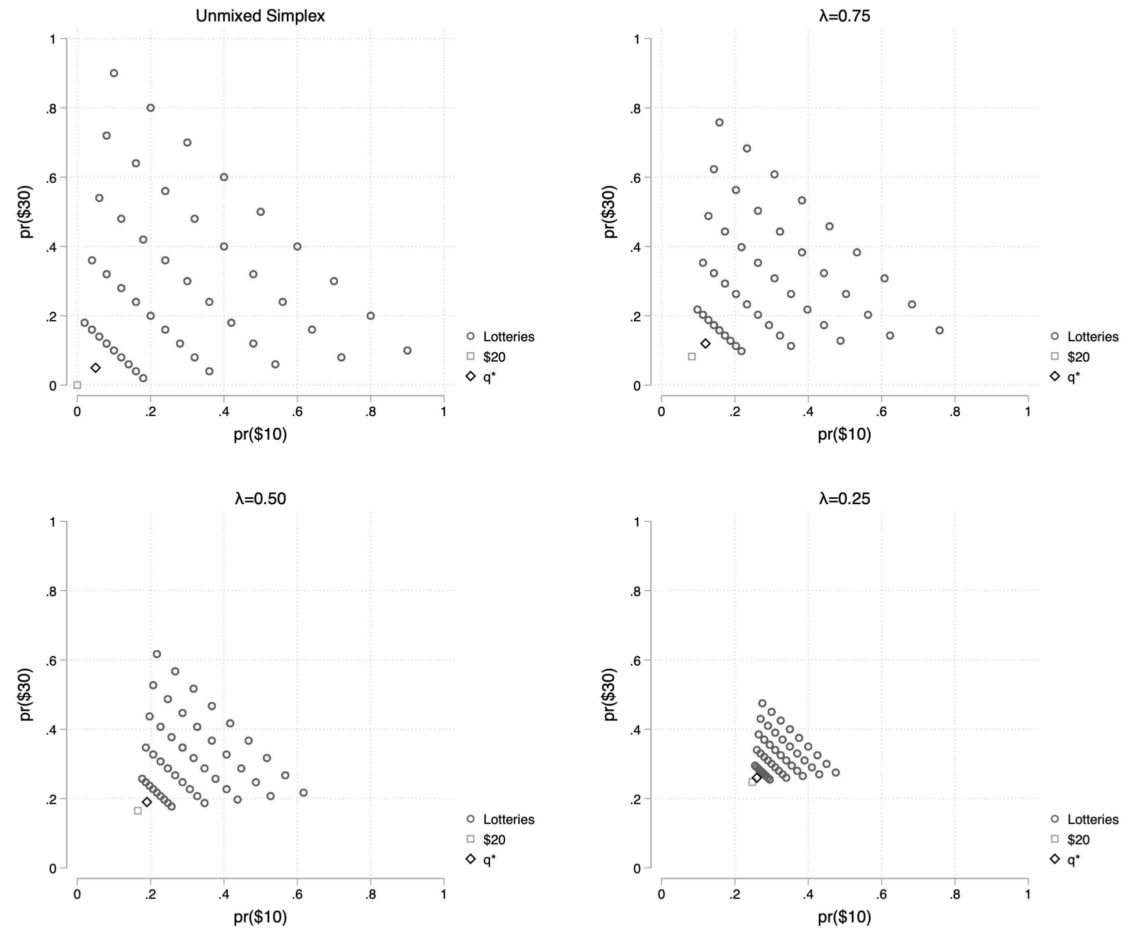


Figure IX: Questions in the Middle Mixture Treatment

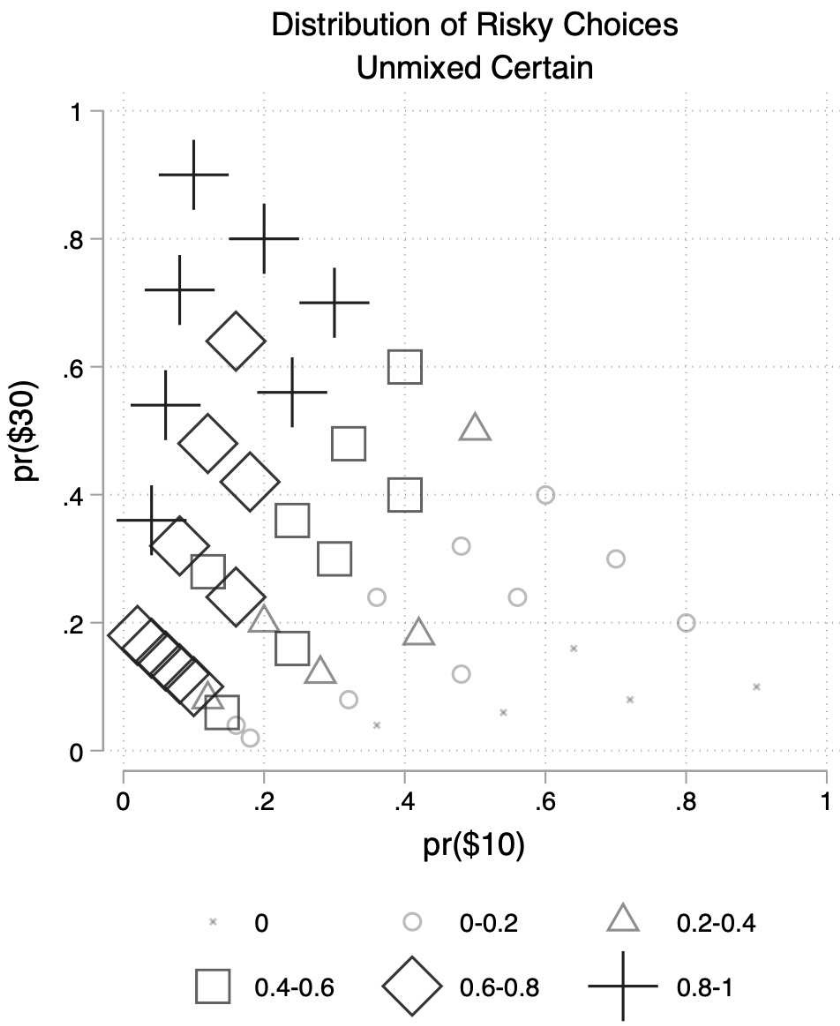
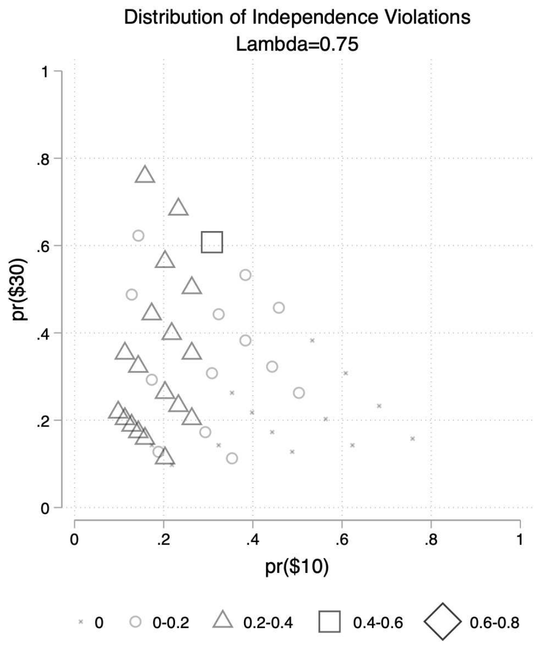
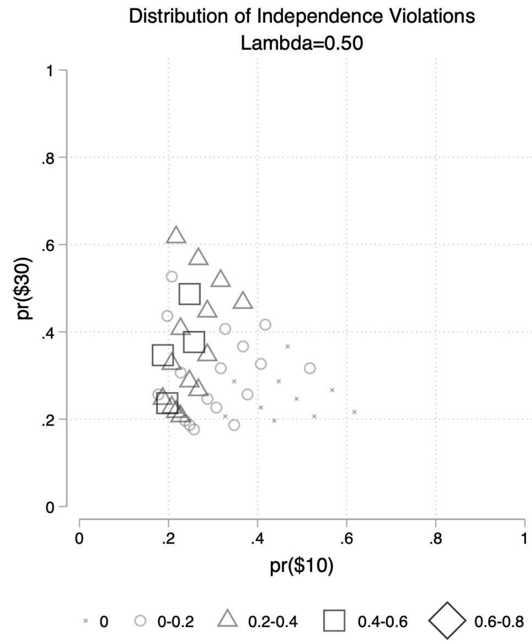


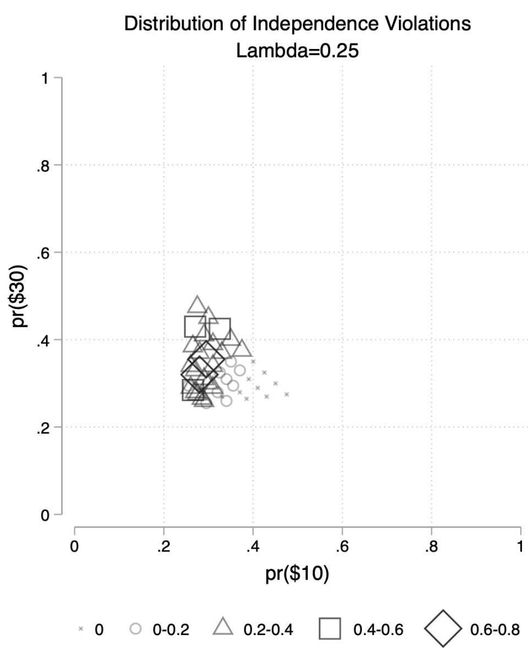
Figure X: Distribution of Risky Choices in the Unmixed Questions



(1) $\lambda = 0.75$



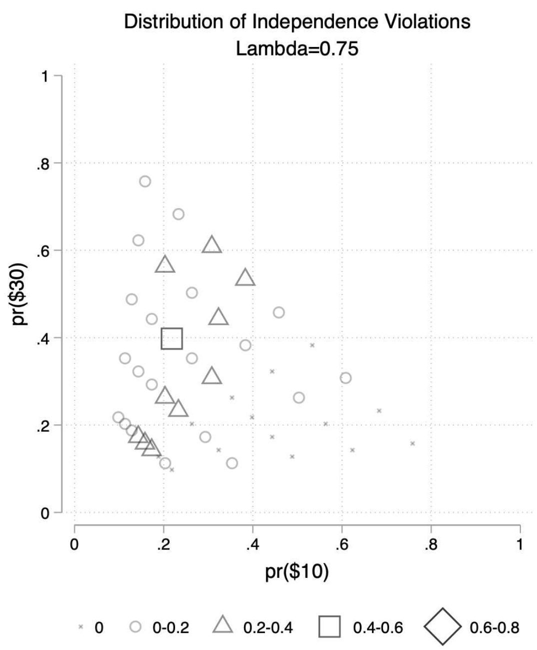
(2) $\lambda = 0.50$



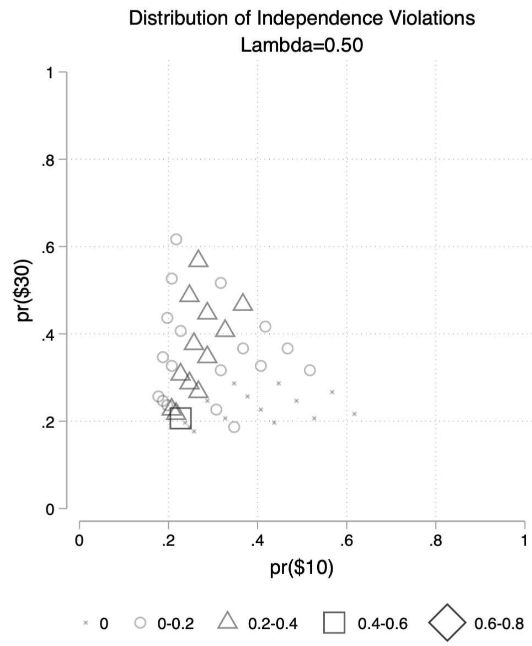
(3) $\lambda = 0.25$

Figure XI: Independence Violations in the Simplex

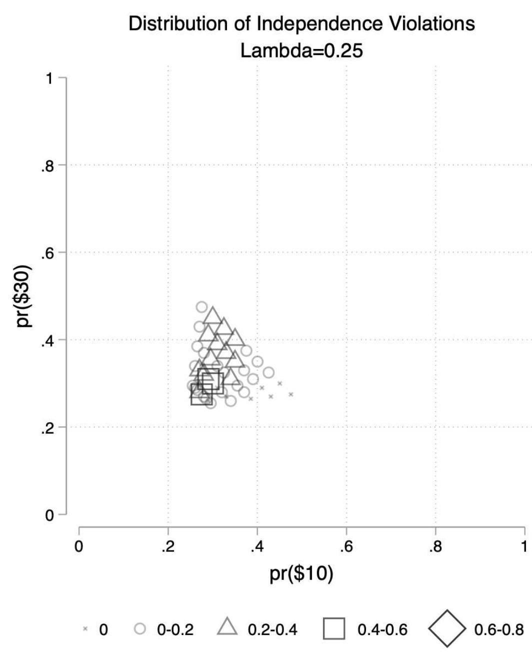
<u>Notes:</u> Figures show percentage of independence violations in the Middle mix questions, compared with certainty. Size of bubbles denote frequency of violations.



(1) $\lambda = 0.75$



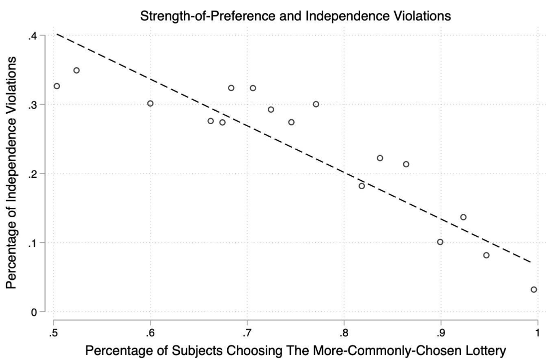
(2) $\lambda = 0.50$



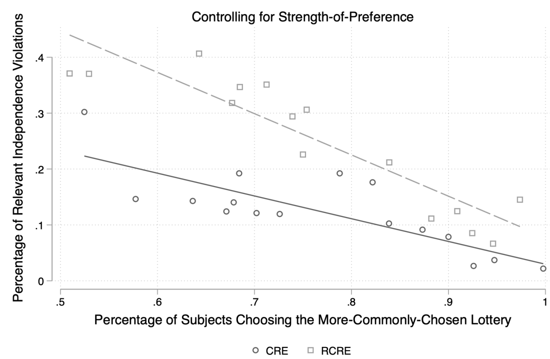
(3) $\lambda = 0.25$

Figure XII: Independence Violations in the Simplex

Notes: Figures show percentage of independence violations in the Middle mix questions, compared with q∗. Size of bubbles denote frequency of violations.



(1) Total Independence Violations



(2) CRE vs. RCRE

Figure XIII: Strength-of-Preference Effects in the Middle Mix Treatment

Notes: The left panel shows the percentage of total independence violations as a function of our strength-of-preference measure, where strength-of-preference is defined as the percentage of subjects in the study who chose the more-commonly-chosen alternative in a given unmixed question. The right panel shows the percentage of CRE violations for questions where a majority of individuals chose the safe option in the unmixed question and shows the percentage of RCRE violations for questions where a majority of individuals chose the risky option in the unmixed question, both as a function of strength-of-preference in the unmixed question. Data include both certain and near-certain comparisons

A.A. Online Replication

Given that the COVID-19 pandemic forced our zero treatment robustness test online, we run our original treatment in this online environment to replicate our main results. We recruited 95 new subjects to take part in the experiment. The online protocols were the same as in the zero treatment, but we used our original {\$10, \$20, \$20} payoffs with no additional show-up fee.

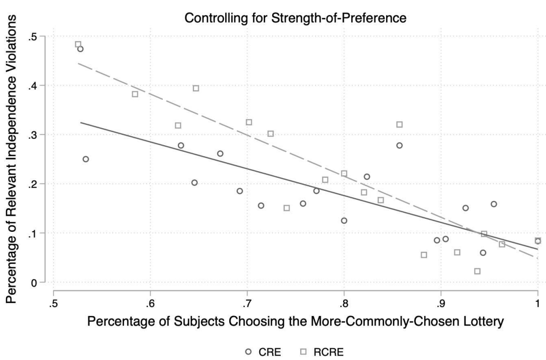
	Percentage of Independence Violations	
	Near-Certain	Certain
Chose Risky in Unmixed	26%	24%
Chose Safer in Unmixed	19%	29%
p-value	< 0.001	< 0.0067

Table V: Percentage of Independence Violations Conditional on Choice in Unmixed Question, by Near-Certain and Certain

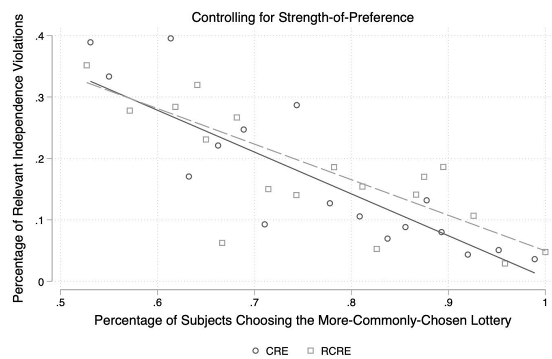
Note: Results are separated by choice in the unmixed question.

In the online replication, we do see significantly more violations of independence conditional on choosing the safer option in the unmixed question for Certain comparisons.

While we do find more CE violations in the online sample, we still find substantial RCE violations. Overall, aggregating certain and near-certain comparisons, among those who chose the riskier alternative in the unmixed comparison, 25% of questions violate independence, and 24% violate independence among those who chose the safer alternative ($p = 0.313$). Further, as Figure XIV shows, we find the same patterns when controlling for strength-of-preference effects, but the difference between CRE and RCRE violations is less dramatic, potentially due to more noisy decision-making online. Thus, while not as stark of a difference as in our other treatments, RCE violations cannot be ignored.



(1) Online $10 Replication



(2) Online $0 Treatment

Figure XIV: Strength-of-Preference Effects in the Online Studies

Notes: The left panel shows the percentage of total independence violations as a function of our strength-of-preference measure, where strength-of-preference is defined as the percentage of subjects in the study who chose the more-commonly-chosen alternative in a given unmixed question. The right panel shows the percentage of CRE violations for questions where a majority of individuals chose the safe option in the unmixed question and shows the percentage of RCRE violations for questions where a majority of individuals chose the risky option in the unmixed question, both as a function of strength-of-preference in the unmixed question. Data include both certain and near-certain comparisons

INSTRUCTIONS

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

LOTTERIES

In this experiment, you will be making choices between “lotteries.” A lottery specifies the chance of receiving certain payoffs. In this experiment, the possible payoffs will be $30, $20, or $10. The chance of each payoff can be anything from 0% to 100%.

For example, one lottery could give you an 80% chance of $30, 10% chance of $20, and a 10% chance of $10. Another lottery might give a 0% chance of $30, a 100% chance of $20, and a 0% chance of $10. There are many different possible lotteries.

In each decision, you will see two lotteries on your screen—one on the left, and one on the right. Your screen will display the written values for the payoffs and probabilities. On top of this, you will see a “pie chart” that shows you the probabilities of each payoff. This pie chart is to help you visualize the various probabilities, but it represents the same exact probabilities and payoffs as what’s written on your screen. The chance of $30 will always be in orange color, the chance of $20 will be in blue, and the chance of $10 will be in green.

Your task is simply to choose the lottery you prefer, either the one on the left or the one on the right. You make your choice by clicking on the pie chart. The computer will record your choice and then will present you again with two lotteries, and so on.

PAYMENT

In this experiment, you will be making 68 decisions, each choosing your more preferred lottery. Each decision will be presented on a different screen. At the end of the experiment, you will be paid for exactly ONE of these decisions. We will roll dice to generate a number 1--68 This will determine the decision that we will pay you for. Note, this means that each of your choices is equally likely to be paid, and one of them actually will be paid, so you should make each decision as if it will determine your payment.

We will pay you the lottery that you chose in the one randomly selected decision. To do this, we will roll dice to generate a number 1—100. If this number is less than the probability of $10, you will receive $10. If the number is greater than the probability of $10 but less than the probability of $10 + the probability of $20, you will receive $20. If it’s greater than the probability of $10 + the probability of $20, you will receive $30.

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

$10 with 30% chance

$20 with 50% chance

$30 with 20% chance

If we roll a number 1—30, you'd receive $10. There are 30 out of 100 possible numbers between 1 and 30, so this corresponds to a 30% chance of $10.

If we roll a number 31—80, you'd receive $20. There are 50 out of 100 possible numbers between 31 and 80, so this corresponds to a 50% chance of $20.

If we roll a number 81—100, you'd receive $30. There are 20 out of 100 possible numbers between 81 and 100, so this corresponds to a 20% chance of $30.