Humphrey, Steven J.; Kruse, Nadia-Yasmine

Article — Published Version

Who accepts Savage's axiom now?

Theory and Decision

Provided in Cooperation with:
Springer Nature

Suggested Citation: Humphrey, Steven J.; Kruse, Nadia-Yasmine (2023) : Who accepts Savage's axiom now?, Theory and Decision, ISSN 1573-7187, Springer US, New York, NY, Vol. 96, Iss. 1, pp. 1-17, https://doi.org/10.1007/s11238-023-09938-8

This Version is available at:
https://hdl.handle.net/10419/307090

Theory and Decision (2024) 96:1–17
https://doi.org/10.1007/s11238-023-09938-8

Who accepts Savage's axiom now?

Steven J. Humphrey[1] · **Nadia-Yasmine Kruse**[1]

Accepted: 2 April 2023 / Published online: 20 June 2023

Abstract
We report the results of an experimental test of whether preaching the normative appeal of the sure-thing principle leads decision-makers to make choices that satisfy it. We use Allais-type decision problems to observe the incentive-compatible choices of 147 subjects, which either violate the sure-thing principle or adhere to it. Subjects are presented with normative arguments that support the counterfactual behaviour and then repeat their decisions. We observe violations of the sure-thing principle are robust to its normative justification. This result replicates a famous small-sample observation using hypothetical tasks that was published by Paul Slovic and Amos Tversky almost half a century ago. We argue that this finding is as relevant now as it was then and that their design can be usefully applied to address contemporary issues in behavioural economics.

Keywords Expected utility theory · Sure-thing principle · Allais Paradox · Normative decision theory

1 Introduction

Whether violations of the *sure-thing principle* are erroneous deviations from an otherwise sound axiom is a matter of long-standing debate which originates in the famous exchanges between Maurice Allais (1953a, b) and Leonard Savage (1954). It is almost half a century since Slovic and Tversky (1974) reported the results of their classic experimental test designed to resolve this dispute. Subjects responded to a version of Allais' (1953a) common consequence effect decision problems (henceforth *CCE* or *Allais paradox*), in which they could either conform to the sure-thing

✉ Steven J. Humphrey
steve.humphrey@uos.de

Nadia-Yasmine Kruse
nadia.yasmine.kruse@uos.de

[1] School of Business Administration and Economics, University of Osnabrück, Rolandstr. 8, 49069 Osnabrück, Germany

principle or violate it. They were then presented with normative arguments favouring the counterfactual behaviour, before being asked to repeat the decision tasks. Violations of the axiom persisted as the modal pattern of behaviour. This observation has been interpreted as undermining the defense of the sure-thing principle as an attainable normative standard that might sometimes be mistakenly violated (e.g. Dietrich et al., 2021).

Although Slovic and Tversky's result provides putative support for Allais' view that violating the axiom is reasonable, the debate remains unresolved. There are several reasons why. Firstly, Slovic and Tversky's findings are subject to the two caveats that choices were hypothetical and the sample size of 29 precluded meaningful statistical analysis (van de Kuilen & Wakker, 2006, p. 158; Dietrich et al., 2021, p. 161; Nielsen & Rehbeck, 2022, p. 2238, fn. 3). Secondly, recent evidence supports Savage's view rather than Allais'. Nielsen and Rehbeck (2022) observe a systematic tendency for decision-makers who violated the sure-thing principle in lottery choices to correct those violations when given the opportunity. Thirdly, there is a general lack of evidence that speaks to the normative force of the sure-thing principle. This is a conspicuous gap in the literature considering the historical significance of the debate and the widely held view that the normativity of choice principles should be settled empirically (e.g. Tversky & Kahneman, 1986, p. S273; Gilboa, 2010, p. 4; Sunstein, 2018, p. 2).[1]

In this paper, we report the results of an experiment designed to address these issues by investigating the reliability of Slovic and Tversky's result. We do so by embedding their test in a modern design that observes the incentive-compatible choices of 147 subjects. This extends Slovic and Tversky's study by enabling an individual-level statistical analysis of the normative force of both Allais' and Savage's arguments. Whilst this contribution is modest in terms of innovation, it is nonetheless important for the following reasons, which we introduce here and elaborate in the next section.

Firstly, the Allais paradox has played a central role in the evolution of risky choice theory (van de Kuilen & Wakker, 2006; Mongin, 2019). By speaking to both the descriptive and normative performance of the sure-thing principle, Slovic and Tversky's result has occupied a significant position in that programme of research. This warrants establishing its reliability. Doing so also matters for the validation of conclusions in the broader literature that are based upon it (e.g. Dietrich et al., 2021, p. 161).

Secondly, the replicability of Slovic and Tversky's result has an important bearing on the implications of Nielsen and Rehbeck's (2022) findings. If Slovic and Tversky's result is robust, then the evidence is mixed. In this case, the pertinent research question is a nuanced inquiry into the factors which influence the normative appeal of the sure-thing principle to decision-makers, rather than simply establishing whether it has normative appeal. Differences between our approach and that followed

[1] At the time of writing, Google Scholar reports that Slovic and Tversky has been cited approximately 850 times, including 21 times in the last year. This impressive citation record attests to the demand for this type of work. Baron (2004) is a notable dissenter, who argues that normative theory should be determined solely through philosophical debate. If normative theory is based upon evidence, that risks undermining its role as a benchmark relative to which behavioural biases can fixed.

by Nielsen and Rehbeck (2022) are relevant to this question, and we discuss these in the next section.

Thirdly, Slovic and Tversky's design was ahead of its time in addressing significant issues in behavioural economics which post-date its publication. For example, public policy involving behavioural nudges is often defended against accusations of undue paternalism on grounds of helping decision-makers avoid choices that they themselves would consider to be mistakes. If Slovic and Tversky's result is robust to modern experimental methods, it would imply that no such defense could be mounted for nudges that encourage conformity with the sure-thing principle.[2]

Our data reveal that Slovic and Tversky's result is robust. The normative arguments yield no overall tendency for decision-makers to conform to the sure-thing principle. In fact, they distill violations of the sure-thing principle such that they become highly patterned in the direction consistent with Allais' (1953a) rationalization. This reveals a systematic Allais paradox in post-argument choices that was not present in pre-argument choices. Therefore, just as was the case in Slovic and Tversky, our data indicate that a sizeable proportion of decision-makers do not accept Savage's axiom.

2 Background

At the May 1952 Paris Symposium on the *Foundations and Applications of the Theory of Risk-Bearing*, Leonard Savage famously violated the sure-thing principle of his own subjective expected utility theory (EUT) in hypothetical decision problems presented to him by Maurice Allais (see Allais, 1953a). Upon reflection, Savage changed his choices to conform with the sure-thing principle, and concluded that he had corrected what had been an erroneous decision (Savage, 1954, p. 103). The debate ignited by the Paris Symposium included, in the following December, the publication of Friedman and Savage's (1952) description of EUT as an empirical hypothesis. In this paper, they make the thought-provoking claim that the hypothesis can be partly evaluated against indirect evidence, which can include the normative appeal of choice axioms. They argue that the introspective appeal of choice axioms means that they would not be deliberately violated, hence doing so would be a mistake. Over half a century later, Starmer (2005) took issue with this claim on grounds of it being incomplete. His reasoning is as follows.

First, assume that a decision-maker finds the sure-thing principle introspectively appealing and accepts that they would not deliberately violate it. Second, consider the proposition that the decision-maker will probably respect the sure-thing principle. The assumption is a statement about a normative judgement and the proposition is about behaviour. Observing the proposition to be true would only validate the assumption if there is a premise that links them (e.g. the decision-maker will rarely violate the sure-thing principle, because they believe it should be respected), *and* that premise is supported by empirical evidence. Both are absent from Friedman and Savage's (1952) argument.

[2] This issue is discussed in Sugden (2016, 2018) and Sunstein (2018).

The implication of Starmer's (2005) critique is that the normativity of a choice axiom cannot be evaluated by evidence unless it is generated under conditions that link choices to underlying normative judgements. Slovic and Tversky's test satisfies this criterion in that the repeated choices embody subjects' reflections on their initial decisions, as well as the normative logic of the sure-thing principle and Allais' (1953a) rationalization of its violation. This point is echoed by Gilboa (2010), who similarly argues that the rationality of decision-making cannot be assessed by merely observing choices. According to Gilboa (2010), the rationality of behaviour can be established by confronting decision-makers with their choices to discover attitudes towards them, such as whether they are regretted.[3] On Gilboa's (2010) view, the persistence of sure-thing principle violations observed by Slovic and Tversky counts as evidence that the axiom is not a requirement of rationality (Dietrich et al., 2021, p. 161).

A different approach to connecting normative judgements with choices is reported by Nielsen and Rehbeck (2022). Whereas Slovic and Tversky, MacCrimmon (1968) and Moskovitz (1974) study how axioms apply in specific decision situations, Nielsen and Rehbeck (2022) employ an incentive-compatible procedure to elicit subjects' preferences over axioms they would like to apply to all choices. Subjects then make choices in situations with the potential for the axioms to be violated, before being confronted with each situation where their preferences over axioms and choices were inconsistent. Finally, they are given the opportunity to reconcile inconsistencies by changing their choices or deselecting the axiom.[4] Nielsen and Rehbeck (2022) observe that 83% of subjects wanted choices to conform to the sure-thing principle, which they attribute to the axiom's normative appeal, but 75% of these subjects violated it in subsequent decision tasks. Of these inconsistent subjects, 16% reconciled the conflict by deselecting the axiom and 34% changed their choices. This leads Nielsen and Rehbeck (2022) to a conclusion that echoes Savage (1954). In their view, the sure-thing principle has normative content and violations are mainly choice errors.

Nielsen and Rehbeck (2022) describe their procedure as being conducive to general conclusions, because it does not entail an explanation of axioms in the context of specific decision problems. They acknowledge that this approach foregoes the simplicity and clarity of the approach employed by Slovic and Tversky, where the normative arguments explain how abstract axioms apply to actual choices. This may matter for two reasons. Firstly, it is conceivable that the clarity of context-specificity will influence conformity with an axiom. Secondly, it is arguably conducive to a precise interpretation of observed behaviour. For example, Gilboa (2010) describes a situation where an individual who persists in committing the Allais paradox is considered to be doing so rationally, even though their persistence stems not from a considered rejection of the sure-thing principle's claim to normativity, but because they did not understand the explanation of it. For these reasons, the data we report in

[3] Without explicitly mentioning Slovic and Tversky, Gilboa (2010, p4, fn.3) notes how this might be done by describing their design.

[4] Subjects could also do nothing.

this paper should be considered complementary to those reported by Nielsen and Rehbeck (2022), with the potential to shed light on these matters.[5]

Nielsen and Rehbeck (2022) also argue that their approach it is less susceptible to experimenter demand effects, because the 'intervention' involves confronting subjects with inconsistencies in their own decisions, rather than an inconsistency between their choices and an argument presented to them by an 'authority' (e.g. the experimenter). However, assuming potential demand effects can be controlled (or their influence otherwise understood), 'intervention by an authority' is a desirable design feature in some situations pertinent to the study of whether decision-makers adhere to principles of rationality.

For example, behavioural nudges are an 'intervention by an authority' which feature prominently in the behavioural welfare economics literature (see Thaler & Sunstein, 2008; Sugden, 2016, 2018; Sunstein, 2018). As mentioned in the introduction, nudges can be defended against accusations of undue paternalism if decision-makers themselves want to be nudged. For instance, to avoid decisions that upon reflection they would accept were mistakes. Sunstein (2018, p. 2) recommends that the matter of whether decision-makers want to be nudged should be settled empirically, and Sugden (2018, p. 12) describes the normative arguments in Slovic and Tversky as examples of a "classic" behavioural nudge.

Gilboa (2010) also believes that intervention is an appropriate response to decisions that fail to respect choice axioms. This viewpoint stems from Kahneman and Tversky (1979, p. 277) noting that axiom violations are to be expected when decision-makers have no opportunity to discover that they have violated principles they would rather respect. Gilboa (2010), therefore, considers it reasonable to preach classical decision theory to help individuals make better choices. On this view, it would be premature to evaluate the normativity of a choice theory unless it has been established whether that theory can be successfully preached (c.f. Dietrich et. al., 2021, p. 146). The argument that favours the sure-thing principle in Slovic and Tversky is a straightforward example Gilboa's (2010) preaching.[6]

Finally in this section, we note that Slovic and Tversky's design speaks to literatures besides those we have discussed, including those on preference purification and learning in the Allais paradox. In terms of the latter, reflection on normative arguments constitutes an opportunity for learning by thought, as opposed to learning by experience (van de Kuilen & Wakker, 2006).[7]

[5] Nielsen and Rehbeck's (2022) design is probably best described as being more efficient than the context-specific approach, rather than being more general. The application of Slovic and Tversky's method to different decision tasks and different axioms would also facilitate general conclusions. For example, Humphrey and Meickmann (2022) apply it to the stationarity principle of the discounted expected utility model, and in a follow-up study to this one (which we hope to report soon), the present authors apply it to the common ratio effect test of the sure-thing principle.

[6] Like Sunstein (2018), Gilboa (2010) is of the view that assessing the rationality of behaviour is an empirical matter, because what counts as rational action may depend on the population being studied.

[7] The preference purification literature typically assumes that error-free preferences are described by standard models of rationality (Hausman, 2012; Infante et al., 2016). Violations of choice axioms are considered to be mistakes, and a normative criterion is taken to be the satisfaction of preferences that have been reconstructed to be free of the psychological mechanisms that gave rise to those mistakes. Whilst

3 Experiment

A total of 147 subjects (62% female) were recruited using ORSEE (Greiner, 2015) from a database of volunteers at the Laboratory for Economics Research (LaER) at the University of Osnabrück. Subjects participated in one of six pre-arranged sessions which lasted for up to 45 min. Each session took place online in a BigBlueButton virtual meeting room, accessed through the university's online portal (StudIP).[8] The experiment was programmed in SoPHIE (Hendriks, 2012). Subjects were informed that payments were underwritten by a research fund held at the university and would be made immediately after the experiment by the finance department. The random lottery incentive system was applied to a set of 32 decision problems, yielding an average payment of € 13.05 (including a show-up fee of € 5.00).

Figure 1 illustrates the decision problems used to test the sure-thing principle. The problems were presented using a strip display with the probabilities of each outcome represented by lottery tickets running from 1 to 100. Each problem is a choice between a riskier (R, R') and a safer lottery (S, S'). In each problem, lottery tickets 26–100 yield the same outcome irrespective of the lottery chosen. The sure-thing principle holds that these events are irrelevant to the decision. Therefore, if indifference is disallowed, expected utility maximization requires that either R and R' are chosen (denoted RR') or S and S' are chosen (denoted SS'). The violation of the sure-thing principle that Allais (1953a) described as being reasonable is SR'. A systematic Allais paradox is observed if there are significantly more SR' choices than RS' choices.

Following instructions (contained in the Appendix), subjects completed 30 choice tasks in random order. These comprised the problems in Fig. 1 and 28 other problems, which we discuss below. Subjects were then shown the two problems in Fig. 1 alongside how they had chosen in those tasks. If subjects had violated the sure-thing principle (SR' or RS'), they were shown an argument in favour of adhering to it ('Savage's position'). If subjects had respected the sure-thing principle (RR' or SS'), they were shown an argument supporting Allais' (1953a) rationalization of SR' choices ('Allais' position'). The arguments are described in Fig. 2. Subjects were reminded of the random-lottery incentive system and that they should choose according to their preferences, before repeating the two decisions.

The other 28 decision problems were taken from six well-known studies that report significant violations of EUT.[9] These tasks were used as an 'instrument check' of our procedures. Although the procedures were as close as practicably possible to those employed in the laboratory, it remains possible that the online implementation induced differences in behaviour relative to that observed in conventional

Footnote 7 continued
related to this literature, our study is not an endorsement of the assumption that error-free preferences are described by standard models.

[8] Subjects were required to activate their microphone and speakers, but not their camera.

[9] Specifically, common consequence effects (Starmer, 1992; Starmer and Sugden, 1991), common ratio effects (van de Kuilen and Wakker, 2006), violations of transitivity consistent with prospect theory (Starmer, 1999) and regret theory (Loomes et al., 1991), and event-splitting effects (Humphrey, 1995).

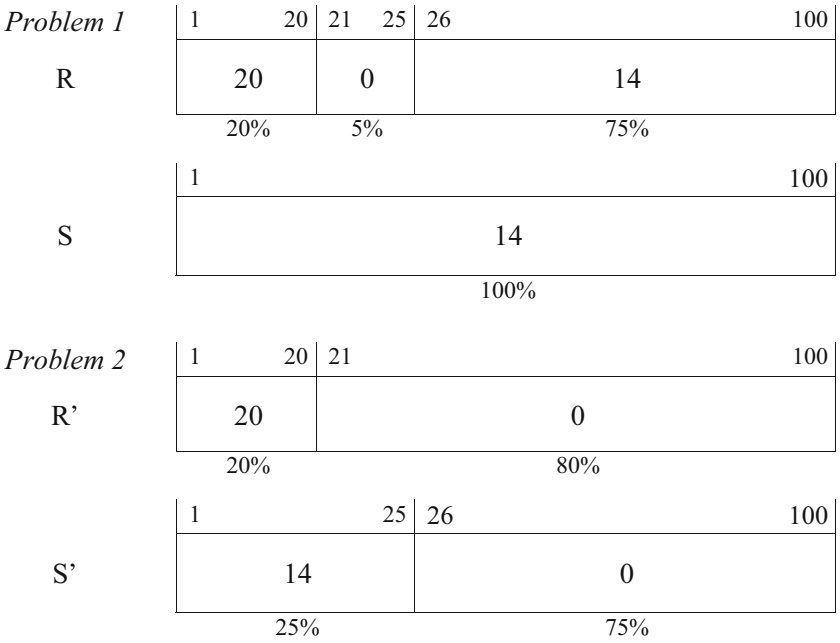


Fig. 1 Common consequence decision problems

Savage's Position:

"*In problem 1, if a ticket between 26 and 100 is drawn it does not matter whether I choose R or S. I would win €14 irrespectively of how I choose. So I will ignore tickets 26 to 100. In problem 2 the same is true. If a ticket between 26 and 100 is drawn, it doesn't matter whether I choose R' or S'. I win nothing, irrespectively of how I choose. So I will ignore lottery tickets 26 to 100. Therefore, because I should always ignore tickets 26 to 100, the problems are exactly the same. In both problems, tickets 1 to 25 in options S and S' always pay me €14. In both problems, tickets 1 to 20 in options R and R' always pay €20, and tickets 21 to 25 pay nothing. It therefore makes no sense to switch choices between the problems. So I would choose either options R and R', or options S and S'.*"

Allais' Position:

"I would choose option S over option R in problem 1 and R' over S' in problem 2. In problem 1, I have the choice between €14 for sure or a gamble where I might end up with nothing. Why gamble? The small probability of missing the chance of winning something seems very unattractive to me. In problem 2, there is a good chance that I will end up with nothing no matter how I choose. The chances of getting €20 are almost as good as getting €14, so I might as well go for the €20 and choose option R' over S'."

Fig. 2 Normative arguments

experiments. Each of the six studies entailed large-sample experiments with incentive-compatible decision tasks, but none were implemented online. Therefore, if we are able to replicate the violations of EUT they report, it would provide

reassurance that observed choices were not induced by the online procedures. This is borne-out by the data. The 28 'instrument check' problems facilitated 12 tests of EUT. Of these, 6 employed the parameters used in the original experiments, and 6 used outcomes adjusted to account for inflation. Significant violations of EUT consistent with those reported in the original literature were observed in 9 of the 12 tests, including all 6 of those with inflation-adjusted outcomes.[10]

4 Results

Table 1 reports aggregate behaviour in pre- and post-argument choices. In pre-argument choices, 59% of subjects respected the sure-thing principle (RR' and SS'). Of the 41% who did not, violations are broadly similarly distributed between SR' and RS' choices. Hence, we do not observe a systematic Allais paradox. This finding is consistent with the empirical fragility of the Allais Paradox discussed by Blavatskyy et al. (2022). In the post-argument data, violations of the sure-thing principle remain stable at around 40–41% of behaviour, but are highly patterned. Whereas 50 subjects violate the sure-thing principle by making choices consistent with Allais' position (SR'), only 11 violate it in the opposite direction (RS'). A two-tailed binomial test of the null hypothesis that each type of violation is equally likely is rejected with significance greater than 1% ($p=0.000$). In the post-argument data, we observe a significant Allais paradox.

The data in Table 1, therefore, replicate Slovic and Tversky's central finding in a large sample of incentivized decision-makers. Reflection upon initial decisions in light of normative arguments favouring counterfactual choices, does not purge aggregate behaviour of violations of the sure-thing principle. Moreover, a particularly interesting feature of our data is that the normative arguments introduce a systematic Allais paradox into post-argument choices that did not exist in pre-argument choices. The origins of this observation can be understood by analyzing how the normative arguments influenced behaviour at the individual level. This is reported in Table 2.

Each of the three panels of Table 2 contains a contingency table that shows how a category of individual behaviour was affected by the normative arguments. The categories are expected utility maximizing choices (respecting the sure-thing principle and manifest in RR' or SS'), the CCE (in line with Allais' position and entailing SR' choices), and the counter-CCE (violating the sure-thing principle in the opposite direction with RS' choices).

The leading diagonal of Panel 1 in Table 2 shows decisions that were unchanged following the normative arguments: 54 (36.7%) subjects respected the sure-thing principle in both decisions, whereas 28 (19%) subjects consistently violated it. Decisions that changed following the normative arguments are reported in the off-leading diagonal. The sure-thing principle was initially violated by 32 (21.8%) subjects who subsequently respected it following the argument in favour of doing so. In contrast, 33 (22.4%) subjects initially respected the sure-thing principle, but

[10] The 28 decision problems and choice data can be found in the supplementary files. We were unable to compare choices elicited online with a contemporaneous laboratory experiment, because the experiment was conducted during the COVID-19 pandemic lockdown.

Table 1 Choice patterns

	Pre-argument	Post-argument
RR′	57 (38.8%)	55 (37.4%)
SS′	30 (20.4%)	31 (21.1%)
RS′	27 (18.4%)	11 (7.5%)
SR′	33 (22.4%)	50 (34.0%)
p-value	0.519	0.000*

An asterisk denotes a significant Allais paradox

violated it following the argument supporting Allais' position. A two-sided McNemar's (chi-square) test of the null hypothesis that behaviour is the same in pre- and post-argument decisions yields $p=0.901$. This confirms the aggregate-level finding in Table 1. The normative arguments do not lead to a systematic change in the overall level of compliance with the sure-thing principle.

Panel 2 in Table 2 reports the number of subjects who switched from committing the Allais paradox to respecting the sure-thing principle (or vice versa) between pre- and post-argument choices. The leading diagonal shows that 17 (11.6%) subjects consistently committed the Allais paradox and 54 (36.7%) subjects consistently respected the sure-thing principle across both sets of choices. The off-leading diagonal shows that 31 (21.1%) subjects switched from respecting the sure-thing principle to committing the Allais paradox following the argument in favour of doing so. However, only 16 (10.9%) subjects switched in the opposite direction. A two-sided McNemar's test of the null hypothesis that behaviour is the same in pre- and post-argument decisions is rejected with 5% significance ($p=0.029$). In these data, the argument supporting Allais' position is systematically more persuasive than the argument supporting Savage's position.

Panel 3 in Table 2 shows that 16 (10.9%) subjects switched from committing the counter-CCE to respecting the sure-thing principle following the argument supporting the latter. The small number of subjects (2) who switched in the opposite direction is consistent with expectations, because subjects were not presented with an argument supporting the counter-CCE (and we are aware of no such argument).[11] The McNemar's test rejects the null hypothesis that behaviour is the same in pre- and post-argument decisions with significance greater than 1% ($p=0.001$). Whilst this result is uninformative in terms of evaluating the relative normative content of Savage's and Allais' positions, it contributes to explaining the aggregate data in Table 1. Despite the support for Allais' position manifest in Panel 2 of Table 2, the normative arguments leave overall compliance with the sure-thing principle unchanged, because Savage's position is persuasive for the majority of subjects who committed the counter-CCE in initial decisions.[12]

[11] The literature occasionally reports observations of significant counter-CCEs (e.g. Starmer, 1992; Humphrey, 2000; see Blavatskyy et al., 2022 for a review).

[12] On the assumption that both Allais' and Savage's positions have normative force, it is perhaps unsurprising that a systematic Allais paradox is observed in post-argument choices. Presenting arguments in support of those positions, but not in support of the counter-CCE, can favour the emergence of a

Table 2 Individual-level contingency tables

Panel 1: expected utility theory (EUT)

		Post-argument EUT	Violation	Totals
Pre-argument	EUT	54 (36.7%)	33 (22.4%)	87 (59.2%)
	Violation	32 (21.8%)	28 (19.0%)	60 (40.8%)
	Totals	86 (58.5%)	61 (41.5%)	147 (100%)

p-value=0.901

Panel 2: EUT violation: the common consequence effect (CCE)

		Post-argument CCE	EUT	Totals
Pre-argument	CCE	17 (11.6%)	16 (10.9%)	33 (22.4%)
	EUT	31 (21.1%)	54 (36.7%)	85 (57.8%)
	Totals	48 (32.7%)	70 (47.6%)	118 (80.3%)

p-value=0.029*

Panel 3: EUT violation: counter-common consequence effect (counter-CCE)

		Post-argument Counter-CCE	EUT	Totals
Pre-argument	Counter-CCE	9 (6.1%)	16 (10.9%)	25 (17.0%)
	EUT	2 (1.4%)	54 (36.7%)	56 (38.1%)
	Totals	11 (7.5%)	70 (47.6%)	81 (55.1%)

p-value=0.001**

Percentages in parenthesis are relative to the total sample of 147. There were 2 subjects who switched from counter-CCE violations of the sure-thing principle to the CCE, and so not appear in the latter two contingency tables. An asterisk denotes significance at the 5% level; two asterisks denote significance at the 1% level

As discussed above, experiments involving opportunities to revisit decisions following an intervention may be susceptible to experimenter demand effects. For example, subjects might switch to conform with a normative argument simply because they do not want to appear resistant to advice, and not because doing so would satisfy their preferences. Incentivizing the revelation of genuine preferences goes some way towards controlling for demand effects, but there remains uncertainty over whether this is completely effective. Two reasons provide reassurance that demand effects of this type do not explain our data.

Firstly, if an experimenter demand effect is defined as switching purely to conform with the argument presented by the experimenter, then switching rates should be

Footnote 12 continued
systematic Allais paradox in the way that we observe. The purpose of the experiment was to evaluate that assumption; there was no a priori presumption that either argument would be normatively persuasive.

independent of the argument seen. According to the data in Table 2, this is not the case. Panel 1 shows that 87 subjects respected the sure-thing principle in initial decisions, and so were presented with the argument supporting Allais' position. Of these, Panel 2 shows that 31 (31/87=35.6%) switched to conform with that position. Likewise, Panel 1 shows that 60 subjects initially violated the sure-thing principle, and therefore saw the argument supporting Savage's position. Of these, Panel 1 shows that 32 (32/60=53.3%) switched to conform with it. On the basis of a two-tailed test of difference in sample proportions, the latter switching rate is significantly greater than former at the 5% level (p=0.0332).

Secondly, Kruse (2022) reports evidence that post-argument decisions in experiments of this type are not susceptible to demand effects. In one treatment, subjects were presented with a single argument supporting the counterfactual behaviour. As in our experiment, the argument was determined by initial choices and supported either respecting EUT or violating it. In another treatment, subjects saw both arguments irrespective of their initial choices. The decision to switch choices following the arguments was found to be independent of treatment. This finding suggests that when subjects are presented with a single argument, they do not simply follow the advice contained in that intervention.[13]

5 Conclusion

Almost half a century ago, Slovic and Tversky asked, 'who accepts Savage's axiom?' They found that a stable proportion of their subjects did not, even after its normative logic had been explained. We also observe that a stable proportion of subjects consistently commit the Allais paradox (and is approximately a third of the size of the proportion who consistently respect the sure-thing principle). Amongst the group of subjects who changed their decisions following the normative arguments, switching from respecting the sure-thing principle to committing the Allais paradox is approximately twice as likely as switching in the opposite direction. This contributes to the Allais paradox in post-argument decisions that establishes the robustness of Slovic and Tversky's key finding. Our data indicate that Maurice Allais was correct: it can be reasonable to violate the sure-thing principle.

This conclusion has important implications for the theory of risky choice. The historical significance of the Allais paradox derives partly from the central role it has played in shaping the programme of generalizing EUT. This programme has produced influential theories such as prospect theory (Kahneman & Tversky, 1979) and rank-dependent theory (Quiggin, 1982; Tversky & Kahneman, 1992). Our results suggest that a major part of the evidential basis for those theories is reliable in that it is robust to Savage's argument. A notable feature of our data is that a Savage-type correction of choice errors may actually engender a systematic violation of the

[13] Analogous results are reported by Nielsen and Rehbeck (2022). They elicit subjects' preferences for 'standard' axioms, which contain normative appeal, and for 'control' axioms which do not. Subjects are significantly less inclined to delegate decisions to the control axioms than the standard axioms. This observation is interpreted as showing that subjects make choices based on the meaning of the axioms, and do not express a preference for an axiom simply because this is what the experimenter has asked them to do.

sure-thing principle. The Allais paradox we observe in post-argument choices is partially attributable to the role played by Savage's argument in reducing counter-CCE violations of the principle to 7.5% of behaviour.

Our experiment has also shed light on Gilboa's (2010) view that the test of the normative status of a canonical axiom is whether it can be successfully preached (or, equivalently in this case, whether subjects can be nudged into conformity). Our data suggest that the sure-thing principle can be successfully preached to some subjects. However, since it is also possible to successfully preach Allais' position, preaching per se does not increase overall respect for the sure-thing principle. Analogous findings can be found elsewhere in the literature. For example, Cox and Grether (1996) report evidence which shows that violations of EUT disappear in market-like experiments that expose decision-makers to the discipline of competition and feedback. They interpret this finding as supporting Plott's (1996) discovered preference hypothesis. However, Braga et al. (2009) show that market experience per se will not purge behaviour of violations of canonical choice axioms. Rather, it depends on the nature of market experience, and this can operate to introduce axiom violations into decisions that did not previously exist. The same can be said of preaching: in post-argument choices, there is a significant violation of the sure-thing principle that did not exist in pre-argument choices.

In terms of future research, there are two natural avenues down which to proceed. Firstly, our findings should be subject to robustness checks. Whilst we have used experimental control to select Allais-type problems that have the characteristics of Slovic and Tversky's problems, it is nevertheless the case that we too have used a single parameter set. Different parameterizations of Allais-type problems are known to influence behaviour in one-shot experiments (e.g. see Blavatskyy et al., 2022). We note, however, that results similar to those reported here are observed in an analogous test of the normative appeal of the stationarity principle of the discounted expected utility model (Humphrey & Meickmann, 2022). The connections between the Allais paradox and present-biased violations of the stationarity principle are well known (e.g. Andreoni & Sprenger, 2012).

Secondly, there are sufficient differences between our findings and those of Nielsen and Rehbeck (2022) to support their call for a controlled comparison of how attitudes towards canonical choice axioms are influenced by the manner in which their normative properties are explained. One line of inquiry would be whether their general explanation of the sure-thing principle contributes to the increase in overall conformity with the axiom following the revision of choices, which they observe and we do not. Another line of inquiry might focus on whether compliance with axioms depends on whether counterfactual choices have normative appeal. Our data suggest that this may be important. The absence of an overall increase in conformity with the sure-thing principle is attributable to the relative normative appeal of the Allais paradox. However, when violations of the sure-thing principle have no obvious normative justification, as is the case for the counter-CCE, subjects behave as if they regard those violations to be mistakes worthy of correction. If violations of the axioms studied by Nielsen and Rehbeck (2022) were devoid of normative justification, this could explain our contrasting conclusions.

Appendix: Instructions and procedures

During the recruitment process, subjects were informed that they would need to provide an IBAN, because payment would be made by bank transfer. Subjects were also informed that participation required a device capable of web browsing and with an audio outlet. After registrations had closed, subjects were provided with a link to a BigBlueButton online meeting which corresponded to their session of the experiment.

In the online meeting room subjects were able to see the experimenter, but no other participant. A link to the experiment software and a participation code were provided to each subject individually, via a chat function. Subjects were then asked to log into the experiment software and await further instructions.

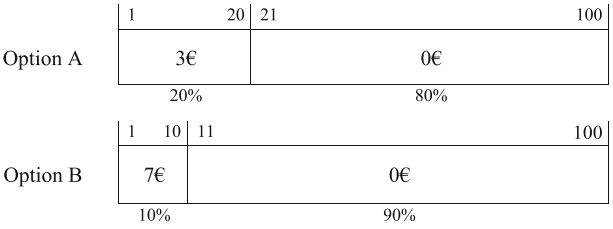
After all subjects had logged-in, they were informed that instructions would be read aloud by the experimenter and simultaneously shown on their screens. Questions could be asked at any time by sending a message to the experimenter via a chat function.

Instructions

Welcome to the experiment and thank you for participating. Before we begin, we would like to provide you with some general information about the procedures in the experiment. The experiment does not involve any "tricks", there are no right or wrong answers, and the experiment is not a "test". We are simply interested in finding out how people make choices in particular situations, and we are prepared to pay you to find out.

All decisions will be made anonymously. No other participant will know your choices and you will not know the choices of other participants. All participants will have their payment transferred to their bank account by the finance department of the university immediately after the experiment. The funds from which you will be paid are from a research account held at the university. You can therefore be sure that you will be paid for your participation.

In the experiment you will be asked to make 32 choices between pairs of lotteries, called Option A and Option B. Each lottery offers real money prizes with an associated probability. The outcome that you can receive from a choice task depends on the choice that you make, and on a number between 1 and 100 that will be randomly and individually determined for each participant. The following question is an example.



❒ I choose Option A

❒ I choose Option B

In this example, if you prefer Option A, you should indicate your choice by ticking the box next to "I choose Option A". If you prefer Option B, you should tick the box next to "I choose Option B".

We will use this example to explain how your payment for participating in the experiment will be determined. Firstly, everyone receives a payment of 5 € just for participating. This does not depend on your decisions. In addition to this, you will receive the outcome of the lottery you chose in one of the 32 choice tasks. You will only find out which of the 32 choice tasks this is after you have made all 32 choices.

When all choices have been made, the computer will generate a random number between 1 and 32, individually for each participant. You will see this happening in real time on your computer screen. The randomization is fair and we have no influence over it. The randomly selected question number is indicated in a "payment question field". This question will determine your payment. For example, if the number 6 is randomly chosen, the option you chose in question 6 will determine your payment.

When your payment question has been chosen, the computer will generate another random number, individually for each participant, this time between 1 and 100. The randomization is fair and we have no influence over it. The number will be shown in a "lottery ticket field" on your computer screen. This number determines the outcome of the lottery you chose in your payment question and therefore the payment you receive.

Imagine that the above example was randomly selected as your payment question. If you chose option A in this problem, the outcome of that lottery is your payment. If the random number selected from 1 to 100 was in the range from 1 to 20, your payment would be 3€. If the random number selected from 1 to 100 was in the range from 21 to 100, your payment would be zero.

Since any of the 32 tasks could be the one that determines your payment, and you will not know which task this is until the end of the experiment, you should answer all 32 questions as if they are for real money (because one of them will be).

Since real money is involved, please consider your choices carefully and make them according to your preferences.

Following an opportunity for questions, subjects completed the 30 decision tasks (28 "instrument check" tasks plus the two tasks in Fig. 1*). When the 30 decision tasks had been completed, the software proceeded to the "Arguments screen".*

Arguments screen

We would like you to reflect on two of the choices you have made in light of how some economists, who are interested in these types of situations, would interpret your behaviour. You will then be asked to make those choices again. You are under no obligation to behave as the economists suggest and there are no right or wrong answers. Since real money is at stake, please consider your choices carefully and decide in the way that you prefer.

The decision problems in Fig. 1 *were shown on the computer screen along with how the subject had chosen in those problems. The appropriate normative argument was then presented, as described in the main text. Subjects then repeated the decisions.*

After the repeated choices, subjects completed a short demographic questionnaire regarding gender, age and their studies. They were then informed of their payoff-relevant decision situation and their resulting payment. Finally, subjects entered their IBAN number and were informed that the experimenters had no access to this information. Subjects were thanked and the experiment ended.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval Our research complies with ethical requirements for decision-making experiments involving human subjects at the Laboratory for Economics Research (LaER) at The University of Osnabrück.

Informed consent All subject participated having given informed consent.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11238-023-09938-8.

References

Allais, M. (1953a). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'ecole Americaine [Rational man's behavior in the presence of risk: Critique of the postulates and axioms of the American school]. *Econometrica, 21*, 503–546. https://doi.org/10.2307/1907921

Allais, M. (1953b). Fondements d'une théorie positive des choix comportant un risqué et critique des postulats et axioms de l'ecole Americaine. *Econom*étrie, *Colloques Internationaux de CNRS,* XL, 257–232.

Andreoni, J., & Sprenger, C. (2012). Risk preferences are not time preferences. *American Economic Review, 102*, 3357–3376. https://doi.org/10.1257/aer.102.7.3357

Baron, J. (2004). Normative models of judgment and decision making. In D. J. Koehler & N. Harvey (Eds.), *Blackwell Handbook of Judgment and Decision Making* (pp. 19–36). Blackwell.

Blavatskyy, P., Ortmann, A., & Panchenko, V. (2022). On the experimental robustness of the Allais Paradox. *American Economic Journal: Microeconomics, 14*, 143–163. https://doi.org/10.1257/mic.20190153

Braga, J., Humphrey, S. J., & Starmer, C. (2009). Market experience eliminates some anomalies – and creates new ones. *European Economic Review, 53*, 401–416.

Cox, J. C., & Grether, D. M. (1996). The preference reversal phenomenon: Response mode, markets and incentives. *Economic Theory, 7*, 381–405.

Dietrich, F., Staras, A., & Sugden, R. (2021). Savage's response to Allais as Broomean reasoning. *Journal of Economic Methodology, 28*, 143–164. https://doi.org/10.1080/1350178X.2020.1857424

Friedman, M., & Savage, L. (1952). The expected-utility hypothesis and the measurability of utility. *Journal of Political Economy, 1*, 463–474.

Gilboa, I. (2010). Questions in decision theory. *Annual Review of Economics, 2*, 1–19. https://doi.org/10.1146/annurev.economics.102308.124332

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association, 1*, 114–125. https://doi.org/10.1007/s40881-015-0004-4

Hausman, D. (2012). *Preference, Value, Choice and Welfare*. Cambridge University Press.

Hendriks, A. (2012). SoPHIE-Software platform for human interaction experiments. Working Paper, University of Osnabrück.

Humphrey, S. J. (1995). Regret-aversion or event-splitting effects? More evidence under risk and uncertainty. *Journal of Risk and Uncertainty, 11*, 263–274.

Humphrey, S. J. (2000). The common consequence effect: Testing a unified explanation of recent mixed evidence. *Journal of Economic Behavior and Organization, 41*, 239–262. https://doi.org/10.1016/S0167-2681(99)00075-X

Humphrey, S.J. and Meickmann, F.C. (2022). Who accepts the stationarity principle? Working paper, University of Osnabrück.

Infante, G., Lecouteax, G., & Sugden, R. (2016). Preference purifiction and the inner rational agent: A critique of the conventional wisdom of behavioural welfare economics. *Journal of Economic Methodology, 23*, 1–25. https://doi.org/10.1080/1350178X.2015.1070527

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*, 263–291.

Kruse, N.Y. (2022). Three experimental investigations of individual decision-making behaviour. Unpublished doctoral thesis, University of Osnabrück.

Loomes, G., Starmer, C., & Sugden, R. (1991). Observing violations of transitivity by experimental methods. *Econometrica, 59*, 425–439.

MacCrimmon, K.R. (1968). Descriptive and Normative Implications of the Decision-Theory Postulates. In: Borch, K., Mossin, J. (eds) Risk and Uncertainty. International Economic Association Conference Volumes, Numbers 1–50. Palgrave Macmillan, London. https://doi.org/10.1007/978-1-349-15248-3_1

Mongin, P. (2019). The Allais paradox: What it became, what it really was, what it now suggests to us. *Economics and Philosophy, 35*, 423–459. https://doi.org/10.1017/S0266267118000469

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

Nielsen, K., & Rehbeck, J. (2022). When choices are mistakes. *American Economic Review, 112*, 2237–2268. https://doi.org/10.1257/aer.20201550

Plott, C. R. (1996). Rational individual behavior in markets and social choice processes: The discovered preference hypothesis. In K. Arrow, E. Colombatto, M. Perlman, & C. Schmidt (Eds.), *The Rational Foundations of Economic Behaviour* (pp. 225–250). London: McMillian.
Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior and Organization, 3*, 323–343.
Savage, L. J. (1954). *The Foundations of Statistics*. Wiley.
Slovic, P., & Tversky, A. (1974). Who accepts Savage's Axiom? *Behavioral Science, 19*, 368–373. https://doi.org/10.1002/bs.3830190603
Starmer, C. (1992). Testing new theories of choice under uncertainty using the common consequence effect. *Review of Economic Studies, 59*, 813–830.
Starmer, C. (1999). Cycling with rules of thumb: An experimental test of a new form of non-transitive behaviour. *Theory and Decision, 46*, 141–158.
Starmer, C. (2005). Normative notions in descriptive dialogues. *Journal of Economic Methodology, 12*, 277–289. https://doi.org/10.1080/13501780500086206
Starmer, C., & Sugden, R. (1991). Does the random lottery incentive system elicit true preferences? An experimental investigation. *American Economic Review, 81*, 971–978.
Sugden, R. (2016). Do people really want to be nudged towards healthy lifestyles? *International Review of Economics, 64*, 113–123. https://doi.org/10.1007/s12232-016-0264-1
Sugden, R. (2018). 'Better off, as judged by themselves': A reply to Cass Sunstein. *International Review of Economics, 65*, 9–13. https://doi.org/10.1007/s12232-017-0281-8
Sunstein, C. R. (2018). "Better off, as judged by themselves": A comment on evaluating nudges. *International Review of Economics, 65*, 1–8. https://doi.org/10.1007/s12232-017-0280-9
Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving Decisions About Health, Wealth and Happiness*. Yale University Press.
Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business, 59*, S251–S278.
Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representations of uncertainty. *Journal of Risk and Uncertainty, 5*, 297–323.
Van de Kuilen, G., & Wakker, P. (2006). Learning in the Allais Paradox. *Journal of Risk and Uncertainty, 33*, 155–164.

Authors and Affiliations

Steven J. Humphrey[1] · **Nadia-Yasmine Kruse**[1]

✉ Steven J. Humphrey
steve.humphrey@uos.de

Nadia-Yasmine Kruse
nadia.yasmine.kruse@uos.de

[1] School of Business Administration and Economics, University of Osnabrück, Rolandstr. 8, 49069 Osnabrück, Germany