Journal of Behavioral and Experimental Finance 49 (2026) 101134

Contents lists available at ScienceDirect

Journal of Behavioral and Experimental Finance

journal homepage: www.elsevier.com/locate/jbef

Full length article

Testing axiomatizations of ambiguity aversion ☆

Daniel L. Chen

Toulouse School of Economics, Institute for Advanced Study in Toulouse, University of Toulouse Capitole, Toulouse, France

ARTICLE INFO

JEL classification:
D81

Keywords:
Ellsberg paradox
Machina paradox
Uncertainty aversion
Independence axiom

ABSTRACT

The study of choice under uncertainty has advanced through key "paradoxes," such as the Ellsberg paradox. We implement Machina's (2014) three-outcome extension, in which four major ambiguity-aversion theories (multiple priors, rank-dependent, smooth ambiguity, variational) all predict indifference between two ambiguous acts. Contrary to these predictions, we find most participants do not express indifference. Our design elicits each subject's certainty equivalent (CE) for an embedded 50–50 lottery and uses that CE in the Machina acts. Under lottery independence—i.e., if individuals apply standard (von Neumann–Morgenstern) expected utility to each objective lottery—these acts map to the same distribution of payoffs and thus should be evaluated identically. Yet we document a robust preference for one act over the other. This preference is associated with violations of lottery independence (e.g., Allais inconsistencies), as well as with disappointment aversion. Our results highlight that Machina's three-outcome paradox is at least as much about failing independence over lotteries as it is about ambiguity aversion.

1. Introduction

The development of the normative and positive theory of behavior under uncertainty is characterized by a series of thought experiments to which scholars or laypersons often give answers that contradict prevailing theory. The St.-Petersburg-Paradox challenged the notion that a lottery will be evaluated by its expected value (de Montmort, 1713). Bernoulli (1738) proposed a concave utility function instead of the payoffs themselves. Allais (1953) subsequently proposed a thought experiment demonstrating that many people do not exhibit the behavior suggested by Bernoulli and von Neumann and Morgenstern's expected utility theory. Ellsberg (1961) further challenged the notion that decision-makers have a single subjective probability distribution (i.e., are probabilistically sophisticated) with a thought experiment involving choice over ambiguity (Feduzi, 2007). Empirical papers (for a survey see Camerer and Weber, 1992) showed that people behave differently than probabilistic sophistication prescribes. New models were proposed to accommodate the ambiguity non-neutrality observed in the Ellsberg experiment. The four prevailing theories are: Schmeidler's (1989) Choquet model (or Rank-Dependent Utility); Gilboa and Schmeidler's (1989) maximin expected utility; Klibanoff et al.'s (2005) smooth ambiguity; and Maccheroni et al.'s (2006) Variational Preferences Model. Ambiguity attitudes are now used to explain puzzles in finance (Erbas and Mirakhor, 2007), promote policies in health (Sutter et al., 2013), law (Segal and Stein, 2005), and the environment (Viscusi and Zeckhauser, 2006), and explain phenomena in the lab (Liu and Colman, 2009; Ball et al., 2012; Baillon et al., 2016).

A thought experiment proposed by Machina (2014) challenges the prevailing four theories of ambiguity aversion. It extends Ellsberg's urn setup to three possible outcomes rather than two. Imagine an urn containing one red ball and two others that may be black or white—either both black (BB), both white (WW), or one black and one white (BW). A decision maker must choose between two acts, denoted L ("ambiguity at low outcomes") and H ("ambiguity at high outcomes"). Each act assigns monetary payoffs to the colors so that, under von Neumann–Morgenstern evaluation of risk, both acts yield the same overall distribution of payoffs—and hence all major ambiguity-aversion models predict indifference between them. Yet Machina conjectured that many people would not be indifferent. The three possible payoffs are \$0, \$c, and \$100, where \$c is the certainty equivalent of a 50–50 lottery paying \$0 or \$100.

☆ This paper builds on an earlier working paper that was coauthored with Martin Schonger. I am grateful to Martin Schonger for his early collaboration on the initial version of this project. Work on this project was conducted while Chen received financial support from Alfred P. Sloan Foundation, United States (Grant No. 2018-11245), European Research Council (Grant No. 614708), Swiss National Science Foundation, Switzerland (Grant Nos. 100018-152678 and 106014-150820), TSE-Partnership and ANITI funding, and IAST funding from the French National Research Agency (ANR) under the Investments for the Future (Investissements d'Avenir) program, grant ANR-17-EUR-0010. There are no relevant or material financial interests that relate to the research described in this paper. Data and code are available here: https://osf.io/zsf4u/.

E-mail address: daniel.chen@iast.fr.

https://doi.org/10.1016/j.jbef.2025.101134
Received 28 July 2025; Received in revised form 28 October 2025; Accepted 9 December 2025
Available online 18 December 2025

Act L			Act H		
1-3 5-7 2 balls		1 ball	1 ball	2 balls	
Black	White	Red	Red	Black	White
\$0	\$*c*	\$100	\$0	\$*c*	\$100

A central feature of Machina's three-outcome paradox is that it combines both subjective and objective uncertainty. The subjective uncertainty involves which state of the world obtains—e.g., whether the unknown balls in the urn are black or white—while the conditional objective uncertainty concerns the probability of drawing a particular ball color within that state (e.g., a 2/3 chance of drawing black in state BB). Purely subjective or purely objective uncertainty alone typically does not generate the same kind of paradoxical prediction.

The broader literature has recognized that blending subjective and objective components can create richer paradoxes that challenge classical expected utility. For instance, Machina (2014) himself emphasizes that many major models of ambiguity (such as multiple priors, rank-dependent, smooth ambiguity, or variational preferences) allow for ambiguity in the state probabilities but still impose von Neumann–Morgenstern (vNM) expected utility on the conditional objective lotteries. Under these models, Machina's act L and act H end up having the same overall distribution of payoffs—hence predict indifference.

This underscores a key limitation in existing ambiguity theories: they typically do not relax how people evaluate the conditional (objective) lottery. As a result, any re-arrangement of payoffs among states that preserves the same final distribution yields indifference. However, according to Machina (2014), "If ambiguity aversion somehow involves 'pessimism', might not an ambiguity averter have a strict preference for [Act] H over [Act] L, just as a risk averter might prefer bearing risk about higher rather than lower outcome levels?" Our key contribution is to implement the Machina "ambiguity at low vs. at high problem with three colors" thought experiment.

We employ a two-part experimental design to elicit each subject's certainty equivalent (CE) of a simple 50–50 lottery and then use that CE as one of the possible outcomes in the Machina (2014) thought experiment. We implement this using the PRINCE (PRior INCEntive) method (Johnson et al., 2021), which is formally equivalent to BDM (Becker et al., 1964) but allows us to pre-randomize the implemented choice. This design choice enables us to present subjects with all instructions at the beginning, avoid additional strategic considerations between parts, and maintain transparent incentives.

In brief, the subject first states, for every possible monetary amount X, whether they would prefer X in cash or the 50–50 lottery (the "CE task"). Then, for the Machina task, each subject makes choices between two ambiguous acts L and H, for each possible value of X. By drawing a single randomly selected value of X before the experiment (and placing it in a sealed envelope), we ensure that there are clear, pre-determined incentives. This approach directly connects the CE task and the Machina task without introducing separate or hidden randomization later. In our experimental implementation, subjects are not indifferent between Acts L and H. On average, subjects prefer Act L over Act H. We use Dillenberger and Segal (2015) in combination with Gul's (1991) disappointment aversion to give conditions under which Act L or Act H is preferred.

We contribute evidence that distinguishes between theoretical foundations of ambiguity aversion. Machina also proposed earlier thought experiments in Machina (2009). Machina distinguishes his 2014 thought experiment, which is based on a single source of purely subjective uncertainty, unlike Machina (2009), which is based on two. Baillon et al. (2011) and L'Haridon and Placido (2010) theoretically and empirically investigated Machina's 2009 two-source thought experiment, whereas the present study implements Machina's (2014) three-outcome single-source problem. These approaches are complementary: theirs tests ambiguity behavior across two independent sources of uncertainty, while ours isolates violations of indifference within a single source under the Anscombe–Aumann framework. Together, they advance the argument that the Machina paradoxes falsify many ambiguity theories, at least in the Anscombe–Aumann framework adopted by those theories with the vNM independence axiom as central. Under the Anscombe–Aumann formulation, decision makers evaluate risky lotteries using von Neumann–Morgenstern expected utility within each state and then aggregate these state utilities according to an ambiguity model. Because this structure implies that all major ambiguity-aversion models predict indifference between the two acts in Machina's (2014) thought experiment, the present study serves as an empirical test of the AA-based axiomatization itself. Yang and Yao (2017) design a mean-preserving experiment involving two draws with replacement from the classic Ellsberg two-color urns and find systematic ambiguity aversion, complementing the present study's focus on Machina's three-outcome extension.

Our findings add to a broader body of evidence questioning the descriptive adequacy of the Anscombe–Aumann (AA) framework for modeling decision making under uncertainty. Previous research has shown that the framework's key assumptions—linearity in probabilities for risk and separability across states—often fail empirically. Trautmann and Wakker (2018) argue that relaxing these assumptions can make AA-based models more suitable for descriptive applications. Schneider and Schonger (2019) experimentally test the AA monotonicity axiom and find systematic violations that mirror those observed here. Oechssler and Roomets (2021) compare the Savage and AA formulations and demonstrate that the paradoxical predictions of AA disappear when ambiguity and risk are treated symmetrically. Taken together, these studies, along with the present results, suggest that the AA framework's simplifying assumptions may be too restrictive for capturing actual behavior and that more flexible formulations are needed to reconcile theory with observed choices.

The remainder of the paper is organized as follows. Section 2 outlines Machina's three-outcome thought experiment and explains why major ambiguity models predict indifference. Section 3 describes our lab study's design and reports key findings. Finally, Section 4 offers concluding remarks.

2. Machina thought experiment

Machina (2014) proposes a paradox involving mixed subjective and objective uncertainty. Consider an urn with three balls, exactly one of which is red, while the other two might be:

- Both black (BB),
- Both white (WW),
- One black and one white (BW).

A decision maker (DM) faces *subjective* uncertainty about which of these states obtains (i.e., the DM may not know the probabilities $q_{\mathrm{BB}}, q_{\mathrm{WW}}, q_{\mathrm{BW}}$), but once a state is realized, the probability of drawing a specific color in that state is *objectively* determined.

Let \$c be the DM's certainty equivalent (CE) of a simple 50–50 lottery paying \$0 or \$100. We study two acts, L (ambiguity at low payoffs) and H (ambiguity at high payoffs). In Act L, ambiguity concerns the lower outcomes—uncertainty applies to whether the decision maker receives \$0 or \$c—while the highest payoff (\$100) is fixed. In Act H, ambiguity concerns the higher outcomes—uncertainty applies to whether the decision maker receives \$c or \$100—while the lowest payoff (\$0) is fixed. We refer to these cases respectively as "ambiguity at low" and "ambiguity at high". Table 1 illustrates, for each of the two acts, the *objective probability distribution* over outcomes for each state, as well as the corresponding ***vNM expected utility***.

The last two columns in Table 1 show the vNM expected utility for each objective lottery given a utility function u over monetary outcomes. Normalizing $u(0) = 0$ and $u(100) = 100$, it follows that $u(c) = 50$. As a result, we have the following. In BB and BW, both

Table 1
vNM Utility for Acts L and H.

State	Act L	Act H	vNM Utility L	vNM Utility H
BB	$(\frac{2}{3}, 0;\ \frac{1}{3}, 100)$	$(\frac{1}{3}, 0;\ \frac{2}{3}, c)$	$\frac{2}{3}u(0) + \frac{1}{3}u(100)$	$\frac{1}{3}u(0) + \frac{2}{3}u(c)$
BW	$(\frac{1}{3}, 0;\ \frac{1}{3}, c;\ \frac{1}{3}, 100)$	$(\frac{1}{3}, 0;\ \frac{1}{3}, c;\ \frac{1}{3}, 100)$	$\frac{1}{3}u(0) + \frac{1}{3}u(c) + \frac{1}{3}u(100)$	$\frac{1}{3}u(0) + \frac{1}{3}u(c) + \frac{1}{3}u(100)$
WW	$(\frac{2}{3}, c;\ \frac{1}{3}, 100)$	$(\frac{1}{3}, 0;\ \frac{2}{3}, 100)$	$\frac{2}{3}u(c) + \frac{1}{3}u(100)$	$\frac{1}{3}u(0) + \frac{2}{3}u(100)$

Table 2
DA Utility for Acts L and H.

State	Act L	Act H	DA Utility L	DA Utility H
BB	$(\frac{2}{3}, 0;\ \frac{1}{3}, 100)$	$(\frac{1}{3}, 0;\ \frac{2}{3}, c)$	$\frac{100}{3+2\beta}$	$\frac{100}{3+\beta}$
BW	$(\frac{1}{3}, 0;\ \frac{1}{3}, c;\ \frac{1}{3}, 100)$	$(\frac{1}{3}, 0;\ \frac{1}{3}, c;\ \frac{1}{3}, 100)$	$\frac{150}{3+\beta}$	$\frac{150}{3+\beta}$
WW	$(\frac{2}{3}, c;\ \frac{1}{3}, 100)$	$(\frac{1}{3}, 0;\ \frac{2}{3}, 100)$	$\frac{(2/3)(1+\beta)50\ +\ (1/3)\,100}{1+\beta\,(2/3)}$	$\frac{200}{3+\beta}$

acts yields a vNM utility of $1/3$, and in WW, both acts yields a vNM utility of $2/3$. Hence, both Acts L and H yield the *same* mapping from states to vNM utilities, and *any aggregator* that just forms a (subjective) weighted sum over those utilities will yield the same overall utility for Acts L and H. For instance, suppose a DM's beliefs about BB, BW, WW lie within some set of priors $p \in \Delta$. Under *multiple-priors utility*, the DM evaluates each act by

$$V(\text{Act}) = \min_{\mathbf{p}\in\Delta} \left[p_{\text{BB}}\, U_{\text{BB}}(\text{Act}) \;+\; p_{\text{BW}}\, U_{\text{BW}}(\text{Act}) \;+\; p_{\text{WW}}\, U_{\text{WW}}(\text{Act})\right].$$

If $U_s(\cdot)$ is the vNM expected utility of the conditional lottery, then each prior $\mathbf{p}$ assigns the *same row-wise utility* for L and H, thus $V(L) = V(H)$.

2.1. Disappointment aversion

Suppose instead that, *rather than* vNM expected utility, the DM evaluates objective lotteries in terms of a disappointment-averse (DA) utility (Gul, 1991). Disappointment aversion generalizes vNM expected utility: in addition to the utility u over monetary outcomes, a parameter $\beta > 0$ that penalizes outcomes below a lottery's mean. DA collapses to vNM if and only if $\beta = 0$. Table 2 gives the DA utility for each of the states for Acts L and H (with the same normalization for u as in Table 1). (The calculations for Table 2 are in our online Appendix A.)

When $\beta > 0$, the DA utility of the objective lottery in states BB or WW differs for acts L and H. As a result, an aggregate of the DA utilities in the states can lead to a different value for the two acts. For instance, suppose a decision maker evaluates lotteries in terms of the DA utilities in Table 2, and entertains multiple priors over the subjective states. Specifically, suppose $P(\text{BW}) = 1/3$ and $P(\text{BB}) \in [0, 2/3]$, with $P(\text{WW}) = 2/3 - P(\text{BB})$. Then, it is easily verified that $V(L) > V(H)$.

2.2. Limitations of the Anscombe–Aumann approach

Like Machina (2014), we adopt the Anscombe–Aumann (AA) framework, which treats risk and ambiguity as separable. Under AA, decision makers are assumed to evaluate objective lotteries within each state using von Neumann–Morgenstern expected utility and then aggregate those state utilities according to an ambiguity model. This "backward-induction" logic imposes linear weighting of known probabilities and separability across states—assumptions that make the framework analytically convenient but descriptively restrictive. Empirically, people often overweight small probabilities, so if a subject distorts known probabilities even slightly, the predicted indifference between Acts L and H will fail. In that sense, Machina's apparent paradox is not truly paradoxical: once probability weighting or non-separability is allowed, strict preferences between L and H naturally emerge. As noted in Wakker (2024, p. 1838), *rank-dependent utility* without the AA assumption could accommodate preferences that break $L \sim H$ by relaxing the requirement of vNM independence for conditional risk. Thus, while we treat $L \sim H$ as a contradiction for multiple priors, rank-dependent/Choquet, smooth, or variational models *under AA*, other non-AA formulations may avoid the paradox.

Indeed, under Savage-based or purely statewise models, such as RDU without AA, a decision maker can be "ambiguity-seeking" for low-likelihood events and "ambiguity-averse" for high-likelihood events, thereby breaking indifference *even* if they hold "Choquet-style" beliefs over states. However, once we impose Anscombe–Aumann's monotonicity or substitutability, the same rank-dependent (or multiple priors, or smooth, etc.) aggregator will yield indifference between Acts L and H if the DM applies standard expected utility *within* each conditional lottery. Thus, the "contradiction" we highlight primarily concerns these four major theories *within* the AA framework (see also Machina, 2014).

Consequently, our experimental findings should be interpreted with the caveat that they directly test these models *under the Anscombe–Aumann assumption* of vNM evaluation for risk. We acknowledge that non-AA versions of RDU or other ambiguity models (i.e., those not imposing vNM on conditional lotteries) may well accommodate the strict preferences we observe without contradiction.

Several authors have questioned the theoretical force of Machina's paradox itself. Trautmann and Wakker (2018) argue that the paradox arises from the restrictive assumptions of the Anscombe–Aumann framework—particularly its separability and linearity in probabilities—and propose modifications that make the framework more descriptively adequate. Oechssler and Roomets (2021) compare Savage- and Anscombe–Aumann-based formulations and find that the paradox disappears when ambiguity and risk are treated symmetrically. These theoretical contributions suggest that Machina's paradox should be viewed not as a refutation of ambiguity theories per se, but as evidence that the Anscombe–Aumann structure may be too narrow to capture observed behavior.

3. Lab study

3.1. Design

We ran the lab experiment at the DeSciL lab following their standard procedures in ETH Zurich using paper-and-pencil, for reasons described below. We had 91 participants across 6 sessions. Rather than replacing \$$c$ with the lottery it is induced by as in Fig. 1, we sought to recover \$$c$ through revealed preference. If the decision-maker has a preference relation which satisfies continuity, then a certainty equivalent is guaranteed to exist; strict monotonicity in the monetary outcomes ensures uniqueness. However, the certainty equivalent of a subject is unknown to the experimenter.

The main challenge is to elicit the subject's certainty equivalent prior to conducting the Machina "ambiguity at low vs. at high problem with three colors" thought experiment. The state-of-the-art method

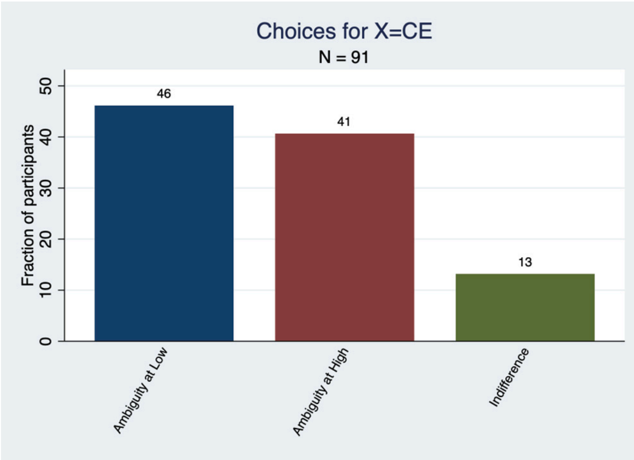

Fig. 1. All participants.
Notes: Fraction of participants who choose ambiguity for the high outcome, ambiguity for the low outcome, or show indifference between the two options.

to experimentally elicit willingness to pay for an object is still BDM (Becker et al., 1964). BDM can be implemented by the mechanism itself or a simplified "list" method. In the mechanism, people are asked to state their true valuation, a price is randomly drawn, and they receive the object at the random price if their stated valuation is above it. In the "list" method, people are presented with a list of choices, each consisting of two options, the object and a valuation, and one of the indicated choices is then selected at random. From a formal point of view, the two are close cousins, the difference being that in the list method the valuation one can state is quite coarse. Regardless of the method, subjects are usually told that correctly stating their true valuation is optimal.

We implement a two-part design. In Part 1 (the "CE task"), each participant states, for every possible value of $X \in \{0, 1, \dots, 20\}$, whether they prefer receiving X in cash or playing a 50–50 lottery that pays 20 or 0 Swiss francs. This allows us to infer each individual's certainty equivalent (CE). In Part 2 (the "Machina task"), participants compare Act L and Act H for every possible value of X.

Before the experiment begins, a sealed envelope with a random integer X (between 0 and 20) is prepared for each participant. That random X is relevant in both parts: in Part 1, subjects effectively give instructions for each possible X; in Part 2, they again indicate whether they prefer Act L or Act H for each possible X. At the end, only the pair of decisions corresponding to the envelope's X is actually implemented (one decision from the CE task and one from the Machina task). This approach simplifies the design by ensuring that all instructions are provided upfront, avoids any follow-up randomization after the first part, and makes the link between the two parts transparent to participants. We follow Johnson et al. (2021) in referring to this approach as "PRINCE", but the main idea is simply to pre-commit to the random draw of X.

In the Machina task, Act L has ambiguity at the lower outcomes (0 and X) and a fixed payoff of 20 if a red ball is drawn, whereas Act H has ambiguity at the higher outcomes (X and 20) but can yield 0 if a red ball is drawn. Participants check a box indicating which act they prefer, for each of the 21 possible values of X. After completing all choices, the envelope is opened to reveal the specific X. We then implement the corresponding choice from Part 1 and Part 2, using the same urn draws and lottery procedures described in Section 2. This yields a clean measure of whether Act L or Act H is chosen when the certain outcome is exactly the participant's own CE.

It is worth highlighting how PRINCE contrasts with the usual BDM. First, we do not directly ask subjects to state their true valuation of a lottery and then ask subjects the Machina (2014) thought experiment where that just-elicited valuation appears to increase the values of the acts. Subjects reading the instructions for the entire experiment would easily realize how the two tasks are related. Our use of the PRINCE method provides full transparency of incentives. Valuations of the lottery from subjects are elicited with their full awareness of the entire experiment. The lottery whose valuation is being elicited appears as "Option A" in the second task. Notice further that the realization of the random draw is inside an envelope that they hold. This realized draw is then used in the Machina thought experiment. We then ask subjects to choose between the acts for every possible value of the draw. The connection of the envelope's content across tasks is maximally salient to subjects. What we use, as the experimenter, is the valuation reported in the second task to locate the actual comparison of interest among the 20 choice decisions in the third task. Thus, we raise minimal suspicion from subjects (there is a clear connection between the second and third tasks) and without deception (we present the full set of instructions prior to subjects making any decisions).

To familiarize subjects with PRINCE, we first used it for a first order stochastic dominance (FOSD) task (See Appendix B.1) and then for CE. Since the Machina experiment is implemented with the list method, we can explore if subjects have a unique switching point. A priori it is not clear that people have a unique switching point nor direction.

3.2. Results

In Part 1 (the "CE task"), each participant is given a list of possible cash amounts $X \in \{0, 1, \dots, 20\}$. For each X, they must choose either "Receive CHF X for sure" or "Play a 50–50 lottery for CHF 0/20." Crucially, participants may specify more precise (non-integer) values if desired—e.g., 7.50—by writing them on the answer sheet. This PRINCE-based approach thus eliminates the need to force participants into pre-determined integer bins. In our dataset, 24 out of 91 participants (approximately 26%) reported a non-integer threshold.

After collecting these responses, we infer each participant's certainty equivalent (CE)—the point at which they switch from preferring the sure amount to the lottery. (If they write, say, 7.50, that is recorded directly as their CE.) This CE then matters in Part 2 (the "Machina task"), where participants compare two ambiguous acts, L and H, at every $X \in \{0, 1, \dots, 20\}$. Importantly, before the experiment starts, each participant receives a sealed envelope that contains a randomly chosen integer X. Only after Part 2 is completed do we open the envelope: the revealed X determines (a) which sure-cash choice from Part 1 applies, and (b) which Act-L-vs.-Act-H choice from Part 2 is implemented.

Specifically, to test whether a subject is indifferent at their own CE, we observe which act they choose in Part 2 for the envelope's X. If this X equals (or is close to) a subject's self-reported CE, we can check whether they are indeed "indifferent" at that exact value. In cases where participants report a non-integer CE (e.g., 7.50), we look at their Part 2 choices for both 7 and 8 to see if their switch between L and H occurs in the interval [7,8]. This procedure ensures that all CEs—integer or not—are appropriately matched to the same scale used for the ambiguous acts. As we show below, most participants exhibit a clear preference in Part 2's ambiguous acts, rejecting the theoretical prediction that they would be indifferent exactly at their CE.

Beginning with explicit statements of indifference, we find that participants tend to prefer the act with ambiguity at the low outcome rather than at the high outcome. Fig. 1 shows that only 13 of 91 subjects check "indifferent", and a binomial test rejects the hypothesis that everyone is indifferent ($p < 0.001$). Next, we use each participant's switching point from our choice-list method in Part 2 to infer "indifference" more broadly. Fig. 2 classifies participants as indifferent if they: 1. Explicitly report indifference at or near their CE (including up to two adjacent X values if they reported a non-integer CE), or 2. Have a clear switching point S that lies sufficiently close to their elicited CE when $\text{CE} \in [S - 1.96\,\text{SD}(\text{CE} - S),\, S + 1.96\,\text{SD}(\text{CE} - S)]$. The SD is calculated

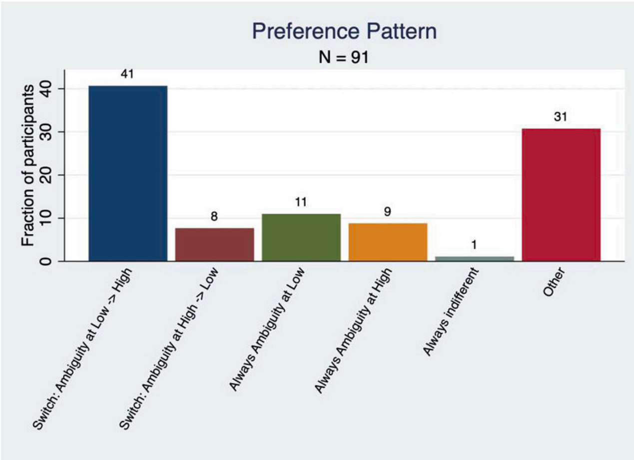

Fig. 2. All participants.
Notes: Fraction of participants who choose to: (1) Switch from ambiguity at low to ambiguity at high, (2) Switch from ambiguity at high to ambiguity at low, (3) Always choose ambiguity at low, (4) Always choose ambiguity at high, (5) Always remain indifferent, or (6) Other.

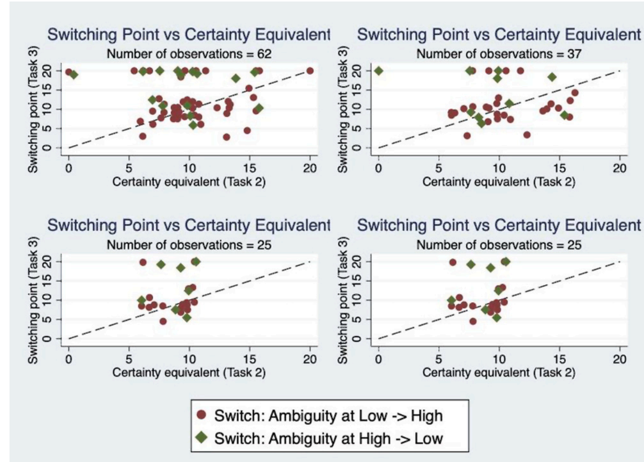

Fig. 3. CE vs. Switching point (raw data).
Notes: In each subplot, the 45 degree line is the $CE = S$ line. This sample includes people who always prefer L or always prefer H (their switching point is represented as 20) and people with single switching points. Each subplot presents a different sample in robustness checks. Clockwise from the upper left: (i) All participants, (ii) $CE \in [4, 10]$, (iii) FOSD, (iv) both.

for $|\text{CE} - S|$ for the population, which biases against our results as this would tend to overestimate the number of people who are indifferent.

Here, S is defined as the midpoint between the last instance of choosing A over B and the first instance of choosing B over A. While this method potentially overestimates indifference by treating any small gap as “measurement error”, the vast majority of participants still fail to qualify as indifferent under these generous criteria. Their CE values and switching points deviate noticeably, suggesting a genuine preference for one act.

We categorize participants into six behavioral types based on their choice patterns across the 21 rows (values of X) within the single Machina task. (i) “Switching from ambiguity at low to ambiguity at high” refers to subjects who initially prefer ambiguity over the lower outcomes (Act L) but later switch to preferring ambiguity over the higher outcomes (Act H) as X increases. (ii) Conversely, “switching from high to low” describes participants who begin by preferring ambiguity at high outcomes but later switch to preferring ambiguity at low outcomes. These classifications capture within-subject changes in preference across values of X, not separate tasks. The remaining four categories correspond to participants who always choose (iii) ambiguity at low, (iv) always at high, (v) always declare indifference, or (vi) display multiple (non-monotonic) switches.

Three key findings emerge: 1. About one-fifth never switch at all, strictly preferring L or strictly preferring H. 2. Among switchers, most move from “ambiguity at low” to “ambiguity at high” as X increases—consistent with wanting to avoid ambiguity at higher payoffs. Even with reversed presentation order, a majority switch in that same direction. 3. Many participants’ CE differs greatly from S, reinforcing that we can reject the Machina-based indifference at CE. A binomial test of the difference between the proportion of expressed indifference and 1 yields $p = 0.000$. Appendix C provides further tabulations supporting this conclusion. To rule out layout effects, we also tested for order dependence: the proportion of participants switching from ambiguity-at-low to ambiguity-at-high outcomes does not differ significantly between normal and reversed presentation orders ($p = 0.65$). These steps confirm that the directional result—preferences tending from ambiguity at low toward ambiguity at high as X increases—is robust and not an artifact of task layout or coding.

Overall, even allowing direct expressions of indifference and employing liberal criteria for classification, the data strongly reject the hypothesis that subjects are broadly indifferent between “ambiguity at low” and “ambiguity at high” outcomes.

To visualize how far participants are from indifference at their own certainty equivalents (CE), Fig. 3 plots each subject’s CE on the x-axis and the switching point (S) on the y-axis. In each panel, the 45-degree line represents perfect alignment of $\text{CE} = S$. The sample includes people who always choose Act L or always choose Act H (coded as $S = 20$) and those with a single switch. We display four different subsets in robustness checks, clockwise from the upper left: 1. All participants, 2. Participants with $\text{CE} \in [4, 10]$, 3. Participants who passed the First Order Stochastic Dominance (FOSD) test, and 4. Participants who satisfy both (2) and (3). (“Passed FOSD” means they consistently chose the stochastic-dominant option in a preliminary choice list.)

Visually, the vast majority of points in Fig. 3 do not lie near the 45-degree line, indicating that for most individuals $\text{CE} \neq S$. A t-test strongly rejects the hypothesis that the mean of $|\text{CE} - S|$ is zero (t=7.8). Appendix D confirms this finding via a regression on “folded” data (i.e., absolute differences–to avoid averaging responses from subjects who switch above their CE with those who switch below their CE), showing that the confidence interval for the slope excludes 1. Together, these results reject the prediction of indifference at CE.

3.3. Allais consistency and machina behavior

We now examine how participants’ Allais consistency relates to their behavior in the Machina thought experiment. To do so, we presented two hypothetical questions modeled on the classic Allais paradox (see appendix for instructions to the subjects). In these questions, each subject chose between:

• Lottery A: $1 Million for sure
• Lottery B: 1% chance of $0, 89% chance of $1 Million, 10% chance of $5 Million

and then:

• Lottery C: 89% chance of $0, 11% chance of $1 Million
• Lottery D: 90% chance of $0, 10% chance of $5 Million

Overall, these findings show that participants systematically reject the indifference predicted by standard ambiguity-aversion models within the Anscombe–Aumann framework. Most subjects prefer ambiguity at low rather than at high outcomes, even when the two acts yield identical distributions of objective payoffs. This asymmetry is closely associated with violations of the von Neumann–Morgenstern independence axiom—particularly Allais-type inconsistencies—suggesting that

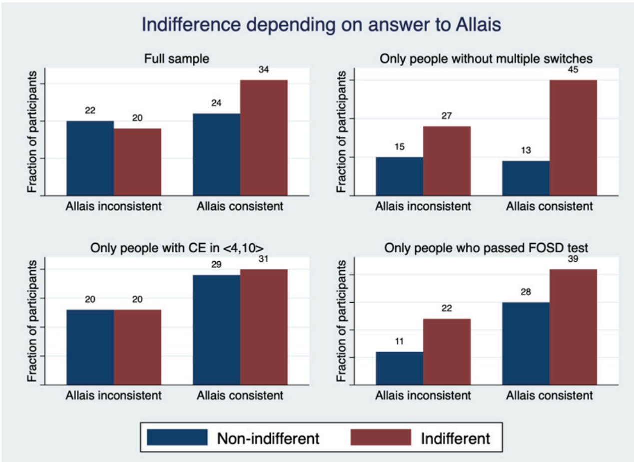

Fig. 4. Allais and Machina paradoxes.
Notes: This figure displays the proportion of subjects who are categorized as indifferent. A subject is considered indifferent under two conditions: (1) They express indifference at their certainty equivalent (CE) or at two neighboring values. (2) They exhibit a clear switching point, with their CE falling within the confidence interval of this switching point. Each panel in the figure represents a sub-sample analysis, segmenting subjects based on their adherence to Allais consistency. Allais consistency is defined by whether subjects satisfy the vNM independence axiom. The sub-sample analyses are conducted as follows: (1) Full Sample: Includes all subjects. (2) No Multiple Switches: Includes only subjects who did not exhibit multiple switching points. (3) Passed FOSD Test: Includes only subjects who passed the First Order Stochastic Dominance (FOSD) test. (4) CE Range: Includes only subjects whose CE falls within a specified range of values.

failures of indifference arise as much from non-EU behavior under risk as from ambiguity attitudes per se. Participants who satisfy independence (e.g., Allais-consistent or FOSD-consistent individuals) tend to exhibit indifference, whereas others display systematic directional preferences. Together, these patterns support disappointment aversion and related non-separable frameworks as more flexible accounts of observed choices.

We also examine participants who pass a First-Order Stochastic Dominance (FOSD) check. Our data indicate that most FOSD passers—those who always choose the stochastically dominant lottery—also tend to be Allais consistent. Moreover, these FOSD passers/Allais-consistent participants are more likely to exhibit indifference in the Machina test, especially near their own certainty equivalent. This outcome matches our theoretical prediction that decision-makers who satisfy vNM independence (or something close to it) do not violate Machina-style indifference.

Hence, while only a subset of our sample passes both FOSD and Allais checks, these subjects align most strongly with the conventional prediction of indifference in Machina's "ambiguity-at-low-vs.-high" scenario. This underscores our interpretation that failing to be Allais consistent (and/or FOSD consistent) is the primary reason most participants do not exhibit Machina-style indifference.

In Fig. 4, subjects are classified as indifferent when they express indifference at their CE (and two neighboring values) or when they have a clear switching point and their CE lies in the confidence interval of this switching point. The phrase "CE falling within the confidence interval of the switching point" refers to cases where an individual's elicited certainty equivalent (CE) lies within the 95 percent confidence interval for that participant's inferred switching point (S). This interval is based on the cross-participant variation in the distribution of CE – S values. We compute these confidence bounds from the standard error of CE – S across subjects and classify a participant as "indifferent" if their CE falls inside this range. The figure shows that indifference appears to depend on the answer to Allais.

Overall, the qualitative alignment with Gul's (1991) disappointment aversion model arises because outcomes below a lottery's reference point are penalized, generating asymmetric evaluation of objective lotteries that can explain the observed preference for ambiguity at low outcomes. A full structural model would be a promising direction for future work, but the present design focuses on testing Machina's (2014) theoretical prediction under the Anscombe–Aumann assumptions.

4. Concluding remarks

Our experiment implemented Machina's (2014) three-outcome thought experiment, in which four major theories of ambiguity aversion predict indifference between two ambiguous acts. Contrary to those predictions, we find that most participants strictly prefer one act over the other, signaling that these classical models do not fully capture behavior when ambiguity involves more than two possible outcomes. Notably, participants who do satisfy vNM independence (e.g., those who are Allais consistent or pass first-order stochastic dominance) are likelier to exhibit indifference, supporting the theoretical link between vNM independence and Machina's paradox.

This divergence between predicted and observed behavior has two major implications. First, it underscores that violations of vNM independence—rather than ambiguity aversion per se—may be the critical driver behind rejections of indifference. Ambiguity models grounded in vNM independence can be accurate only to the extent that decision-makers themselves adhere to that axiom. Second, our data provide empirical support for more flexible frameworks that can accommodate strict preferences in Machina's scenario once the von Neumann–Morgenstern assumption for risk is relaxed, such as models using disappointment aversion, prospect theory for ambiguity, rank-dependent utility, and source-dependent approaches. We emphasize disappointment aversion in particular because it offers the most parsimonious extension of expected utility—introducing a single parameter that captures the observed asymmetry between gains and losses—while remaining formally compatible with the Anscombe–Aumann framework tested here. These findings suggest that relaxing linearity and separability assumptions, as in these broader models, can reconcile observed behavior with theoretical predictions and motivate future experimental work. These frameworks can accommodate decision-makers whose beliefs or risk attitudes differ at higher outcomes versus lower outcomes.

Looking ahead, further theoretical modeling should examine how alternative axioms or preference structures interact with vNM independence to produce deviations from classical predictions. Extending or refining models to capture non-independence, multi-stage ambiguities, or psychological factors (such as disappointment, regret, or other emotional responses) appears essential. Additionally, exploring broader forms of uncertainty—including those beyond classical Ellsberg-type setups—promises to deepen our understanding of how real-world decision-making departs from standard rational-choice benchmarks. Our findings thus highlight the importance of incorporating both institutional (e.g., multiple outcomes, uncertain states) and psychological (e.g., disappointment aversion) factors into the design of ambiguity theories and experiments.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jbef.2025.101134.

References

Allais, Maurice, 1953. Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine. Econ.: J. Econ. Soc. 503–546.

Baillon, Aurélien, Koellinger, Philipp D., Treffers, Theresa, 2016. Sadder but wiser: The effects of emotional states on ambiguity attitudes. J. Econ. Psychol. 53, 67–82.

Baillon, Aurélien, L'Haridon, Olivier, Placido, Laetitia, 2011. Ambiguity models and the Machina paradoxes. Am. Econ. Rev. 101 (4), 1547–1560.

Ball, Linden J., Bardsley, Nicholas, Ormerod, Tom, 2012. Do preference reversals generalise? Results on ambiguity and loss aversion. J. Econ. Psychol. 33 (1), 48–57.

Becker, Gordon M., DeGroot, Morris H., Marschak, Jacob, 1964. Measuring utility by a single-response sequential method. Behav. Sci. 9 (3), 226–232.

Bernoulli, Daniel, 1738. Exposition of a new theory on the measurement of risk. Comment. Acad. Sci. Imp. Petropolitanae Transl. Econom. V, 175–192, 22(1), 1954.

Camerer, Colin, Weber, Martin, 1992. Recent developments in modeling preferences: Uncertainty and ambiguity. J. Risk Uncertain. 5 (4), 325–370.

de Montmort, Pierre Rémond, 1713. Essay of Analysis on Games of Chance. J. Quillau.

Dillenberger, David, Segal, Uzi, 2015. Recursive Ambiguity and Machina's Examples. Technical Report, University of Pennsylvania, pp. 55–61.

Ellsberg, Daniel, 1961. Risk, ambiguity, and the savage axioms. Q. J. Econ. 75 (4), pp. 643–669.

Erbas, S. Nuri, Mirakhor, Abbas, 2007. The Equity Premium Puzzle, Ambiguity Aversion, and Institutional Quality. Technical Report, IMF Working Paper.

Feduzi, Alberto, 2007. On the relationship between Keynes conception of evidential weight and the Ellsberg paradox. J. Econ. Psychol. 28 (5), 545–565.

Gilboa, Itzhak, Schmeidler, David, 1989. Maxmin expected utility with non-unique prior. J. Math. Econom. 18 (2), 141–153.

Gul, Faruk, 1991. A theory of disappointment aversion. Econometrica 59 (3), 667–686.

Johnson, Cathleen, Baillon, Aurélien, Bleichrodt, Han, Li, Zhihua, Van Dolder, Dennie, Wakker, Peter P., 2021. Prince: An improved method for measuring incentivized preferences. J. Risk Uncertain. 62 (1), 1–28.

Klibanoff, Peter, Marinacci, Massimo, Mukerji, Sujoy, 2005. A smooth model of decision making under ambiguity. Econometrica 73 (6), 1849–1892.

L'Haridon, Olivier, Placido, Laetitia, 2010. Betting on Machinas reflection example: an experiment on ambiguity. Theory and Decision 69 (3), 375–393.

Liu, Hsin-Hsien, Colman, Andrew M., 2009. Ambiguity aversion in the long run: Repeated decisions under risk and uncertainty. J. Econ. Psychol. 30 (3), 277–284.

Maccheroni, Fabio, Marinacci, Massimo, Rustichini, Aldo, 2006. Ambiguity aversion, robustness, and the variational representation of preferences. Econometrica 74 (6), 1447–1498.

Machina, Mark J., 2009. Risk, ambiguity, and the rank-dependence axioms. Am. Econ. Rev. 99 (1), 385–392.

Machina, Mark J., 2014. Ambiguity aversion with three or more outcomes. Am. Econ. Rev. 104 (12), 3814–3840.

Oechssler, Jörg, Roomets, Alex, 2021. Savage vs. Anscombe-Aumann: an experimental investigation of ambiguity frameworks. Theory and Decision 90 (3), 405–416.

Schmeidler, David, 1989. Subjective probability and expected utility without additivity. Econometrica 57 (3), 571–587.

Schneider, Florian H., Schonger, Martin, 2019. An experimental test of the Anscombe–Aumann monotonicity axiom. Manag. Sci. 65 (4), 1667–1677.

Segal, Uzi, Stein, Alex, 2005. Ambiguity aversion and the criminal process. Notre Dame L. Rev. 81, 1495.

Sutter, Matthias, Kocher, Martin G., Glaetzle-Ruetzler, Daniela, Trautmann, Stefan T., 2013. Impatience and uncertainty: Experimental decisions predict adolescents' field behavior. Am. Econ. Rev. 103 (1), 510–531.

Trautmann, Stefan, Wakker, Peter P., 2018. Making the Anscombe-Aumann approach to ambiguity suitable for descriptive applications. J. Risk Uncertain. 56 (1), 83–116.

Viscusi, W. Kip, Zeckhauser, Richard J., 2006. The perception and valuation of the risks of climate change: A rational and behavioral blend. Clim. Change 77 (1–2), 151–177.

Wakker, Peter P., 2024. Annotated bibliography. https://personal.eur.nl/wakker/refs/webrfrncs.pdf. (Accessed 30 December 2024).

Yang, Chun-Lei, Yao, Lan, 2017. Testing ambiguity theories with a mean-preserving design. Quant. Econ. 8 (1), 219–238.