Title

Contingent Thinking and the Sure-Thing Principle: Revisiting Classic Anomalies in the Laboratory#

Permalink

https://escholarship.org/uc/item/32j4d5z2

Authors

Esponda, Ignacio
Vespa, Emanuel

Publication Date

2023-11-10

DOI

10.1093/restud/rdad102

Contingent Thinking and the Sure-Thing Principle: Revisiting Classic Anomalies in the Laboratory[*]

Ignacio Esponda (UCSB)

Emanuel Vespa (UCSD)

April 24, 2023

Abstract

We present an experimental framework to study the extent to which failures of contingent thinking explain classic anomalies in a broad class of environments, including overbidding in auctions and the Ellsberg paradox. We study environments in which the subject's choices affect payoffs only in some states, but not in others. We find that anomalies are in large part driven by incongruences between choices in the standard presentation of each problem and a 'contingent' presentation, which focuses the subject on the set of states where her actions matter. Additional evidence suggests that this phenomenon is in large part driven by people's failure to put themselves in states that have not yet happened even though they are made aware that their actions only matter in those states.

---

[*] We thank Eduardo Azevedo, Ted Bergstrom, Gary Charness, Benjamin Enke, Larry Epstein, Erik Eyster, Drew Fudenberg, Xavier Gabaix, Itzhak Gilboa, Daniel Gottlieb, Yoram Halevy, Paul J. Healy, Shachar Kariv, Shengwu Li, Bart Lipman, Mark Machina, Paulo Natenzon, Muriel Niederle, Ryan Oprea, Pietro Ortoleva, Matthew Rabin, Collin Raymond, Ariel Rubinstein, Marciano Siniscalchi, Joel Sobel, Charlie Sprenger, Tomasz Strzalecki, Severine Toussaert, Georg Weizsäcker, Leeat Yariv, and Sevgi Yuksel for helpful comments. Esponda: Department of Economics, UC Santa Barbara, 2127 North Hall, Santa Barbara, CA 93106, iesponda@ucsb.edu.; Vespa: Department of Economics, UC San Diego, 9500 Gilman Dr. #0508, La Jolla, CA 92093, evespa@ucsd.edu.

1 Introduction

In many economic environments, people need to anticipate the consequences of their actions under different contingencies. A recent economics literature attributes several mistakes to failures in contingent thinking.[1] Some of this literature makes progress in a piecemeal fashion, studying one environment at a time. In this paper, we propose a common experimental framework to study the extent to which several classic anomalies can be attributed to failures of contingent thinking. By focusing on a specific (and relevant) type of contingent thinking that can be formally defined and tested in the laboratory, we are able to connect environments that have been the focus of the contingent-thinking literature (e.g., auctions) to environments that have not (e.g., the Ellsberg paradox).

To illustrate the main ideas and results, consider one of Ellsberg's (1961) classic experiments. There is an urn with 90 balls, 30 of which are red (R) and 60 of which are yellow (Y) or blue (B). One ball is selected from the urn and its color is used to determine the payoff. In Question 1 (Q1), the subject must choose between two options, f (win \$10 if R or B, \$0 otherwise) and g (win \$10 if Y or B, \$0 otherwise). The options and payoffs are depicted in the top portion of Figure 1. In this question, the payoff if the ball is blue is the same regardless of the choice of the subject (\$10). Therefore, when facing this question, it is natural to ask oneself what choice to make contingent on the ball not being blue. Let's say that in Q1 you would prefer f over g in the hypothetical case that the ball were not blue (i.e., it is red or yellow). Then, a natural principle is that you should prefer f over g in Q1, since the payoff in case the ball is not blue is the same regardless of your action.

This principle is an instance of what Savage (1972) called the sure-thing principle (STP), and he stated that "...except possibly for the assumption of simple ordering, I know of no other extralogical principle governing decisions that finds such ready acceptance." Motivated by STP, Savage introduced his central postulate P2 (henceforth, we refer to Savage's P2 as Separability, or simply SEP). When defining SEP, Savage relied on STP and added *one more restriction*: that the choice contingent on the ball not being blue in Q1 should not depend on the specific payoff obtained if the ball is blue, that is, the case where f and g pay the same amount. These two restrictions (STP

---

[1] For recent experiments with a focus on contingent reasoning, see Esponda and Vespa (2014), Louis (2015), Dal Bó, Dal Bó, and Eyster (2018), Li (2017), Martínez-Marquina, Niederle, and Vespa (2019), Ngangoué and Weizsäcker (2021), Bayona, Brandts, and Vives (2020), Martin and Munoz-Rodriguez (2019), Moser (2019), Ali, Mihm, Siga, and Tergiman (2021) and Calford and Cason (2021). Failures in contingent reasoning may create difficulties with understanding correlations (Eyster and Weizsäcker (2010; 2016), Enke and Zimmermann (2019) and Rees-Jones, Shorrer, and Tergiman (2023)) and/or selection effects (Esponda and Vespa (2018), Araujo, Wang, and Wilson (2021), Barron, Huck, and Jehiel (2019) and Enke (2020)).

		A		A^c
		R	Y	B
Question 1	f	$10	$0	$10
	g	$0	$10	$10
Question 2	f'	$10	$0	$0
	g'	$0	$10	$0

Figure 1: Example - Ellsberg problem (ELLS)

Notes: There are 30 red (R) balls and 60 balls each of which can be yellow (Y) or blue (B). In each question, a ball is selected (with replacement). In both questions, the subject's choice only matters if the ball is R or Y.

and the additional preference restriction) are embodied by Savage's SEP postulate, which in this problem implies that changing the payoff if the ball is blue from $10 to something else should not affect the (noncontingent) choice.[2] For this reason, Ellsberg's experiment uses a second question, where the payoff if the ball is blue changes from $10 to $0. Specifically, in Question 2 (Q2), the subject is asked to choose between two other options, f' (win $10 if R, $0 otherwise) and g' (win $10 if Y, $0 otherwise). In this example, the restriction imposed by SEP rules out preference reversals in Q1 and Q2, since in both questions the choice is between $f = f'$ and $g = g'$ in case the ball is not blue, and a payoff that is the same regardless of the choice and differs by question ($10 in Q1 and $0 in Q2).

SEP has been shown to be violated in several experiments, in line with Ellsberg's thought experiment, where the typical preference reversal is to prefer g over f in Q1 but f' over g' in Q2. This violation of SEP is overwhelmingly attributed to non-standard preferences in the literature (e.g., ambiguity aversion).[3] But psychologists have argued that hypothetical thinking is challenging (e.g., Nickerson, 2015), and people may have difficulty putting themselves in situations that have not occurred (such as "the ball is not blue" in the Ellsberg problem). So a violation of SEP may result not from an explicit preference to respond to the common consequence in case the ball is blue, but rather from a failure of contingent thinking. In other words, it is possible that SEP is violated because the sure-thing principle fails, that is, a person prefers f to g contingent on the ball being blue, but then reports that they prefer g to f in the overall problem.

We design an experiment to test the extent to which violations of classical (subjective) expected utility postulates, such as SEP and Dominance (DOM, discussed in more detail in Section 2), arise due to failures in contingent thinking. The results indicate that if the problem is described in a con-

---

[2] The sure-thing principle is so embedded into the notion of Savage's P2 that scholars often refer to P2 itself as the sure-thing principle (e.g., Slovic and Tversky (1974)), when in fact P2 (or SEP in our terminology) includes an additional restriction on preferences, as explained above.

[3] These choices are consistent with an aversion to ambiguity, because in each case the subject chooses the alternative with the known probability of success (2/3 in the first question and 1/3 in the second question).

tingent form (to be explained below), then the reversal leading to a violation of SEP significantly decreases in the Ellsberg problem. Moreover, we show that a similar failure of contingent thinking arises in a broad set of problems where the literature has documented violations of the classical postulates. These problems all have the feature that there is an event, which we denote by A, where the choice matters, and an event, A^c, where the payoff is the same regardless of the choice. The examples we study include the above Ellsberg problem and two versions (common-consequence and common-ratio) of the Allais paradox (where in case a ball of a certain color is drawn, the payoffs are the same regardless of one's choice), auctions (where a bidder's decision to bid above one's valuation is irrelevant if the maximum of others' bids is below the bidder's valuation), and elections (where one's vote in a committee only matters in a pivotal event).

Our strategy is to elicit subjects' choices under two frames for each of the five problems we study. The first frame (*noncontingent*) replicates existing experimental anomalies by presenting each problem in the standard way it appears in the literature. The second frame (*contingent)* frames the problem in the contingent manner that Savage envisioned when motivating his central postulate. In particular, the contingent frame asks the subject to make a choice contingent on event A being true. The subject is also told that, if event A^c is true, then she will receive the specific amount of money corresponding to that event (recall this amount is the same, regardless of her choice). And, if event A is true, the subject will be paid according to her contingent choice. Crucially, in both treatments the subject must make the choice *before* any uncertainty about the state is resolved, and so the difference between treatments is just a framing effect. This is crucial because in the discussion above, when the agent determines her preference conditional on the hypothetical case where event A is true, she does not yet know if event A is actually true.

We can assess failures of STP in each problem by comparing responses between the noncontingent and contingent frames, since a subject who gives different responses in the two frames violates STP. Moreover, we can assess the extent to which anomalies (i.e., failures of SEP or DOM) persist in the contingent treatment. The failure to replicate the anomalies in the contingent treatment would cast doubt on the claim that anomalies are mainly driven by a preference to respond to the common consequence in event A^c.

We first successfully replicate existing anomalies in all the problems that we study, which involve 1,164 subjects who participated in our experimental sessions. We then document that a large part of the anomalies are driven by the failure of STP. In particular, we report violations of STP in *all* problems, ranging from 20% to 70%. Additionally, in most environments we document large reductions of failures of Savage's P2 (SEP) and Dominance in the contingent frame. However, we also document heterogeneity in subjects' response to the treatment. While most subjects go from being inconsistent with standard postulates in the noncontingent frame to being consistent in the contingent frame, several subjects are converted in the opposite direction, particularly for SEP. The net effect of the treatment is that failures of Dominance and SEP drop by half in all problems, except in common-consequence Allais, where conversions in the opposite direction essentially cancel out.

Our results are robust to several variations of each problem as well as to between- and within-subjects designs. Overall, a clear picture emerges that, in a broad set of problems, one should be cautious about drawing inferences about people's preferences from anomalies. Subjects in the noncontingent frame have all the information to reconstruct the contingencies and behave as they do in the contingent frame, but the fact that they don't indicates that non-standard behavior is largely driven by the manner in which subjects mentally represent the problem.

We conclude by examining the possible mechanisms driving the treatment effect between the contingent and noncontingent frames. Broadly speaking, the contingent frame can operate through two channels. First, it highlights the payoff structure of the problem. Specifically, it highlights that because the payoffs are the same in event A^c, the decision only matters in event A. Moreover, it computes the payoffs in event A^c, which seems trivial in some problems (e.g., Ellsberg) but not others (e.g., election). Second, the contingent treatment asks subjects to put themselves in the hypothetical position in which event A is true and make a choice. This type of hypothetical thinking may be challenging, and it is plausible that even if people know that their choice only matters in A, they may still not mentally place themselves in a world in which A is hypothetically true.

To examine the validity of these two channels, we conduct a third frame for two of our representative problems, the Ellsberg and the election problems. In this new treatment, we turn off the second feature described above, and simply highlight the payoff structure and the fact that the choice can only affect payoffs in event A. We find a not statistically significant decrease in anomalies in this new treatment relative to the noncontingent frame. This suggests that the differences in anomalies between the noncontingent and contingent frame are in large part driven by people's failure to put themselves in situations that have not yet happened.[4]

These findings imply that several classic anomalies are in large part driven by failure of contingent thinking. Our approach of unpacking failures of standard postulates into failures of STP or genuine failures of standard postulates can provide a new fruitful perspective on the modeling

---

[4] In the conclusion, we argue that this failure to put oneself in an event that has not happened is likely to be due to an inability to do so more than a preference for not doing so, though we acknowledge that the question of whether what looks like a mistake could be represented as a constrained form of optimization (for example, because it is costly to do hypothetical thinking) is not completely settled.

of anomalies. The standard approach in the literature to deal with anomalies has been to relax Savage's axioms to allow for nonstandard preferences, but these extensions maintain the implicit assumption that subjects can run through the state space and satisfy STP. Our work, however, shows that a larger issue is not so much what kind of preferences people have over final payoffs, but rather the way in which they mentally represent the problem. Finally, our finding that people have difficulty with hypothetical thinking in such a wide range of settings suggests that this is a pervasive issue and likely to manifest in several other economic contexts.

Connections to the literature

The paradoxes of Allais (1953) and Ellsberg (1961) inspired a great amount of experimental and theoretical literature testing for and relaxing the classical postulates, with a particular emphasis on SEP.[5] While SEP relies on the more basic STP, the specific role of STP that we highlight in this paper in connection to SEP has largely been overlooked. Evidence from psychology on the difficulty of contingent thinking, however, suggests that satisfying STP may not be trivial.[6] In economics, some early contributions showing the importance of contingent thinking include Shafir and Tversky (1992) and Croson (1999) in the context of the prisoner's dilemma, and Charness and Levin (2009) in environments with adverse selection. Yet a recent literature has expanded the scope of these ideas to several environments, including elections (Esponda and Vespa, 2014), auctions (Li (2017), Moser (2019), Martínez-Marquina, Niederle, and Vespa (2019), Martin and Munoz-Rodriguez (2019)), markets (Bayona, Brandts, and Vives (2020), Ngangoué and Weizsäcker (2021)), and environments where selection plays a central role (Esponda and Vespa (2018), Barron, Huck, and Jehiel (2019), Enke (2020), Ali, Mihm, Siga, and Tergiman (2021), Araujo, Wang, and Wilson (2021)).[7]

Evidence of failures of contingent thinking have also motivated a theoretical literature. Li (2017) defines the notion of an obviously dominant strategy as one that a cognitively challenged

---

[5] A limited sample of the large experimental literature includes MacCrimmon and Larsson (1979), the survey by Camerer (1995), Wakker (2001), Halevy (2007), Ahn, Choi, Gale, and Kariv (2014), Andreoni, Schmidt, and Sprenger (2014), Dean and Ortoleva (2015), and Kovářík, Levin, and Wang (2016). For theoretical responses to the paradoxes, see the surveys by Machina (2008), Gilboa and Marinacci (2011) and Machina and Siniscalchi (2013). For a critical assessment of the ambiguity aversion literature, see Al-Najjar and Weinstein (2009) and Siniscalchi (2009),

[6] A large literature in psychology finds that people have difficulty with various forms of hypothetical thinking. The early evidence dates back to Wason selection tasks (Wason 1966, 1968). See Evans (2007) and Nickerson (2015) for recent textbook treatments.

[7] There is also a large literature documenting mistakes in strategic settings, though these are not always directly linked to contingent thinking: Bazerman and Samuelson (1983), the survey by Kagel and Levin (2002), Charness and Levin (2009), Ivanov, Levin, and Niederle (2010), Esponda and Vespa (2014), and Levin, Peck, and Ivanov (2016) for mistakes in common value settings, such as auctions or elections, and Kagel, Harstad, and Levin (1987), Kagel and Levin (1993), and Harstad (2000) for overbidding in second-price private value auctions.

player who cannot perform a certain kind of contingent thinking can recognize as dominant. He shows experimentally that subjects play obviously dominant strategies at higher rates than non-obviously dominant ones.[8] In our election and auction problems, there is weak dominance of one action over another, so there is no obviously dominant action. But, in the contingent frame, assuming the subject prefers not to respond to the common consequence in event A^c, then there is an obviously dominant choice in event A and this might help explain why more subjects satisfy dominance in the contingent treatment. More recently, Cohen and Li (2022) present an extension of the notion of cursedness (Eyster and Rabin, 2005) designed to accommodate experimental findings on failures of contingent thinking such as the ones presented in this paper. Meanwhile, Chew and Wang (2022) rely on STP to extend the notion of obvious dominance. Piermont (2022) develops a model in which the decision-maker misperceives or ignores certain hypothetical events, which allows for STP to fail and can accommodate several of our experimental findings (see also Piermont and Zuazo-Garin (2022)). Finally, de Clippel (2022) provides implications of failures of STP in game theory and mechanism design.

While we focus on a specific aspect of contingent thinking (the sure-thing principle), a few papers formalize other aspects of contingent thinking, including models of local thinkers who focus on salient aspects of the decision problem (e.g., Gennaioli and Shleifer, 2010 and Gabaix, 2014).[9] In particular, Bordalo, Gennaioli, and Shleifer (2012) experimentally show that, in line with predictions from salience theory, the preference reversal of the common-consequence Allais problem is drastically mitigated if the randomization is correlated relative to the standard presentation.[10] Since our focus is instead on failures of STP, we use the correlated presentation for both frames (contingent and noncontingent) of the common-consequence Allais problem. Consistent with Bordalo, Gennaioli, and Shleifer (2012), we find relatively few preference reversals. But in our comparison between contingent and noncontingent frames, we document a large proportion of STP failures.

The focus of our paper is on testing possible failures of STP and the tool we use involves a

---

[8] Zhang and Levin (2017) extend these results to more general partitions of the state space.

[9] Other related work includes: Gilboa and Schmeidler (1995), who develop case-based decision theory (where agents choose acts based on past performance) partly motivated by environments where probabilities or states of the world are not salient (or easily accessible) features of the problem; Glazer and Rubinstein (1996), who illustrate how the extensive form can facilitate contingent thinking relative to the normal form of a game; and the experimental findings of Dal Bó, Dal Bó, and Eyster (2018), who document under-appreciation of equilibrium effects, and Martínez-Marquina, Niederle, and Vespa (2019), who show that mistakes in contingent thinking occur when these contingencies are framed in a probabilistic manner.

[10] Correlated here roughly means that the outcomes are presented state-by-state. Starmer (1992), Starmer and Sugden (1993) and Leland (1998) show that regret theory also predicts that behavior depends on the correlation between lotteries. However, the mechanism implied by regret theory differs from the mechanism under salience, and Frydman and Mormann (2016) provide further evidence in support of salience theory.

comparison between different frames, which connects our paper to the literature on framing effects. The potential difference between the noncontingent and contingent treatments has been previously examined in the specific context of the common-ratio Allais paradox. Tversky and Kahneman (1981) and Holler (1983) find significant differences between treatments, but Cubitt, Starmer, and Sugden (1998) find no significant differences in the marginal responses across treatments, like we do for this specific problem. Their focus, however, is different from ours and so they do not test if there is a reduction of reversals (we find there is). We also introduce a conceptual framework that allows us to go beyond the common-ratio Allais paradox and study anomalies in both decision and strategic environments.[11] Moreover, while there is a large literature documenting the importance of framing effects (e.g., Tversky and Kahneman, 1986), we find that a specific frame, linked to people's ability to put themselves in situations that have not occurred, has a systematic effect across a wide range of problems that seemed previously unrelated.

2 Experimental framework

We study five classic problems in the laboratory: Ellsberg (ELLS), common-consequence Allais (CC ALLAIS), a private-values second-price auction (AUCT), a common-value election (ELECT), and common-ratio Allais (CR ALLAIS). We begin by describing each of the problems and then we introduce the common experimental framework that applies to all problems.

2.1 Five problems

Each problem has the property that there is an event A in which the choice affects payoffs and an event A^c in which it does not. We now describe each problem separately.

Ellsberg (ELLS). Figure 1 depicts the classical one-urn Ellsberg discussed in the introduction, where there are 30 red (R) balls and 60 yellow (Y) or blue (B) balls. Here and in all related problems, a ball is randomly drawn (with replacement) for each question. In Question 1 (Q1), the subject must choose between two options, f (win \$10 if R or B, \$0 otherwise) and g (win \$10 if Y or B, \$0 otherwise). In Question 2 (Q2), the subject is asked to choose between two other options, f' (win \$10 if R, \$0 otherwise) and g' (win \$10 if Y, \$0 otherwise). A typical response pattern is to prefer g in Q1 and f' in Q2, which as explained in the introduction violates Savage's P2, which we call separability (SEP).

---

[11] In a theory paper, Eliaz, Ray, and Razin (2006) make a connection between anomalies in decision and game theory environments by establishing a formal equivalence between violations of expected utility theory and choice shifts in groups.

		A		A^c
		R (1)	Y (10)	B (89)
Question 1	f	\$100m	\$100m	\$100m
	g	\$0	\$500m	\$100m
Question 2	f'	\$100m	\$100m	\$0
	g'	\$0	\$500m	\$0

Figure 2: CC ALLAIS

Notes: There are 1 red (R), 10 yellow (Y), and 89 blue (B) balls. In each question, a ball is selected (with replacement). In both questions, the subject's choice only matters if the ball is R or Y.

Common-consequence Allais (CC ALLAIS). There is an urn with 100 balls. Of the 100 balls, 1 is R, 10 are Y, and 89 are B. In Q1, the subject must choose between f, which gives \$100 million for sure, and g, which gives \$500 million if the ball is yellow and \$100 million if it is blue.[12] In Q2, the subject must choose between f', which gives \$100 million if the ball is red or yellow, and g', which gives \$500 million if the ball is yellow. Figure 2 illustrates these payoffs. The paradox is that a significant number of subjects choose f (safe option) in Q1 and g' (risky option) in Q2. This reversal also constitutes a violation of SEP.

Auction (AUCT). Consider an auction where the highest bidder wins and pays the bid of the second-highest bidder. There are two bidders, the subject and one other bidder that is represented by a computer and is equally likely to bid amounts \$0.50, \$4.50, or \$8.50. The subject must choose an integer bid from \$1 to \$8. She gets \$5.50 if she wins the auction and an outside value of \$3 if she does not.[13]

Question 1 in Figure 3(a) depicts the choices faced by the subjects. There are three states of the world, corresponding to the three possible bids of the competitor, and the decision maker essentially chooses one of two options. Option f' corresponds to a bid of \$1, \$2, \$3, or \$4 while option g' corresponds to a bid of \$5, \$6, \$7, or \$8. Question* in the figure contrasts two options, f and g, that deliver constant payoffs of \$3 and \$1, respectively. We will not ask Question* in the experiment but will rather simply assume that more money is preferred to less, i.e., f is preferred

---

[12] This is the only decision problem for which we use hypothetical payoffs. Huck and Müller (2012) study three alternative ways of implementing the common-consequence Allais problem. They find few violations of SEP with small payoffs, but that violations become more prevalent with hypothetical payoffs expressed in millions of euros. For a recent meta study in the common-consequence Allais problem, see Blavatskyy, Ortmann, and Panchenko (2022). The meta study documents under which settings the paradox is more likely and less likely to appear. We also conducted a version of this problem with real payoffs; see footnote 22.

[13] While it is standard to normalize the outside value to be zero, here we illustrate the ideas using the payoffs from the actual experiment, where an outside value of \$3 guarantees that subjects do not incur losses. Moreover, we used a neutral, non-auction language to describe this problem to subjects, in order to avoid the issue pointed out by Cason and Plott (2014) that subjects may mistakenly interpret the second-price auction as a first-price auction.

		A	A^c	
		$4.50	$0.50	$8.50
Quest.*	f	$3	$3	$3
	g	$1	$1	$1
Quest. 1	f'	$3	$5	$3
	g'	$1	$5	$3

(a) AUCT

		A	A^c
		bWB	**wWW**
Quest.*	f	$5	$5
	g	$0	$0
Quest. 1	f'	$5	$5
	g'	$0	$5

(b) ELECT

		RY	B
Quest.2	f	x	x
	g	y	y
Quest. 1	f'	x	$0
	g'	y	$0

(c) CR ALLAIS

Figure 3: Dominance: Payoffs in the Auction, Election, and CR Allais problems

Notes: (i) AUCT: There are three possible bids of the computer bidder, $4.50, $0.50, and $8.50, but the subject's choice only matters if the bid of the computer is $4.50. Option f' corresponds to a bid of $1, $2, $3, or $4 while option g' corresponds to a bid of $5, $6, $7, or $8.
(ii) ELECT: In event A, the ball is black, computers vote differently, and the subject's vote matters. In event A^c, the ball is white, both computers vote White, and the subject earns $5 regardless of her vote.
(iii) CR ALLAIS :x is a lottery that gives a sure payoff of $4, y is a lottery that gives $5.30 with probability .8 and nothing otherwise, and 0 is a lottery that pays $0 for sure. In Q1, the subject's choice matters in state RY (the ball is red or yellow) but not in state B (blue). In Q1, the jar has 12 red, 3 yellow, and 85 blue balls. In Q2, the jar has 80 red balls and 20 yellow balls.
(iv) In AUCT and ELECT, Question* represents a trivial choice between a lower and a higher amount of money, illustrating that a preference reversal violates dominance. We do not ask this question but simply assume more money is preferred to less.

to g.[14]

In the auction, the choice of an integer bid from $1 to $8 is only relevant in the event that the other bidder bids $4.50. In this case, however, it is optimal to lose the auction by submitting a bid of $1, $2, $3 or $4, which is represented by f'. The paradox in this problem is that, in a treatment where subjects in the role of the decision maker are asked to submit an integer bid from $1 to $8, a significant fraction of subjects end up overbidding and choose g' over f'. Note that, under the assumption that f is preferred to g in Question*, which is the assumption that more money is preferred to less money, then this choice constitutes a reversal: f is preferred to g but g' is preferred to f'. This reversal is a violation of dominance (DOM).

Common-value election (ELECT). There is an urn with 7 white balls and 3 black balls, and one ball is randomly drawn. There are two computers. If the drawn ball is white (w), both computers vote White (WW). If the drawn ball is black (b), computers vote for different colors (WB). Without observing either the color of the drawn ball or the votes of the computers, the subject must choose between voting for Black and voting for White. If the color chosen by the majority matches the color of the drawn ball, the subject gets $5; otherwise, she gets $0. It is optimal for the subject to

---

[14] By stating Question* we make explicit an assumption that is commonplace in experimental research in the laboratory. Namely, that showing up to participate in an incentivized experiment is taken to indicate that subjects prefer more money to less. However, a recent paper by Gupta, Rigotti, and Wilson (2021) provides some direct evidence. One of their environments is a symmetric game in which subjects select between A that entails a payoff of either $15 or $16, and B that would deliver a payoff of $10 or $11. While A is obviously dominant in the sense of Li (2017), selecting it still involves contemplating more moving pieces than a problem like Question*. There is another player involved, and there is more than one possible payoff per action. With undergraduates in a laboratory setting, Gupta, Rigotti, and Wilson (2021) find that 97.3% of subjects select A.

vote for Black, since, if her vote matters, it must be that the ball is indeed black.

Question 1 in Figure 3(b) represents the payoffs. Voting for Black is represented by f' and voting for White is represented by g'. We also include Question* with options f (a sure payment of \$5) and g (a sure payment of \$0), but we do not ask this question and instead assume that f is preferred to g. The paradox in this problem is that a significant fraction of subjects choose g' over f', which is a violation of DOM.[15]

Common-ratio Allais (CR ALLAIS). There is a jar with 100 balls, and the subject answers two questions, Question 1 (Q1) and Question 2 (Q2). In Q1, the jar has 12 red, 3 yellow, and 85 blue balls. The subject must choose an option that gives \$4 if the drawn ball is red or yellow, and an option that gives \$5.30 if it is red. In Q2, the jar has 80 red balls and 20 yellow balls. The subject must choose between an option that gives \$4 for sure and an option that gives \$5.30 if it is red. Note that the ratio of red to yellow balls is the same in both jars, which explains the term "common-ratio" in this experiment.

To depict this problem in Savage's framework, let the space of consequences be given by $Z = \{x, y, 0\}$, where x is a lottery that gives a sure payoff of \$4, y is a lottery that gives \$5.30 with probability .8 and nothing otherwise, and 0 is a lottery that pays \$0 for sure. The set of states is $S = \{RY, B\}$, where RY is the state where the ball drawn from the urn is red or yellow and B is the state where it is blue. Figure 3(c) depicts the choices faced by the subject in each question. In Q1, the subject must choose between f' and g'. In Q2, the subject must choose between f (a sure payoff of \$4, represented by x) and g (a lottery that pays \$5.30 with probability .8, represented by y).[16] The typical paradox is that many subjects choose g' in Q1 and f in Q2. This is a violation of dominance (DOM).

2.2 Experimental design

For each of these five problems, we conduct two treatments. The *noncontingent treatment* is the benchmark treatment and it is intended to replicate existing results in the literature. In contrast, in the *contingent treatment* we elicit choices contingent on the event for which the consequences of the two options differ, and we investigate if this manipulation has a systematic effect across problems.

---

[15] In Online Appendix B, we describe the complete payoff table, including states that have zero probability, and argue that depending on the subject's beliefs about zero probability events, then either DOM or an objective version of DOM is violated. We will be content with either interpretation (DOM or objective DOM) since our main focus will be on STP.

[16] Even though the objective probability of the blue state is zero in Q2, we include it in the table to ease comparison with Q1. The important feature is that the subject chooses between x and y in Q2, which is the same choice contingent on event RY in Q1.

In the contingent treatment, we first describe the payoffs in event A^c and highlight that these payoffs do not depend on the subject's choice. Then, we ask the subject to make a choice contingent on event A being true. Crucially, we ask the subject to commit to a choice in case event A is realized *before* she knows whether or not the state of the world will be in A. Thus, the subject faces the same choice she faces in the noncontingent problem. In particular, as in the noncontingent treatment, the subject (i) also faces a static choice problem in the contingent treatment, and (ii) can also decide whether or not to take into account the payoff in event A^c.[17]

To illustrate the contingent treatment, consider Q1 in ELLS. We tell a subject that, if the ball is $(B)lue$, then she will get \$10. And, if the ball is $(R)ed$ or $(Y)ellow$, then she has a choice between the option that pays \$10 if the ball is R or \$0 if Y (f), and the option that pays \$10 if it is Y and \$0 if R (g). But we ask the subject to commit to a choice in case event $A = \{\text{R}, \text{Y}\}$ is realized *before* she knows whether or not the state of the world will be in A. Thus, the subject faces the same choice between f and g that is faced by a subject in the first question of the noncontingent treatment. For the second question, Q2, we proceed in the same way except that we tell the subject that she will get \$0 if the ball is B. Thus, the subject faces the same choice between f' and g' that is faced by a subject in the second question of the noncontingent treatment.

Our experimental framework is motivated by the observation that there are two underlying principles behind the failure of separability (SEP) and dominance (DOM). To illustrate, consider first the case of SEP in ELLS, where the typical reversal is to choose g in Q1 but f' in Q2. There are two underlying principles behind this reversal. The first is a contingent version of SEP. For example, if the subject continues to choose g and f' in the contingent treatment, then we say that a contingent version of separability, which we call C-SEP, is violated.

The second principle, and the main focus of the paper, is the sure-thing principle. As an example of this principle, consider Q1 and suppose that the subject prefers f over g in the contingent treatment. Assuming that the subject is indifferent between f and g contingent on A^c, then the sure-thing principle says that she should prefer f over g in the problem where the choice is not contingent, that is, in the noncontingent problem. In the experiment, f and g give exactly the same monetary payoffs in A^c. To focus on the issue of contingent thinking and to avoid testing for indifference between two options that deliver the same consequences under A^c, we test a version of the sure-thing principle where we elicit choice contingent on A and the subject is told exactly what payoff she would get if a state in A^c is realized, a payoff that does not depend on her choice.[18]

---

[17] For a discussion of *dynamic* choice problems where the subject is asked to move *after* receiving information about an event, see Online Appendix B.

[18] In addition, this choice not to test for indifference allows us to compare our results to the existing literature, which also does not ask for indifference.

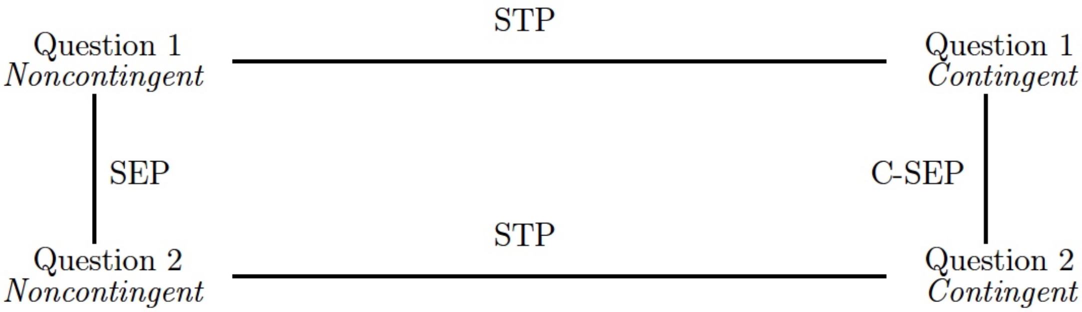


Figure 4: Experimental design and tests of STP, SEP, and C-SEP.

Notes: STP fails if there are differences between Question 1 noncontingent and Question 1 contingent, or differences between Question 2 noncontingent and Question 2 contingent. SEP fails if there are differences between Question 1 noncontingent and Question 2 noncontingent. C-SEP fails if there are differences between Question 1 contingent and Question 2 contingent.

Thus, we test a version of the sure-thing principle, which we abbreviate by STP, that says that if f is preferred to g contingent on A and f is *equal* to g outside of A, then f should be preferred to g. In other words, if f is preferred to g in the contingent treatment, then STP says that f should be preferred to g in the noncontingent treatment. The same is true regarding Q2 between f' and g'.

Figure 4 illustrates the testable hypotheses in the problems where the typical reversal violates SEP (i.e., ELLS and CC ALLAIS). A comparison between Q1 and Q2 in the noncontingent treatment is the standard way of testing for SEP in the literature. A comparison between Q1 and Q2 in the contingent treatment provides a test of C-SEP, our contingent version of SEP. A comparison between Q1 in the noncontingent treatment and Q1 in the contingent treatment provides a test of STP; the same is true for the comparison between Q2 across treatments.

Similarly, there are two principles underlying the failure of dominance (DOM). The first is a contingent version of DOM. In particular, if there is a preference reversal (i.e., if the subject chooses the dominated option) in the contingent treatment, then we say that a contingent version of dominance, which we call C-DOM, is violated. The second is STP as introduced above, which says that, for each question, the options chosen in the contingent and noncontingent framing must coincide.

Figure 5 illustrates the testable hypotheses in the problems where the typical reversal violates DOM (i.e., AUCT, ELECT,and CR ALLAIS). A comparison between Q1 and Q* (or Q2 in CR ALLAIS) in the noncontingent treatment tests DOM. This is the standard test in the literature. A comparison between Q1 and Q* (or Q2 in CR ALLAIS) in the contingent treatment tests C-DOM, our contingent version of DOM. Finally, a comparison between Q1 in the noncontingent treatment and Q1 in the contingent treatment tests STP.

In Online Appendix A, we provide a simple formalization of the connections between the dif-

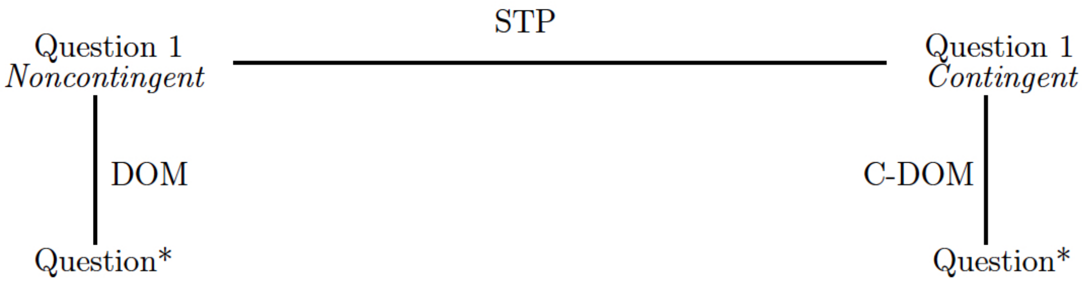


Figure 5: Experimental design and tests of STP, DOM, and C-DOM.

Notes: STP fails if there are differences between Question 1 noncontingent and Question 1 contingent. DOM fails if there are differences between Question 1 noncontingent and Question* noncontingent. C-DOM fails if there are differences between Question 1 contingent and Question* contingent. In AUCT and ELECT, Question* was not presented to subjects as we directly assumed that subjects prefer more money to less. In CR ALLAIS, Question* corresponds to Question 2 in Figure 3(c).

ferent concepts depicted in Figures 4 and 5. In particular, we show that, as the informal discussion in this section suggests, if STP and C-SEP (or C-DOM) hold, then SEP (or DOM) must also hold. These results formalize the claim that a very particular form of contingent thinking, embodied by the notion of STP, underlies behavior in classical environments from both decision and game theory. In the experiment, we will apply this simple framework to test to what extent it is failure of STP that explains the classic anomalies, as opposed to failures of C-DOM or C-SEP.

2.3 Treatments

For each of the five problems, we conduct two treatments, noncontingent and contingent, as described above. We conducted both *between-subjects* and *within-subject* designs. In the between-subjects design, each subject participated in one of the five problems and in one of the two treatments only. We conducted *three* experiments with a within-subjects design. In the first version, which we refer to as the *within* design, each subject participated in both treatments for all of the five problems using the same parameterization of the between-subjects design. Subjects first go over the noncontingent versions of all problems before going over the contingent versions of all problems. Provided we can replicate the between-subjects design, the within design provides valuable information to assess the effect of the treatment.

For robustness purposes, we conducted a second version of a within-subjects design, which we refer to as as the *within+* design, where we exposed subjects to several parameterizations of each of the five problems. We also conducted a third version that is identical to the first version except that the subjects were exposed to all five problems using the contingent treatment first and the noncontingent treatment second. We refer to this version as the *withinCNC* design.

Finally, in order to understand the mechanisms behind the treatment effects, we conducted an *auxiliary treatment* that uses a different presentation of the Ellsberg and Election problems that we introduce in Section 4.

The between-design and all the within-design experiments were conducted at the University of California, Santa Barbara and subjects were recruited using ORSEE (Greiner, 2015). There were 1,093 subjects who participated only once: 742 participants in the between-subjects design, a total of 351 in the within-subjects designs (131 in the *within*, 119 in *within+*, and 101 in the *withinCNC*). The auxiliary treatment was conducted at the University of California, San Diego and 71 subjects participated. The experiment was conducted using zTree (Fischbacher, 2007).[19] At the end of the experiment, each subject was asked additional questions to assess her level of risk or ambiguity aversion and cognitive ability. We describe these questions and demographic information in Online Appendix C and in the Procedures Appendix.

3 Results

In subsection 3.1 we use the between-subjects design to document failures of STP. Finding #3, which rejects the hypothesis that STP holds in all of the five problems that we study is the central finding. The following subsection shows that the findings hold in the within-subject experiment where a participant faces both contingent and noncontingent frames for each problem. The *within* design also allows us to document (finding #5) that the rates of STP failures at the individual level are between 19.9% and 72.5%, depending on the problem. In addition, we can show that subjects that are affected by the contingent/noncontingent frames largely go from being inconsistent with SEP or DOM to being consistent with C-SEP or C-DOM (finding #7), though we report that there is some degree of heterogeneity at the individual level.

Subsection 3.3 shows that the results hold for several parameterizations of each problem and studies correlation of inconsistencies across problems.

3.1 Between-subjects results

Figure 6 summarizes the findings for the between-subjects design. For each of the five problems and for each treatment (noncontingent, NC, and contingent, C), we report the number of observa-

---

[19] A session in the between-subjects (*within*, *within+*, *withinCNC*, *auxiliary*) design lasted approximately 30 (90, 120, 90, 30) minutes and on average subjects received \$9.50 (\$19.10, \$27.10, \$18.60, \$9.73) in compensation. In Online Appendix C we show that in terms of observables there are no significant differences between participants depending on their university.

tions and the percent of subjects making each of *four* (in ELLS, CC ALLAIS and CR ALLAIS, where subjects face two questions) or *two* (in AUCT and ELECT, where subjects face one question) possible choices. Based on these choices, we compute and report the percent of subjects failing SEP or DOM (in the noncontingent treatment) and C-SEP or C-DOM (in the contingent treatment). In the last row, we report, separately for each problem, the p-value of a test of the null hypothesis that percentage failure of SEP or DOM equals percentage failure of C-SEP or C-DOM. There are three main takeaways from this table and we start by documenting that if for each problem we focus on noncontingent treatments, we replicate previous results in the literature.

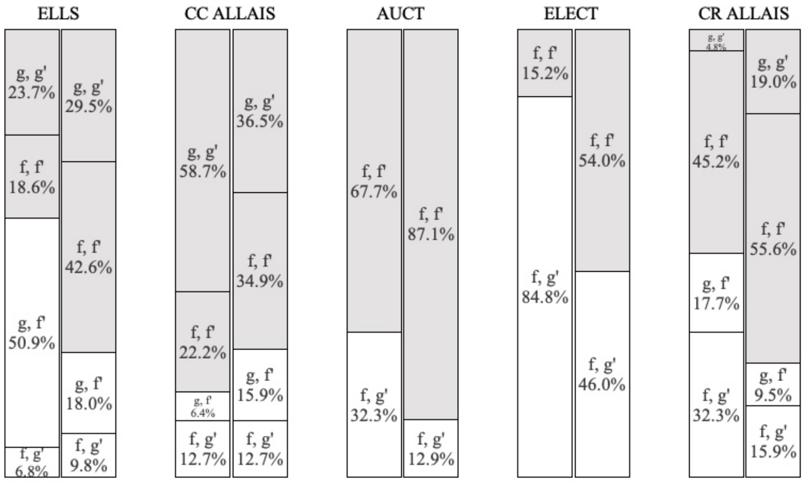
Finding #1. *We replicate the anomalies pointed out in the literature for all noncontingent versions of the problems.*

This finding is observed by looking at each column labeled NC (noncontingent treatment). Consider first the case of ELLS. Most subjects (50.9%) select g in Q1 and f' in Q2, a result that is in line with previous literature, as reviewed by Camerer (1995). These choices are consistent with the heuristic of ambiguity aversion (Ellsberg, 1961): subjects in this group prefer g in the first question (where the probability of receiving \$10 is known to be 2/3) and f' in the second question (where the probability of receiving \$10 is known to be 1/3). There is also a 6.8% of subjects who fail SEP by selecting f in Q1, but g' in Q2. Overall, the percent of subjects with choices that are inconsistent with SEP in ELLS is 57.7%. In the case of CC ALLAIS, in line with the literature, we find that the most common violation of SEP occurs when subjects select f in Q1 and g' in Q2. Such violation is consistent, for example, with the heuristic of regret aversion (Loomes and Sugden, 1982). The percent of subjects with choices that are inconsistent with SEP is 19.1%. This relatively low proportion of inconsistencies is in line with the findings of Bordalo, Gennaioli, and Shleifer (2012).[20,21]

---

[20] Our implementation of CC ALLAIS uses lotteries that are perfectly correlated. Once the state (i.e. a ball) is selected for one of the questions, it determines the payment for any of the two options. An alternative implementation simply describes the probability of obtaining each payoff for each option without explicitly describing a state space. It is thus possible to have a different randomization device to determine payoffs for each lottery, which would be uncorrelated. Bordalo, Gennaioli, and Shleifer (2012) compare both environments and find that inconsistencies are lower when lotteries are correlated (in an implementation similar to ours). This result, which is consistent with Salience Theory, rationalizes the larger inconsistencies observed when lotteries are uncorrelated. That is, in the case of CC ALLAIS our findings are consistent with the literature: when lotteries are described as correlated, deviations from SEP are relatively low. But our design allows us to evaluate to what extent STP is satisfied even when there are relatively few deviations from SEP.

[21] In addition, the experimental literature has documented that inconsistencies in CC ALLAIS may be larger or smaller depending on aspects of the implementation; see Huck and Müller (2012) for a study using college students and a broader population, and the recent meta-study of Blavatskyy, Ortmann, and Panchenko (2022). One concern is that our results may change if instead of using hypothetical payoffs we had used real payoffs. To evaluate such possibility, we replaced \$100m with 10 dollars and \$500m with \$20 dollars in Figure 2. We recruited 58 subjects for the noncontingent treatment and 60 subjects for the contingent treatment. We found that 39.6% of subjects violated



Treatment	NC	C	NC	C	NC	C	NC	C	NC	C
# of observations	59	61	63	63	62	62	66	63	62	63
% fail SEP/C-SEP	57.7%	27.8%	19.1%	28.6%	-	-	-	-	-	-
% fail DOM/C-DOM	-	-	-	-	32.3%	12.9%	84.8%	46.0%	50.0%	25.4%
p-value	0.001		0.258		0.011		0.000		0.001	

Figure 6: Between-subjects design: summary of results

Notes: 1) NC: Noncontingent treatment, C: Contingent treatment.
2) In ELLS and CC ALLAIS, (f, f') indicates the proportion of subjects who selected f in Q1 and f' in Q2. In AUCT and ELECT, (f, f') indicates choices of f in Q* and f' in Q1. We assume that all subjects prefer more money to less and so we impute that all subjects select f in Q*. In CR ALLAIS, (f, f') indicates choices of f in Q2 and f' in Q1.
3) Shaded cells correspond to choices that are consistent (inconsistent) with SEP/C-SEP or DOM/C-DOM, respectively. % fail SEP/C-SEP and % fail DOM/C-DOM presents the addition of subjects who chose (f, g') and subjects who chose (g, f').
4) The reported p-value results from a regression in which the unit of observation is a subject. The dependent variable is a dummy that takes value 1 if the subject's choices fail to satisfy the corresponding postulate, and the right-hand side includes a constant, a treatment dummy that takes value 1 if the subject participated in the contingent treatment, and the demographic variables detailed in Table 3. The p-values we report correspond to the coefficient estimated for the treatment dummy.

Choices of g' in Q1 of AUCT and ELECT are inconsistent with DOM under the assumption that subjects prefer more money to less. In Figure 6, we force this assumption by imposing that all subjects would select f in Q*. We find that 32.3% and 84.8% of subjects select g' in AUCT and ELECT, respectively. In AUCT, the observed overbidding is consistent with the illusion that it increases the chance of winning with little cost because the winner pays the second highest bid (Kagel, Harstad, and Levin, 1987). In ELECT, the mistake is consistent with the heuristic that subjects choose the color that is more prevalent in the jar. These results are in line with previous evidence on overbidding in second-price auctions (Kagel (1995)) and non-pivotal voting (Esponda and Vespa (2014; 2018)).

Finally, in CR ALLAIS, DOM is violated if subjects' choices are (f, g') or (g, f'). Figure 6 shows that 50% of subjects make choices inconsistent with DOM, qualitatively in line with previous findings; see Camerer (1995). As in ELLS and CC ALLAIS, the most prevalent anomaly in CR ALLAIS coincides with the one found in the literature, which corresponds to choosing (f, g'). These choices are consistent with the certainty effect heuristic (Tversky and Kahneman, 1986).

In terms of Figures 4 and 5, our first finding focuses on the left-most comparison: contrasting answers to Question 1 of the noncontingent to Question 2 of the contingent treatment, or Question* for SEP and DOM, respectively. We can subsequently compute to what extent inconsistencies arise for C-SEP and C-DOM (right-most comparison in the Figures) by focusing for each problem on subjects who participated in the contingent treatment. Finally, we can compare inconsistencies in noncontingent relative to contingent treatments, which is what we state in our second finding.

Finding #2. *In all problems except* CC ALLAIS*, the anomalies drop by about half in the contingent treatment; in* CC ALLAIS*, anomalies increase but the difference is not statistically significant.*

In particular, reversals decrease from 57.7% to 27.8% (p-value of .001) in ELLS and from 50.0% to 25.4% (p-value of .001) in CR ALLAIS. Inconsistent choices decrease from 32.3% to 12.9% in AUCT and from 84.8% to 46% in ELECT, with both differences being statistically significant.[22]

SEP in the noncontingent treatment. In the contingent treatment, violations of C-SEP are almost identical, at 41.7%, and the difference is not statistically significant (p-value .826). Moreover, the three findings listed in this section are robust to using hypothetical or real payoffs in CC ALLAIS. The percentages of choices between the noncontingent and contingent treatments were 19.0 vs 28.3 for (f, f'), 41.4 vs 30.0 for (g, g'), 29.3 vs 10.0 for (f, g'), and 10.3 vs 31.7 for (g, f'). We find no significant difference in the choice of f in Q1 (48.3% vs. 38.3%; p-value of .280), but we find a significant difference in the choice of f' in Q2 (29.3% vs. 60.0%; p-value of .001).

[22] We conduct a linear regression in which the unit of observation is a subject. The dependent variable is a dummy that takes value 1 if the subject's choices are inconsistent with the corresponding postulate, and the right-hand side includes a constant, a treatment dummy that takes value 1 if the subject participated in the contingent treatment, and the demographic variables detailed in Table 3. The p-values we report correspond to the coefficient estimated for the treatment dummy. A similar approach is followed for the other tests reported in the paper except that when the dataset

Finally, the proportion of subjects making inconsistent choices in CC ALLAIS increases from 19.1% to 28.6%, but this difference is not statistically significant (p-value of .258).

Our second finding indicates that, when it makes a difference, the contingent treatment works to reduce the proportion of inconsistencies. This result suggests that STP fails. However, observing that the proportion of inconsistencies do not drop (as in CC ALLAIS) is not evidence that STP holds. Upon inspection of Figure 4 it is possible to conceive a situation where the proportion of subjects satisfying SEP and C-SEP are not different but where STP for the first and/or second questions fails. Our third and main finding directly tests whether STP holds in each problem.

Finding #3. *We find a treatment effect in all problems. In particular, by comparing the joint distribution of responses across treatments, we can reject the hypothesis that STP holds in all of the five problems.*

We can also look at the extent to which differences in Q1 or Q2 are driving the failures of STP. In Q1 of ELLS, f is preferred by 25.4% of subjects (this is the sum of $(f, f')\%$ and $(f, g')\%$) in the noncontingent treatment and by 52.4% of subjects in the contingent treatment; the difference is statistically significant (p-value of .005). We also find significant differences for Q1 of ELECT (15.2% vs. 54%; p-value of .000) and Q1 of AUCT (67.7% vs. 87.1%; p-value of .011). For Q1 of CC ALLAIS, there is a difference but it is marginally not significant (34.9% vs 47.6%; p-value of .25). Moreover, there is a significant difference for Q2 of CC ALLAIS (28.6% vs. 50.8%; p-value of .013). Finally, we do not find a significant difference for either Q1 or Q2 of CR ALLAIS. The result for Q2 is reassuring because Q2 is the same question in the noncontingent and contingent treatments for this problem. Interestingly, despite no difference in the marginal distributions, the joint distribution for CR ALLAIS is significantly different, as evidenced by the drop in half in anomalies.

Our last finding of this section documents that STP fails broadly in all the problems that we study. This suggests that re-framing the same problem (by emphasizing the states where the consequences of actions differ) does affect subjects' choices. Moreover, finding #3 in conjunction with finding #2 indicates that the difference in responses across contingent and noncontingent treatments have directional consequences. In problems where we observe failures of STP and changes in inconsistencies comparing SEP to C-SEP (or DOM to C-DOM), we observe fewer inconsistencies in the contingent relative to the noncontingent treatment.

The strength of the between-subjects comparison is that we can use uncontaminated answers to test for STP. For example, a subject who participates in the noncontingent ELLS treatment faces Q1 of that problem without having answered anything prior to that. A similar situation corresponds to a subject who participates in the contingent ELLS treatment. Comparing the choices of subjects

---

involves more than one observation per subject (e.g *within* design), we cluster standard errors by subject.

who participate in the noncontingent to the contingent version allows us to cleanly assess whether STP holds in the aggregate. But a clear shortcoming of the between-subjects design is that we cannot assess to what extent STP fails at the individual level because no subject faced the same question in the contingent and in the noncontingent frames. To gain further insight we use the *within* design to which we turn next.

3.2 Failures of postulates at the individual level

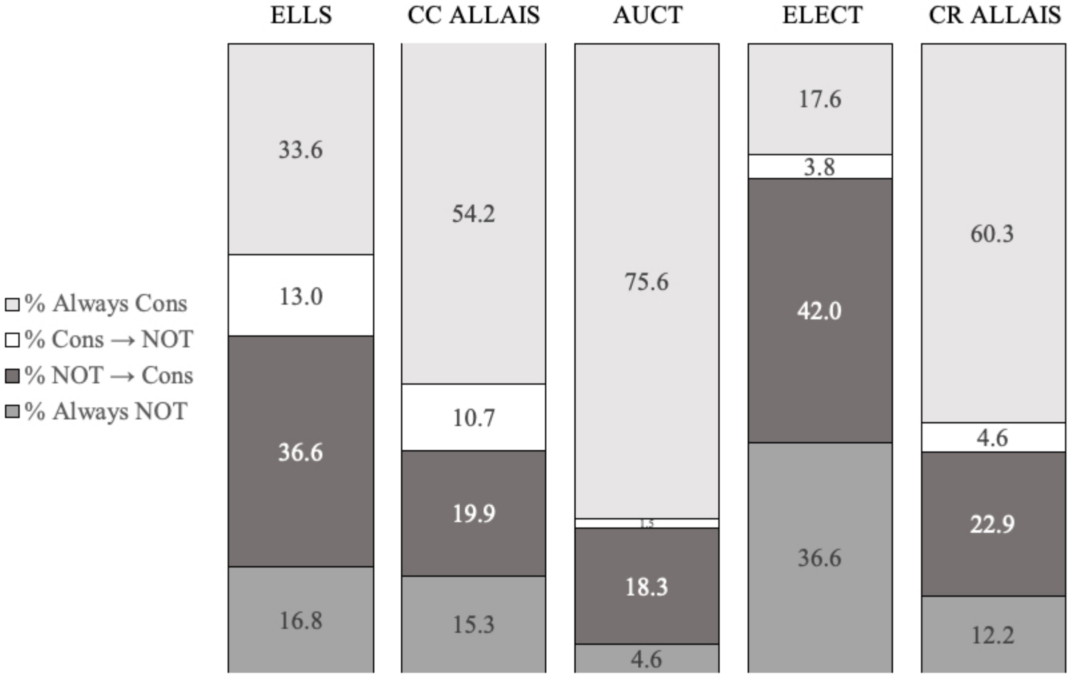
Figure 7 summarizes the findings for the *within* design. Recall that this is an experiment with 131 subjects, where each subject participates first in the noncontingent version of all problems and then in the contingent version of all problems. Since all subjects in the *within* design face the noncontingent and the contingent version of each problem, we can evaluate whether each subject satisfies SEP and C-SEP (or DOM and C-DOM) in each problem. Prior to presenting our findings in the *within* design, we describe how we summarize the data.

For each problem, there are four possible outcomes regarding consistency with the relevant postulate: always consistent (i.e., satisfies the relevant postulate in both treatments; e.g., satisfies SEP and C-SEP in ELLS), always not consistent, inconsistent in the noncontingent treatment but consistent in the contingent treatment ($NOT \rightarrow Cons$ in the figure) and the other way around ($Cons \rightarrow NOT$ in the figure). Based on these four outcomes, we compute and report the percent of subjects failing SEP/DOM and C-SEP/C-DOM. In particular, failure of SEP/DOM is given by the sum of $\%Always\, NOT$ and $\%NOT \rightarrow Cons$, since these were the subjects who were not consistent in the noncontingent treatment. Similarly, the failure of C-SEP/C-DOM (i.e., the failure of the relevant postulate in the contingent treatment) is given by the sum of $\%Always\, NOT$ and $\%Cons \rightarrow NOT$. In the table below the figure, we also report the p-value from a test of the null hypothesis of an equal proportion of failures of SEP/DOM and C-SEP/C-DOM.

Towards the bottom of the table in Figure 7, we report the percent of STP violations for each problem. For the problems that have two questions to test STP (ELLS and CC ALLAIS), we say that the subject fails STP if it fails for at least one of the questions. The bottom three rows in the table will be explained later. There are four main takeaways from Figure 7.

Finding #4. *The findings of the within design are in line with the previous findings from the between-subjects design.*

Inspection of Figures 6 and 7 shows that the levels of failure of SEP/DOM and C-SEP/C-DOM are comparable to the levels observed in the between-subjects design. In particular, failures of consistency drop by about one-half in all problems except CC ALLAIS, where the difference continues



	ELLS	CC ALLAIS	AUCT	ELECT	CR ALLAIS
% fail SEP/DOM	53.4	35.1	22.9	78.6	35.1
% fail C-SEP/C-DOM	29.8	25.9	6.1	40.5	16.8
(▲) p-value	0.000	0.058	0.000	0.000	0.000
% fail STP	72.5	43.5	19.9	45.8	25.2
$q_{C\|I}$	0.69	0.57	0.80	0.53	0.65
$q_{I\|C}$	0.28	0.16	0.02	0.18	0.07
(⊗) p-value	0.000	0.000	0.000	0.000	0.000

Figure 7: *Within* design: summary of results

Notes: 131 Observations. Always Cons: consistent with the relevant postulate in noncontingent and contingent. Cons → NOT: consistent with postulate in noncontingent, but inconsistent in contingent, NOT → Cons: inconsistent with postulate in noncontingent, but consistent in contingent. Always NOT: inconsistent with postulate in noncontingent and contingent.
(▲): p-values for the null hypothesis that failures of SEP/DOM are not different than failures of C-SEP/C-DOM. Details in footnote 23.
$q_{C|I}$: estimated probability of switching to being consistent (contingent treatment) conditional on being inconsistent in the noncontingent treatment; $q_{I|C}$: estimated probability of switching in the opposite direction (becoming inconsistent conditional on being consistent in the original treatment).
(⊗): p-values for the null hypothesis that $q_{C|I} = q_{I|C}$. For details see footnote 26.

to be not significant.[23,24] Based on this evidence, we feel more confident in taking advantage of the additional information that the *within* design provides.

Finding #5. *There is a large treatment effect across all problems, as evidenced by the number of failures of STP documented.*

The range of STP violations goes from a low of 19.9% in AUCT to a high of 72.5% in ELLS. The fact that there is a large number of STP violations and that inconsistencies with separability and dominance are lower in the contingent version of the problem suggests that most subjects move from not being consistent to being consistent (i.e. $NOT \rightarrow Cons$). The *within* design can help us understand to what extent this is indeed the case. Note that a lower bound for failure of STP is given by those subjects whose consistency is affected by the treatment (either $NOT \rightarrow Cons$ or $Cons \rightarrow NOT$), except in CR ALLAIS, where only Q1, and not Q2, can be used to test for STP. This bound is tight in the AUCT and ELECT problems. But for ELLS and CC ALLAIS, where there are two questions to test for STP, a subject who is either always or never consistent may still provide different responses to these questions and, therefore, fail STP.

Finding #6. *Subject heterogeneity: In all problems, most subjects whose consistency is affected by the treatment go from being inconsistent to consistent, but there is also a considerable number of subjects who go from consistent to inconsistent in the case of SEP.*

To the extent that the desire to satisfy SEP is determined by the preferences of a subject, there is no a priori reason to believe that every subject should be more likely to satisfy this postulate once faced with the contingent treatment. Indeed, while 36.6% of subjects go from inconsistent to consistent in ELLS, there are also 13% of subjects who go from consistent to inconsistent.

By definition, the difference between $\%NOT \rightarrow Cons$ and $\%Cons \rightarrow NOT$ is exactly the difference between subjects failing the noncontingent and contingent versions of the postulate. For example, in the case of CC ALLAIS, 19.9% of subjects change consistency in one direction and 10.7% in the other direction; these effects essentially cancel out and the difference of 9.2

---

[23] The test proceeds in the following manner. Let $j \in \{$ELLS, CC ALLAIS, AUCT, ELECT, CR ALLAIS$\}$ capture each of the problems. For each subject D_{Inc} is a dummy variable that takes value 1 if the subject is not consistent with the corresponding postulate (SEP/DOM in the noncontingent versions and C-SEP/C-DOM in the contingent versions). D^C is a dummy that takes value 1 if the observation is from a Contingent problem and the dummy variable D_j takes value 1 if the answers are for problem j. We run the following regression, in which for each subject we have ten observations (corresponding to the contingent and noncontingent answer to each of the five problems): $D_{Inc} = \sum_j \left(\delta_j D_j + \phi_j (D^C \times D_j)\right) + \upsilon$, where υ is an error term. The null hypothesis of interest is that there is no effect of the contingent treatment for each j. The (▲) p-values reported in the table of Figure 7 correspond to the null hypothesis that the corresponding coefficient $\phi_j = 0$. Standard errors are clustered by subject.

[24] See Online Appendix D for a more detailed analysis of Findings 1-3 using the data from the *within* design. In addition, Table 11 of Online Appendix D provides a detailed comparison of the *between* design and *within* design.

percentage points, which is exactly the difference between %fail SEP (35.1%) and %fail C-SEP (25.9%), is not statistically significant. But, despite these effects cancelling out, there are still a total of 30.6% of subjects in CC ALLAIS whose consistency is affected by the treatment. Interestingly, when the relevant postulate is dominance (as in AUCT, ELECT, and CR ALLAIS), there are essentially no switches in the direction $Cons \to NOT$.

Another way to interpret these numbers is to note that, if subjects' actual preferences were more clearly elicited by the contingent treatment, then we would be incorrectly misclassifying a large fraction of subjects using the noncontingent treatment, where this large fraction is equal to the sum of $\%NOT \to Cons$ and $\%Cons \to NOT$, a figure that is as high as 49.5% in ELLS. Taken together, the two previous findings indicate that, while the treatment can move subjects in either direction, in all problems except CC ALLAIS, the net effect is to decrease inconsistencies.

The data analyzed so far, however, does not account for base rates of consistency and, in particular, does not yet prove that the contingent treatment is more likely to convert an inconsistent subject into consistent than vice versa. To investigate this issue, we define $q_{C|I}$ to be the estimated probability of switching to being consistent (contingent treatment) conditional on being inconsistent in the noncontingent treatment. Similarly, let $q_{I|C}$ be the estimated probability of switching in the opposite direction: becoming inconsistent conditional on being consistent in the original treatment.

Finding #7. *In all problems, the estimated probability of switching to being consistent conditional on being inconsistent ($q_{C|I}$) is significantly higher than the estimated probability of switching to being inconsistent conditional on being consistent ($q_{I|C}$).*

These probabilities, $q_{C|I}$ and $q_{I|C}$, are reported in the last rows of the table at the bottom of Figure 7, pg. 20, together with the p-value for the test of the null hypothesis that $q_{C|I} = q_{I|C}$.[25,26] For ELLS, CC ALLAIS, and ELECT, $q_{C|I}$ is about three times higher than $q_{I|C}$. For AUCT and CR ALLAIS, the differences are even larger, as almost all subjects who are originally consistent

---

[25] For example, for ELECT,

$$q_{C|I} = \frac{42}{36.6 + 42} = .53 > .18 = \frac{3.8}{3.8 + 17.6} = q_{I|C}.$$

[26] To test this null hypothesis we proceed in the following manner. Let $j \in$ {ELLS, CC ALLAIS, AUCT, ELECT, CR ALLAIS} capture each of the problems. For each subject in the contingent (noncontingent) treatment of each problem D^{C}_{Inc} (D^{NC}_{Inc}) is a dummy that takes value 1 if the subject is not consistent. We run the following regression, in which for each subject we have five observations (corresponding to the five problems): $D^{C}_{Inc} = \sum_{j} \left(\alpha_j D_j + \beta_j (D^{NC}_{Inc} \times D_j)\right) + \epsilon$, where ϵ is an error term. The null hypothesis $q_{C|I} = q_{I|C}$ implies that $2\alpha_j = 1 - \beta_j$. The reported p-values in the last row of the table in Figure 7 correspond to a Wald test of this equality for the corresponding j. In the estimation we cluster standard errors by subject.

remain consistent in the contingent treatment.

The calculation of $q_{C|I}$ and $q_{I|C}$ can also help us understand why there is no (significant) net effect on consistency for CC ALLAIS, despite large changes in response to the treatment, as described earlier. In particular, the difference between $q_{C|I} = .57$ and $q_{I|C} = .17$, while considerable, is not sufficiently large to lead to a statistically significant reduction in inconsistencies given that the baseline rate of inconsistency in the noncontingent treatment is 35.1%. It is interesting to compare these figures to the ones for AUCT, where the difference between $q_{C|I} = .80$ and $q_{I|C} = .02$ is sufficiently large to lead to a significant reduction in inconsistencies, despite a baseline rate of inconsistency of only 22.9% in the noncontingent treatment.

Figure 13 of Online Appendix G reproduces Figure 7, but using data from the *withinCNC* design that is identical to *within* except that it inverts the order of the frames: Subjects face all problems and parameterizations in the contingent frame first and in the noncontingent frame second. Consistent with Finding #5, there is a large fraction of subjects who fail STP, ranging from 20.8% (AUCT) to 69.3% (ELLS). There is also similar evidence of heterogeneity, consistent with Finding #6 and in all problems $q_{C|I}$ is higher than $q_{I|C}$, in line with Finding #7. Overall, the evidence suggests that failures of STP remain when treatments are faced in different order but that some of the differences become smaller.

3.3 Multiple parameterizations of each problem

The previous results were obtained for specific parameterizations of five classical problems. To check the robustness of our results and to be able to assess correlations between problems, we conducted an additional, within-subjects experiment where subjects were exposed to both the noncontingent and contingent versions of *multiple* (rather than just one, as before) parameterizations for each of the five problems. In particular, we considered five parameterizations for ELLS, three for CC ALLAIS, ten for AUCT, ten for ELECT, and four for CR ALLAIS. Following the *within* design reported earlier, each subject in this new experiment participated in all parameterizations of all five problems, with all noncontingent versions taking place before all the contingent versions. We refer to this experiment as the *within+* design.

In Online Appendix E we describe the details of the different parameterizations and we also contrast the results of this *within+* experiment with previous results. As we discuss in the online appendix, the findings of the *within+* design are qualitatively in line with the previous findings from the *within* design, with two minor differences. First, average inconsistencies continue to drop in the contingent relative to the noncontingent treatment, but not as dramatically as before. Presumably, this is driven by the fact that the *within+* design is much longer, lasting two hours. Second, for

CC ALLAIS, there is now a significant drop in inconsistencies in the contingent treatment driven by our third parameterization, where we used real payoffs as high as \$50 to incentivize subjects. In addition, for all parameterization of CC ALLAIS, regardless of whether the drop in inconsistencies is significant or not, we continue to document that $q_{C|I}$ is significantly higher than $q_{I|C}$.

One advantage of having multiple measures of consistency for each subject and problem in the *within+* design is that we can obtain reliable measures of correlations across problems. We focus on correlations between dummy variables that describe if a subject satisfies or not the relevant postulate for each problem (separability or dominance). Correlations provide further evidence about the stability of preferences within a frame. For example, if there is a frame where people like to satisfy separability in ELLS but not in CC ALLAIS, one should be more cautious about drawing inferences about people's preference to satisfy separability from that frame.

Table 1 shows correlations of consistency with separability and dominance both for the noncontingent version (SEP or DOM) and for the contingent version (C-SEP or C-DOM) of the problems (see Online Appendix E for a detailed explanation of how these correlations are computed). Our main finding is that there is a positive correlation across problems for both separability and dominance in the contingent treatment (where there are fewer of these failures), but no correlation in the noncontingent treatment. This suggests that the noncontingent treatment (i.e., the typical treatment in the literature) may provide a misleading picture of people's preference to satisfy or not the standard postulates. It also suggests that the contingent treatment provides a more coherent picture of people's tendency to satisfy these postulates.

Finding #8. *There is a positive correlation in satisfaction of separability across problems and also in satisfaction of dominance across problems for the contingent versions, but not so for the noncontingent versions of the problems.*

For ELLS and CC ALLAIS, consistency with (contingent) separability in one problem is correlated with consistency in the other problem. The correlation is .482. For AUCT, ELECT, and CR ALLAIS, consistency with (contingent) dominance in one problem is correlated with consistency in each of the other problems. In particular, the correlation between AUCT and ELECT is .419, the correlation between ELECT and CR ALLAIS is .389, and the correlation between AUCT and CR ALLAIS is .415. Across postulates, the correlation is very close to zero or very small, with the exception of the correlation between CC ALLAIS and CR ALLAIS of .692. Meanwhile, the correlations are very different for the noncontingent treatment. For example, ELLS and CC ALLAIS are negatively correlated (not significantly different from zero), even though separability underlies both problems, and several other correlations are negative or have no clear pattern.

	ELLS	CC ALLAIS	AUCT	ELECT	CR ALLAIS
ELLS	—	*-.560* **.482****	*-.006* **-.060**	*.111* **-.030**	*.298* **.086**
CC ALLAIS	—	—	*.254* **-.070**	*-.250* **.047**	*-.248* **.692*****
AUCT	—	—	—	*.198** **.419*****	*.006* **.415*****
ELECT	—	—	—	—	*-.429* **.389****
CR ALLAIS	—	—	—	—	—

Table 1: Correlations across problems

Notes: Correlations in Italics are for noncontingent versions. Correlations in Bold are for contingent versions. Correlations are computed using standard corrections for multiple measures; see Online Appendix F for details on how correlations are computed. Standard errors are computed by bootstrapping. The stars indicate that each is significantly different fro zero at the 1%(***), 5%(**) and 10%(*) level. Standard errors are clustered by subject.

4 Evidence on the mechanism behind the drop in anomalies

We documented differences in behavior between the noncontingent and contingent frames that are systematic across a wide range of settings. In particular, subjects typically behave in a manner that is more consistent with classical rationality postulates (dominance and separability) in the contingent frame. In this section, we examine the possible mechanisms underlying this treatment effect.

There are two channels through which the contingent frame may be affecting decisions. The first channel highlights the payoff structure of the problem. Specifically, it emphasizes that the decision only matters in event A because the payoffs are the same in event A^c. In addition, it computes the payoffs in event A^c, which may be nontrivial in problems such as the election, where where the payoff in case the subject is not pivotal depends on the state of the world and how the computers vote. The second channel is more closely connected to the notion of hypothetical thinking. In particular, the contingent treatment asks subjects to make a choice conditional on the event that A realizes. That is, subjects are asked to put themselves in the hypothetical position in which event A is true, and this is something they may not do on their own even if they were aware that their choice only matters in A.

To understand to what extent the treatment effect between the contingent and noncontingent frames is driven by these two channels, we introduce a third frame, which we will call the *auxiliary* frame (Aux). The aim of this third frame is to clearly inform subjects of the payoff structure of the problem, without asking them to make a choice contingent on event A being true. In other words, of the two channels described above, the first one is still present but the second one is absent in this new frame.

We focus on two problems, the Ellsberg (ELLS) and common-value election (ELECT) problems. These are the two problems in which anomalies show the largest drop (29.9 and 38.8 percentage points in ELLS and ELECT, respectively; see Table 6), and the highest rates of STP failures (72.5 and 48.8 percent in ELLS and ELECT, respectively; see Table 7). Given that the reductions in anomalies are large in these two problems, there is room to better understand what is driving the treatment effect.

In the case of Q1 in ELLS, subjects in this new Aux treatment are told that, if the ball is Red, their payoff is \$10 if they pick option 1 and \$0 if they pick option 2; if the ball is Yellow, their payoff is \$0 if they pick option 1 and \$10 if they pick option 2; and that if the ball is Blue their payoff is \$10 if they pick option 1 and \$10 if they pick option 2. In fact, we mention that if the ball is blue their payoff is \$10 regardless of their choice. But we then simply ask them to make a choice between options 1 and 2, as opposed to asking them to make a choice that is contingent on the event $\{Red, Yellow\}$ being true. Similarly, for ELECT, we break out the payoff structure into an event A in which their vote is relevant and an event A^c in which it is not. We then tell subjects that their vote is only relevant in event A and compute the payoffs for them in event A^c.[27] But we then simply ask them to cast a vote, as opposed to asking them to cast a vote that is contingent on the event A that their vote is pivotal.

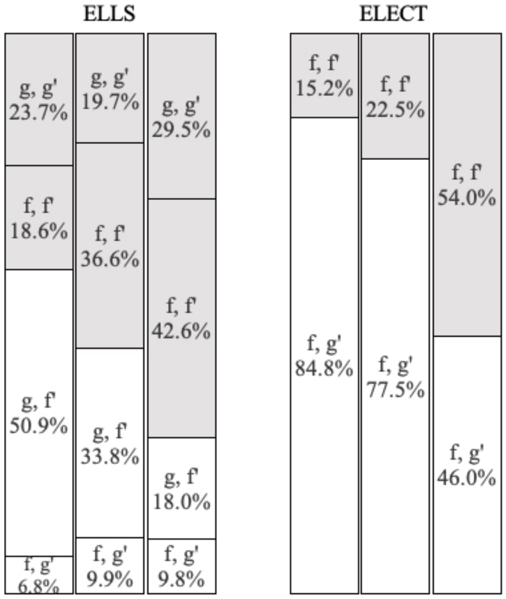
Our previously discussed Findings #2 and #3 show that failures of the underlying postulate (SEP or DOM, respectively) fall in the contingent treatment relative to the noncontingent in both ELLS and ELECT. If failures of the underlying postulate in Aux are close to those in NC, but remain higher than those in C, this would suggest that the drop in failures observed in C is in large part driven by the second channel, providing direct evidence about the failure of contingent thinking. If, on the other hand, failures of SEP or DOM in Aux are close to those observed in C, but far from NC, then it suggests that the treatment effect is mainly driven by the first channel.

We conducted five sessions with 71 participants that faced the Aux frame of both the ELLS (first) and the ELECT (second) problems.[28] Figure 8 presents the findings and, as a reference, also reproduces the findings for the noncontingent (NC) and contingent (C) frames in the Between design. In both problems, treatment Aux is in between NC and C in terms of failures of the

---

[27] Specifically, we tell participants that if the computers vote for different colors, the subject's choice will determine the color chosen by the majority. In this case, if computers vote for different colors and the subject votes for White (Black), they would receive \$5 if the ball drawn from the jar is White (Black) and \$0 otherwise. Meanwhile, we tell them that if computers vote for the same color, then that color would be selected by the majority and that the subject's payoff will not depend on how they voted. We also compute payoffs in this case for them. We tell them that if both computers vote for White (Black), they will get \$5 if the ball drawn from the jar is White (Black), and \$0 otherwise.

[28] Sessions for Aux were conducted at UC San Diego. Table 6 in Online Appendix C provides a comparison between these participants and participants in the between design, which was conducted at UC Santa Barbara. There is no difference in the samples in terms of our measures of cognitive ability, gender, or stage of their college studies (a proxy for age). There is a small difference in terms of risk aversion, with the UC San Diego sample being slightly more risk loving. More participants in UC San Diego are majoring in economics. The p-values that compare across samples that we report in the text control for these observables.



Treatment	NC	Aux	C	NC	Aux	C
# of observations	59	71	61	66	71	63
% fail SEP postulate	57.7%	43.7%	27.8%	-	-	-
% fail DOM postulate	-	-		84.8%	77.5%	46.0%
p-value NC v. Aux	0.215			0.279		
p-value Aux v. C		0.024			0.000	

Figure 8: Aux frame vs. contingent & noncontingent frames

Notes: 1) NC: Noncontingent treatment, C: Contingent treatment, Aux: Auxiliary treatment
2) In ELLS, (f, f') indicates the proportion of subjects who selected f in Q1 and f' in Q2. In ELECT, (f, f') indicates choices of f in Q* and f' in Q1. We assume that all subjects prefer more money to less and so we impute that all subjects select f in Q*.
3) Shaded cells correspond to choices that are consistent (inconsistent) with SEP/C-SEP or DOM/C-DOM, respectively. % fail SEP/C-SEP and % fail DOM/C-DOM presents the addition of subjects who chose (f, g') and subjects who chose (g, f').
4) The reported p-values results from regressions in which the unit of observation is a subject. The dependent variable is a dummy that takes value 1 if the subject's choices fail to satisfy the corresponding postulate, and the right-hand side includes a constant, the demographic variables detailed in Table 3, and a treatment dummy that takes value 1 if the subject participated in the NC treatment and 0 if the subject participated in the Aux treatment (for the p-value NC v. Aux) or a treatment dummy that takes value 1 if the subject participated in the Aux treatment and 0 if the subject participated in the C treatment (for the p-value Aux v. C). The p-values we report correspond to the coefficient estimated for the corresponding treatment dummy.

corresponding SEP or DOM postulate. In ELLS, failures of SEP drop from 57.7 percent in NC to 43.7 percent in Aux, and further to 27.8 percent in C. In ELECT, failures of DOM fall from 84.8 percent in NC to 77.5 percent in Aux, and further to 46 percent in C. In both cases, however, failures in NC are not statistically different than failures in Aux (p-values of 0.215 and 0.279, in ELLS and ELECT, respectively), but failures in Aux are statistically different than failures in C (p-values of .024 and <0.000).

The evidence suggests that the reduction in anomalies when comparing NC to C mainly comes from the second channel, that is, from the fact that in the contingent frame subjects are asked to make a choice as if the relevant state is in set A. This suggests that the treatment effect is mainly driven by people's failure to put themselves in situations that have not yet happened but that are the relevant ones for the decision problem at hand.

5 Conclusion

We proposed a common experimental framework to study the extent to which failures of classical rationality postulates (SEP and DOM) can be attributed to failure of contingent thinking. Our motivating observation is that these postulates rely on the more basic sure-thing principle (STP), and that the specific role of STP that we highlight in this paper in connection to SEP and DOM has largely been overlooked.

We find that some of the most common anomalies uncovered in laboratory experiments, including overbidding in auctions, naive voting in elections, and Ellsberg and Allais types of paradoxes, spring, at least in large part, from the failure of STP. Our strategy is to run subjects through standard versions of each of these canonical problems (noncontingent frame), and then to run subjects through slight alterations of each problem (contingent frame), which focus on the subset of states where their choices affect payoffs. STP would be satisfied if choices were the same across the two frames, but we find that many subjects behave differently in the two versions of the problem, and that this difference explains about half of the anomalies in many classic problems.

Moreover, we provide evidence that the differences in anomalies between the noncontingent and contingent frame are in large part driven by asking subjects to make choices contingent on the event for which their choice is relevant. In other words, to eliminate differences in anomalies between the noncontingent and contingent frames, it does not suffice to simply tell people that their choice is relevant in a particular event and to describe the payoffs in the complementary event. For there to be a treatment effect, we also need to ask people to make a choice conditional on the hypothetical event that the relevant event occurred. This suggests that people fail to put themselves in situations that have not occurred, which is a key aspect of contingent thinking.

Our findings cast doubts on claims that anomalies are mainly driven by (nonstandard) preferences. Instead, our findings suggest that it is important to examine other aspects of a decision problem, such as the manner in which people mentally represent the problem. More specifically, our approach of unpacking failures of standard postulates into failures of STP or genuine failures of standard postulates can be applied to understand related anomalies by developing theoretical frameworks that allow STP to fail. Recent theoretical approaches along this dimension include Cohen and Li (2022), Chew and Wang (2022), Piermont (2022), Piermont and Zuazo-Garin (2022), and de Clippel (2022).

Finally, our paper does not fully answer the question of whether failure to mentally place oneself in a situation that has not yet happened is due to an inability to do so more than a preference for not doing so. For example, we cannot rule out a situation where it is costly to engage in contingent thinking, and people are optimally choosing not to engage in contingent thinking given the costs. Alternatively, people may prefer not to engage in contingent thinking despite negligible costs, for instance because they prefer to keep the "whole picture" in mind when evaluating choices. But, if this were the case, then they could still choose to keep this whole picture in mind even if we ask them to make a choice contingent on a hypothetical event. The finding that they no longer do so in this case suggests that the story that people have difficulty with contingent thinking is more plausible than the story that people prefer not to engage with contingent thinking. While completely settling the issue of mistakes vs. preferences is outside the scope of this paper, we hope that further work pays attention to this issue and the welfare implications of different mental representations (see Nielsen and Rehbeck (2022) for recent work along these lines).

References

AHN, D., S. CHOI, D. GALE, AND S. KARIV (2014): "Estimating ambiguity aversion in a portfolio choice experiment," *Quantitative Economics*, 5, 195–223.

AHN, D. S. AND H. ERGIN (2010): "Framing contingencies," *Econometrica*, 78, 655–695.

AL-NAJJAR, N. I. AND J. WEINSTEIN (2009): "The ambiguity aversion literature: a critical assessment," *Economics and Philosophy*, 25, 249–284.

ALI, S. N., M. MIHM, L. SIGA, AND C. TERGIMAN (2021): "Adverse and Advantageous Selection in the Laboratory," *American Economic Review*, 111, 2152–78.

ALLAIS, M. (1953): “Le comportement de l’homme rationnel devant le risque: critique des postulats et axiomes de l’école américaine,” *Econometrica: Journal of the Econometric Society*, 503–546.

ANDREONI, J., T. SCHMIDT, AND C. SPRENGER (2014): “Measuring Ambiguity Aversion: Experimental Tests of Subjective Expected Utility,” *mimeo*.

ARAUJO, F. A., S. W. WANG, AND A. J. WILSON (2021): “The times they are a-Changing: Dynamic Adverse Selection in the Laboratory,” *American Economic Journal: Microeconomics*, 13, 1–22.

BARRON, K., S. HUCK, AND P. JEHIEL (2019): “Everyday econometricians: Selection neglect and overoptimism when learning from others,” *Working Paper*.

BAYONA, A., J. BRANDTS, AND X. VIVES (2020): “Information frictions and market power: A laboratory study,” *Games and Economic Behavior*, 122, 354–369.

BAZERMAN, M. H. AND W. F. SAMUELSON (1983): “I won the auction but don’t want the prize,” *Journal of Conflict Resolution*, 27, 618–634.

BLAVATSKYY, P., A. ORTMANN, AND V. PANCHENKO (2022): “On the experimental robustness of the Allais paradox,” *American Economic Journal: Microeconomics*, 14, 143–63.

BORDALO, P., N. GENNAIOLI, AND A. SHLEIFER (2012): “Salience theory of choice under risk,” *The Quarterly journal of economics*, 127, 1243–1285.

BRANDTS, J. AND G. CHARNESS (2011): “The strategy versus the direct-response method: a first survey of experimental comparisons,” *Experimental Economics*, 14, 375–398.

CALFORD, E. M. AND T. N. CASON (2021): “Contingent Reasoning and Dynamic Public Goods Provision,” *Working Paper*.

CAMERER, C. (1995): “Individual Decision Making. Chapter 8 in John Kagel and Al Roth (eds.), The Handbook of Experimental Economics,” .

CASON, T. N. AND C. R. PLOTT (2014): “Misconceptions and game form recognition: Challenges to theories of revealed preference and framing,” *Journal of Political Economy*, 122, 1235–1270.

CHARNESS, G. AND D. LEVIN (2009): "The origin of the winner's curse: a laboratory study," *American Economic Journal: Microeconomics*, 1, 207–236.

CHEW, S. H. AND W. WANG (2022): "Generalizing Obvious Dominance using the Sure-thing Principle," *Working Paper*.

COHEN, M., I. GILBOA, J.-Y. JAFFRAY, AND D. SCHMEIDLER (2000): "An experimental study of updating ambiguous beliefs," *Risk Decision and Policy*, 5, 123–133.

COHEN, S. AND S. LI (2022): "Sequential Cursed Equilibrium," *Working Paper*.

CROSETTO, P. AND A. FILIPPIN (2013): "The 'bomb' risk elicitation task," *Journal of Risk and Uncertainty*, 47, 31–65.

CROSON, R. T. (1999): "The disjunction effect and reason-based choice in games," *Organizational Behavior and Human Decision Processes*, 80, 118–133.

CUBITT, R., C. STARMER, AND R. SUGDEN (1998): "Dynamic choice and the common ratio effect: An experimental investigation," *The Economic Journal*, 108, 1362–1380.

CUBITT, R. P. (1996): "Rational dynamic choice and expected utility theory," *Oxford Economic Papers*, 48, 1–19.

DAL BÓ, E., P. DAL BÓ, AND E. EYSTER (2018): "The demand for bad policy when voters underappreciate equilibrium effects," *The Review of Economic Studies*, 85, 964–998.

DE CLIPPEL, G. (2022): "Departures from Preference Maximization, Violations of the Sure-Thing Principle, and Relevant Implications," *Working Paper*.

DEAN, M. AND P. ORTOLEVA (2015): "Is it all connected? a testing ground for unified theories of behavioral economics phenomena," *Working Paper*.

DOMINIAK, A., P. DÜRSCH, AND J.-P. LEFORT (2012): "A dynamic Ellsberg urn experiment," *Games and Economic Behavior*, 75, 625–638.

ELIAZ, K., D. RAY, AND R. RAZIN (2006): "Choice shifts in groups: A decision-theoretic basis," *The American economic review*, 96, 1321–1332.

ELLSBERG, D. (1961): "Risk, ambiguity, and the Savage axioms," *The Quarterly Journal of Economics*, 643–669.

ENKE, B. (2020): "What you see is all there is," *The Quarterly Journal of Economics*, 135, 1363–1398.

ENKE, B. AND F. ZIMMERMANN (2019): "Correlation neglect in belief formation," *The Review of Economic Studies*, 86, 313–332.

ESPONDA, I. AND E. VESPA (2014): "Hypothetical Thinking and Information Extraction in the Laboratory," *American Economic Journal: Microeconomics*, 6, 180–202.

——— (2018): "Endogenous sample selection: A laboratory study," *Quantitative Economics*, 9, 183–216.

EVANS, J. S. B. (2007): *Hypothetical thinking: Dual processes in reasoning and judgement*, vol. 3, Psychology Press.

EYSTER, E. AND M. RABIN (2005): "Cursed equilibrium," *Econometrica*, 73, 1623–1672.

EYSTER, E. AND G. WEIZSÄCKER (2010): "Correlation neglect in financial decision-making," *Working Paper*.

——— (2016): "Correlation neglect in portfolio choice: Lab evidence," *Available at SSRN 2914526*.

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

FREDERICK, S. (2005): "Cognitive reflection and decision making," *The Journal of Economic Perspectives*, 19, 25–42.

FRYDMAN, C. AND M. MORMANN (2016): "The role of salience and attention in choice under risk: An experimental investigation," *Working Paper*.

GABAIX, X. (2014): "A sparsity-based model of bounded rationality," *The Quarterly Journal of Economics*, 129, 1661–1710.

GENNAIOLI, N. AND A. SHLEIFER (2010): "What comes to mind," *The Quarterly journal of economics*, 125, 1399–1433.

GHIRARDATO, P. (2002): "Revisiting Savage in a conditional world," *Economic theory*, 20, 83–92.

GILBOA, I. (2009): *Theory of decision under uncertainty*, vol. 1, Cambridge university press.

GILBOA, I. AND M. MARINACCI (2011): "Ambiguity and the Bayesian paradigm," *Chapter*, 7, 179–242.

GILBOA, I. AND D. SCHMEIDLER (1995): "Case-based decision theory," *The Quarterly Journal of Economics*, 605–639.

GLAZER, J. AND A. RUBINSTEIN (1996): "An Extensive Game as a Guide for Solving a Normal Game," *Journal of Economic Theory*, 1, 32–42.

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

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

HALEVY, Y. (2007): "Ellsberg revisited: An experimental study," *Econometrica*, 75, 503–536.

HAMMOND, P. J. (1988): "Consequentialist foundations for expected utility," *Theory and decision*, 25, 25–78.

HARSTAD, R. M. (2000): "Dominant strategy adoption and bidders' experience with pricing rules," *Experimental economics*, 3, 261–280.

HOLLER, M. J. (1983): "Do economics students choose rationally? A research note." *Social Science Information/sur les sciences sociales*.

HUCK, S. AND W. MÜLLER (2012): "Allais for all: Revisiting the paradox in a large representative sample," *Journal of Risk and Uncertainty*, 44, 261–293.

IVANOV, A., D. LEVIN, AND M. NIEDERLE (2010): "Can relaxation of beliefs rationalize the winner's curse?: an experimental study," *Econometrica*, 78, 1435–1452.

KAGEL, J., R. HARSTAD, AND D. LEVIN (1987): "Information impact and allocation rules in auctions with affiliated private values: A laboratory study," *Econometrica*, 1275–1304.

KAGEL, J. AND D. LEVIN (2002): *Common value auctions and the winner's curse*, Princeton Univ Pr.

KAGEL, J. H. (1995): "Auctions: A survey of experimental research. John H. Kagel, Alvin E. Roth, eds., The Handbook of Experimental Economics," .

KAGEL, J. H. AND D. LEVIN (1993): "Independent private value auctions: Bidder behaviour in first-, second-and third-price auctions with varying numbers of bidders," *The Economic Journal*, 103, 868–879.

KOVÁŘÍK, J., D. LEVIN, AND T. WANG (2016): "Ellsberg paradox: Ambiguity and complexity aversions compared," *Journal of Risk and Uncertainty*, 52, 47–64.

LELAND, J. W. (1998): "Similarity judgments in choice under uncertainty: A reinterpretation of the predictions of regret theory," *Management Science*, 44, 659–672.

LEVIN, D., J. PECK, AND A. IVANOV (2016): "Separating bayesian updating from non-probabilistic reasoning: An experimental investigation," *American Economic Journal: Microeconomics*, 8, 39–60.

LI, S. (2017): "Obviously strategy-proof mechanisms," *American Economic Review*, 107, 3257–87.

LOOMES, G. AND R. SUGDEN (1982): "Regret theory: An alternative theory of rational choice under uncertainty," *The economic journal*, 92, 805–824.

LOUIS, P. (2015): "The barrel of apples game: Contingent thinking, learning from observed actions, and strategic heterogeneity," .

MACCRIMMON, K. R. AND S. LARSSON (1979): "Utility theory: Axioms versus paradoxes," in *Expected utility hypotheses and the Allais paradox*, Springer, 333–409.

MACHINA, M. (2008): "Non-Expected Utility Theory. In Steven Durlauf and Lawrence Blume (eds.), The New Palgrave Dictionary of Economics," .

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

MACHINA, M. J. AND M. SINISCALCHI (2013): "Ambiguity and ambiguity aversion," *Economics of Risk and Uncertainty, Handbook in Economics, North Holland, Machina and Viscusi eds.*

MARTIN, D. AND E. MUNOZ-RODRIGUEZ (2019): "Misperceiving Mechanisms: Imperfect Perception and the Failure to Recognize Dominant Strategies," .

MARTÍNEZ-MARQUINA, A., M. NIEDERLE, AND E. VESPA (2019): "Failures in contingent reasoning: The role of uncertainty," *American Economic Review*, 109, 3437–74.

MOSER, J. (2019): "Hypothetical thinking and the winner's curse: an experimental investigation," *Theory and Decision*, 87, 17–56.

NEBOUT, A. AND D. DUBOIS (2014): "When Allais meets Ulysses: Dynamic axioms and the common ratio effect," *Journal of Risk and Uncertainty*, 48, 19–49.

NGANGOUÉ, M. K. AND G. WEIZSÄCKER (2021): "Learning from unrealized versus realized prices," *American Economic Journal: Microeconomics*, 13, 174–201.

NICKERSON, R. (2015): *Conditional Reasoning: The Unruly Syntactics, Semantics, Thematics, and Pragmatics of "If"*, Oxford University Press.

NIELSEN, K. AND J. REHBECK (2022): "When choices are mistakes," *American Economic Review*, 112, 2237–68.

PIERMONT, E. (2022): "Hypothetical Expected Utility," *Working Paper*.

PIERMONT, E. AND P. ZUAZO-GARIN (2022): "Failures of Contingent Thinking," *Working Paper*.

REES-JONES, A., R. I. SHORRER, AND C. TERGIMAN (2023): "Correlation Neglect in Student-to-School Matching," *Working Paper*.

SAMET, D. (2015): "The sure-thing principle in epistemic terms," *Working Paper*.

SAVAGE, L. J. (1972): *The foundations of statistics*, Courier Corporation.

SHAFIR, E. AND A. TVERSKY (1992): "Thinking through uncertainty: Nonconsequential reasoning and choice," *Cognitive psychology*, 24, 449–474.

SINISCALCHI, M. (2009): "Two out of three ain't bad: A comment on 'the ambiguity aversion literature: A critical assessment'," *Economics and Philosophy*, 25, 335–356.

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

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

STARMER, C. AND R. SUGDEN (1993): "Testing for juxtaposition and event-splitting effects," *Journal of risk and uncertainty*, 6, 235–254.

TVERSKY, A. AND D. KAHNEMAN (1981): “The framing of decisions and the psychology of choice,” *Science*, 211, 453–458.

——— (1986): “Rational choice and the framing of decisions,” *Journal of business*, S251–S278.

WAKKER, P. P. (2001): “Testing and characterizing properties of nonadditive measures through violations of the sure-thing principle,” *Econometrica*, 69, 1039–1059.

WASON, P. C. (1966): “Reasoning,” *New horizons in psychology, Vol. 1, BM Foss ed.*

——— (1968): “Reasoning about a rule,” *The Quarterly journal of experimental psychology*, 20, 273–281.

ZHANG, L. AND D. LEVIN (2017): “Partition Obvious Preference and Mechanism Design: Theory and Experiment,” *Working Paper*.