Journal of Behavioral and Experimental Economics 107 (2023) 101950

Contents lists available at ScienceDirect

Journal of Behavioral and Experimental Economics

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

A criticism of Bernheim & Sprenger's (2020) tests of rank dependence☆

Peter P. Wakker

Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, The Netherlands

ARTICLE INFO

JEL-Classification:
D81
C91

Keywords:
Prospect theory
Rank dependence
Complexity aversion

ABSTRACT

Bernheim and Sprenger (2020, *Econometrica*; SB) claimed to experimentally falsify rank dependence in prospect theory. This paper criticizes SB's results and novelty claims. Their experiments only captured well-known heuristics and not genuine preferences. Many falsifications of rank dependence have been made before, and SB's equalizing reductions have also been used before. SB thought to identify probability weighting and utility where they are unidentifiable, invalidating all SB's related claims. SB used an incorrect formula of original prospect theory. Their suggested alternative of rank-independent probability weighting with dependence on the number of outcomes (their "complexity aversion;" a misnomer) has long been discarded.

1. Introduction

Bernheim and Sprenger (2020) (SB henceforth[1]) claimed to falsify rank-dependent probability weighting. Rank dependence was introduced independently by Quiggin (1982) and Schmeidler (1989; first version 1982). Tversky and Kahneman (1992) used it to improve their prospect theory (CPT). This paper criticizes SB and the follow-up paper Bernheim, Royer, and Sprenger (2022) (§9). The main text focuses on criticisms relevant to the main conclusions. Appendix OA.4 (Supplementary Material) lists 15 further inaccuracies in SB.

Several papers, not cited by SB, have reported violations of rank dependence and prospect theory more serious than those claimed by SB (§5; §8). Wakker's (2022) keyword "PT falsified" gives 59 violations, reproduced here in Appendix OA.5 (Supplementary Material). Still, there are so many more positive findings that CPT is the most popular descriptive risk theory available today; no better alternative is available.[2]

SB's main experimental problem is that their stimuli are too complex while the stakes (payoff *differences*) are too low. Several preceding papers considered tests as in SB but avoided these problems and then did find rank dependence (§§4–5). Hirshman and Wu (2022) will provide a replication of SB that corrects their experimental mistakes. SB claimed to introduce a new measurement method of decision weights. However, they were preceded by Diecidue, Wakker, and Zeelenberg (2007), not cited by SB, who showed that linear utility is needed for this method, contrary to SB's claim of general validity (§5).

SB suggested probability weighting of original prospect theory (Kahneman & Tversky 1979), which they called rank-independent, as an improvement over rank-dependent weighting. Unfortunately, they used an incorrect formula (§2.1). Further, this approach has long been known to be too problematic (§2.2). SB (their §3.2) thought to identify probability weighting and utility from stimuli where they were in fact unidentifiable (§2.3), invalidating all their related claims.

SB suggested, as a second improvement, preference functionals that depend on the number of outcomes in a lottery, which they called complexity aversion. However, this idea needs further specification before being operational and has often been considered not to be promising (§6). Thus, SB did not provide a viable alternative to the rank-dependent weighting they criticized.

2. Three problems for SB's treatment of 1979 prospect theory

By

$$(p : X,\ q : Y,\ 1 - p - q : Z), \tag{1}$$

called *lottery*, we denote a probability distribution over *outcomes* (monetary gains; $\mathbb{R}^+$) that assigns probability p to outcome X,

☆ This paper heavily builds on Abdellaoui, Li, Wakker, and Wu (2020), often verbatim. However, I alone am responsible for errors.

E-mail address: Wakker@ese.eur.nl.

[1] We avoid the abbreviation BS for linguistic reasons.

[2] See Barberis (2013 p. 2068, finance); Dean & Ortoleva (2017 p. 386, economics); Fehr-Duda & Epper (2012 §6, economics); Murphy & ten Brincke (2018 p. 309); Pachur et al. (2018 p. 408, psychology); Ruggeri et al. (2020 abstract, general).

https://doi.org/10.1016/j.socec.2022.101950
Received 1 August 2022; Received in revised form 14 October 2022; Accepted 17 October 2022
Available online 18 October 2023

probability q to outcome Y, and probability $1-p-q$ to outcome Z ($p \geq 0,\ q \geq 0,\ p+q \leq 1$).[3] SB only considered lotteries with three or fewer outcomes. Fewer outcomes result if some of the probabilities in Eq. (1) are 0. By $u : \mathbb{R}^{+} \rightarrow \mathbb{R}^{+}$ we denote a *utility function* (or value function). It is strictly increasing and continuous, with $u(0) = 0$. By $\pi : [0,1] \rightarrow [0,1]$ we denote a *weighting function*. It is strictly increasing and continuous with $\pi(0) = 0$ and $\pi(1) = 1$.

2.1. Incorrect formula

SB defined *rank-independent probability weighting* as the following evaluation of the lottery in Eq. (1):

$$\pi(p)u(X) + \pi(q)u(Y) + \pi(1-p-q)u(Z). \tag{2}$$

Contrary to SB's claims, and as explained by Wakker (2023), this is not Kahneman & Tversky's 1979) original prospect theory.[4] It is known as *separable prospect theory*. When Nilsson, Rieskamp, and Wagenmakers (2020) discovered that an earlier paper by them had erroneously used Eq. (2) instead of original prospect theory, they and the Journal of Mathematical Psychology took a principled and exemplary stance and published a correction using the right formula.

2.2. Problematic violations of stochastic dominance

Rank-independent weighting (our Eq. (2)) was popular in the psychological literature until the 1980s, but was abandoned when Fishburn (1978) discovered that it violates stochastic dominance, as does original prospect theory. However, it has not always been understood that these violations are extreme and absurd, rendering the theory both theoretically and empirically unacceptable. To illustrate this point, Wakker (2023) considered the following lottery:

$$\left(0.01 : 1 + 1 \times 10^{-5}, \ldots, 0.01 : 1 + 99 \times 10^{-5}, 0.01 : 0\right).$$

With the parametric estimates of Tversky and Kahneman (1992)[5] and under the natural extension of Eq. (2) and also of original prospect theory to multiple outcomes (Wakker, 2023), agreeing here with separable prospect theory, the certainty equivalent of this lottery is 6.9, exceeding the maximal outcome of the lottery more than six-fold. This does not make any sense. The basic problem of rank-independent weighting is that the requirements for lotteries with few outcomes are entirely incompatible with those for many outcomes.[6]

Rieger and Wang (2008) further showed that extensions of rank-independent weighting to continuous distributions are not well possible, confirming preceding conjectures in the literature (Quiggin, 1982 p. 330).[7] Some studies suggested rank-independent weighting may work just as well as CPT for fitting student-lab choices between lotteries with no more than, say, four outcomes (Gonzalez & Wu 2022; Peterson, Bourgin, Agrawal, Reichman, & Griffiths, 2021; Starmer, 1999; Wu, Zhang, & Abdellaoui, 2005). However, rank-independent weighting does not work well beyond. Violations include Bernasconi (1994 p. 69), Blondel (2002 pp. 260–261), Edwards (1996 §III.B), Fennema and Wakker (1997), Fudenberg and Puri (2022 p. 422), L'Haridon (2009 p. 548), Levy (2008 all tasks), Loehman (1998 p. 293), Schneider and Lopes (1986 p. 546 first para), and Sonsino, Benzion, and Mador (2002) p. 946). Rank-independent weighting cannot be used in monotonic economic theories. Whereas SB (p. 1364) once acknowledged the problem of violation of stochastic dominance by rank-independent probability weighting, the rest of their paper was invariably positive about it and recommended a return to this old psychological theory. Section 6 returns to this point.

One of the main empirical findings for decision under risk concerns the overweighting of best and worst outcomes and the underweighting of intermediate outcomes (Fehr-Duda & Epper 2012; l'Haridon & Vieider 2019; Luce & Suppes 1965 §4.3; Starmer, 2000). This finding fits well with rank dependence but falsifies rank-independent probability weighting.

2.3. Non-identifiability of probability weighting and utility

SB aimed to measure probability weighting and utility. To do so, they only considered lotteries with one nonzero outcome in both their experiments. However, a joint power of probability weighting and utility is then unidentifiable. Thus, $\pi(p)u(x)$ is empirically indistinguishable from $\pi(p)^r u(x)^r$ for any $r > 0$ (Cohen & Jaffray 1988 Eqs. (5) and 7a). For this reason, Fehr-Duda and Epper (2012 p. 583) strongly advised against using only such stimuli.

SB claimed that

$$(p : x,\ 1-p : 0) \rightarrow \frac{p^{0.715}}{\left(p^{0.715} + (1-p)^{0.715}\right)^{1/0.715}} x^{0.941}$$

fits their data best in Experiment 1. However, $\left(\frac{p^{0.715}}{\left(p^{0.715}+(1-p)^{0.715}\right)^{1/0.715}}\right)^{\frac{1}{0.941}} x$ fits their data exactly as well. In particular, the power family they assume for utility can never rule out linear (or convex or concave) utility as best fitting. Similarly, in SB's second experiment, $\left(\frac{p^{0.766}}{\left(p^{0.766}+(1-p)^{0.766}\right)^{1/0.766}}\right)^{\frac{1}{0.982}} x$ fits the data exactly as well as their claimed optimal fit. Hence, SB could not really identify utility and probability weighting.

The incorrect measurements complicate SB's claims on rank-independent probability weighting. We, therefore, do not discuss them further except in the last paragraph of §6. We instead focus on SB's analyses of rank dependence.

3. Deterministic analysis of SB's experiments and linear utility

This section gives some preparatory mathematical definitions. In particular, it identifies an assumption of linear utility that SB took no account of, discussed further in later sections. We assume CPT, with the following evaluation of lotteries:

$$(p : X,\ q : Y,\ 1-p-q : Z) \rightarrow w_X u(X) + w_Y u(Y) + w_Z u(Z). \tag{3}$$

Here, u is as above, and w_X, w_Y, and w_Z are *decision weights*. Decision weights are nonnegative and add to 1. They are *rank-dependent*. For example, w_X depends on whether X is the best, middle, or worst outcome. We follow SB in using the term rank informally and not expressing rank dependence in notation. A complete definition of general CPT, including reference dependence and a formalization and notation for ranks and rank dependence, is in Wakker (2010).

[3] I deviate from SB's notation of lotteries because their use of braces to denote arrays rather than sets violates common conventions. In general, I follow their terminology and notation as much as possible, sometimes reluctantly, to facilitate communication.

[4] SB (footnote 11) cited Camerer & Ho (1994) for Eq. 2. However, Camerer & Ho's endnote 16 pointed out that Eq. 2 deviates from prospect theory for strictly positive lotteries. SB (footnote 11) also cited Fennema & Wakker (1997) for Eq. 2. However, Fennema & Wakker (p. 54) pointed out that they used Eq. 2 only for mixed lotteries, which assign positive probabilities to both gains and losses. Then Eq. 2 does agree with original prospect theory (Wakker 2023).

[5] Because original and new prospect theory agree on the two-outcome gain lotteries used there, these estimates are also valid for original prospect theory. They may be somewhat different for separable prospect theory, but this does not affect the essence of the example here.

[6] Contrary to SB's suggestions, dependence on number of outcomes provides no solution (§6).

[7] Such extensions depend too much on the particular discrete approximations chosen and depend on π only through $\pi'(0)$.

SB's first and main experiment concerned indifferences of the form

$$(p:X,\ q:Y,\ 1-p-q:Z) \sim (p:X,\ q:Y+m,\ 1-p-q:Z-k). \quad (4)$$

X, the common outcome, was varied across SB's main experiment, with m and k so small that the ranking of outcomes is the same for the two lotteries in Eq. (4) (i.e., "comonotonicity" is satisfied). Throughout, $Y = 24$, $Z = 18$, $m = 5$, and $q = 0.3$. Price lists were used to elicit k (called an *equalizing reduction* by SB) for three values of p: $p = 0.1$, $p = 0.4$, and $p = 0.6$.

SB used seven values of X. In some instances, X was the best-ranked outcome ($X = 34$, $X = 32$, or $X = 30$). In other cases, X, which we denote X', was ranked in the middle ($X' = 23$, $X' = 21$, or $X' = 19$). When X is ranked best, the weights of outcomes Y and $Y + m$ are denoted w_Y. When X is ranked middle, they are denoted w_Y'. We similarly write k'. Under linear utility, $\frac{w_Y}{w_Z} = \frac{k}{m}$ and $\frac{w_Y'}{w_Z} = \frac{k'}{m}$. The ratio

$$\frac{k'}{k} = \frac{w_Y'}{w_Y} \quad (5)$$

captures the proportional change of the decision weight and, hence, rank dependence. This ratio, or its log, was used in SB's analysis.

SB repeatedly claimed that Eq. (5) could be used for all differentiable utility functions. However, this claim was based on marginal rates of substitution involving infinitesimal changes m and k, which cannot be implemented empirically.[8] Empirically, we have to work with moderate outcome changes and the following assumption:

Assumption 1. [linear utility for moderate outcome changes]. For outcome changes within a small interval $[A, B]$, utility is approximately linear.[9] □

SB's second experiment concerned indifferences of the form

$$(p:X,\ q:Y,\ 1-p-q:Z) \sim (p:X+m,\ q:Y-k,\ 1-p-q:Z-k). \quad (6)$$

In this experiment, $p = 0.4$ and $q = 0.3$, or $p = 0.6$ and $q = 0.2$, with $Y = 36$, $Z = 18$, and $m = 4$ throughout. Finally, $X = 2,\ 3,\ 4,\ 20,\ 21,\ 22,\ 38,\ 39$, or 40, with price lists again used to elicit k, which SB again called an *equalizing reduction*. Under linear utility and Eq. (3),

$$w_X = \frac{k}{m+k}. \quad (7)$$

Consider $X = 4$ (with k) and $X' = 20$ (with k'). In the lottery with $X = 4$, X has the worst rank, whereas $X' = 20$ has the middle rank in the corresponding lottery. The rank of Z changed from middle to worst in these two lotteries. Such rank changes affect the decision weight in Eq. (7). SB again captured the change by the ratio $\frac{k'}{k}$.[10] Eq. (7) again uses Assumption 1. More precisely, it used linearity of utility in the intervals $[\min\{18-k, 18-k'\}, 18]$ and $[\min\{36-k, 36-k'\}, 36]$.

[8] SB's Footnote 13 even claimed validity for infinitesimals for every strictly increasing continuous utility, dispensing with differentiability. However, this is not correct. For singular Cantor-type functions, the *positive* right derivatives claimed by SB may exist *nowhere*, let be at the points where needed. See Paradís, Viader, & Bibiloni (2001; their Theorem 3.1 and its proof are also valid for right and left derivatives).

[9] For the studies considered in this paper, $[A, B] = [0, 100]$ suffices. More precisely, for Eq. 5 and SB's first experiment, linearity of utility is used on all intervals $[\min\{Z-k, Z-k'\}, Z]$.

[10] This result is not exact. More precisely, the ratio of decision weights is $\frac{k'}{k} \times \frac{m+k}{m+k'}$ which is, roughly, a monotonic nonlinear transformation of k'/k. Importantly, it does not affect being larger or smaller than 1. Probably, SB still used k'/k, or its log, as index in their analysis for this reason.

4. Small payoff changes: Ramsey's trifle problem in SB's experiment

Ramsey (1931) pointed out a difficulty that applies to SB's implementation of Assumption 1, which we call *Ramsey's trifle problem*:

> Since it is universally agreed that money has a diminishing marginal utility, if money bets [to measure decision weights (subjective probabilities) through ratios] are to be used, it is evident that they should be for as small stakes as possible. *But then again the measurement is spoiled by introducing the new factor of reluctance to bother about trifles*. [Italics added] [p. 176]

Samuelson (1960 Footnote 5) also referred to this trifle problem. SB were not aware of it. They used very small payoff changes m, k so as to approximate infinitesimal changes with perfect linearity. But these changes were too small to motivate subjects, making choice options almost identical. SB's subjects had to bother about trifles. Bear in mind that differences in outcomes, rather than outcomes themselves, give motivation for careful preferences.

Abdellaoui et al. (2020) §4) argued in detail that SB's experiments measured only heuristics, not preferences. In brief, there were too many choices with too small incentives. Smith (1982, "dominance"), Wilcox (1993), and many others cautioned against this. Ariely, Loewenstein, and Prelec (2001) cautioned against coherent arbitrariness (called the shaping hypothesis by Loomes, Starmer, & Sugden, 2003) in such cases, where subjects develop coherent heuristics rather than coherent preferences. Appendix OA.1 (Supplementary Material) shows mathematically that this happened in SB. Further, the layout in SB's Experiment 1 made *cancellation* (ignoring common outcomes heuristically rather than as reflecting true preference) too salient. Weber and Kirsner (1997 top of p. 57) showed that Wakker, Erev, and Weber (1994) suffered from cancellation, explaining the absence of rank dependence there. Weber & Kirsner avoided it and then did find rank dependence. These papers used stimuli as in SB's first experiment (Eq. (4)). Thus, SB's first experiment replicated Wakker, Erev, & Weber's cancellation problem.

To avoid cancellation, SB (§5.3) conducted a second smaller experiment, based on Eq. (6). Now, there was no common outcome to be canceled. The format in SB's second experiment was used before but never became popular because of its reliance on linear utility (§5). Hence, less is known about the presence or absence of heuristics. Still, too many problems of SB's Experiment 1 remained. The complexity of the lotteries increased due to the absence of a common outcome. This augmented Ramsey's trifle problem. The layout of the stimuli (their Online Appendix, Fig. 5) with the same format for hundreds of choices over several pages, again induced coherent arbitrariness.

5. A measurement of equalizing reductions preceding SB

Diecidue et al. (2007; DWZ henceforth) used the same equalizing reductions as in SB's second experiment, i.e., indifferences as in our Eq. 6; see their Eq. 3.2. Thus, following Weber and Kirsner (1997), they avoided the cancellation in Wakker et al. (1994). DWZ's primary purpose was also to test rank dependence (their main hypothesis on p. 185 and p. 195 2nd para) using quantitative parameter-free estimations rather than counting statistics, preceding SB. To avoid Ramsey's trifle problem, their outcome differences k, m were considerably larger than SB's, with m never below €20 (≈ $30 in 2020 value). To be able to still use Assumption 1, these differences were chosen not to be very large though. Using moderate outcome differences and linear utility (Assumption 1) is the only way SB's equalizing reductions can be used, and this is what DWZ did.

DWZ discussed and minimized all aforementioned experimental problems of SB (§4).[11] Their main findings supported rank dependence (DWZ p. 192 last para) but DWZ also found violations of rank dependence. Decision weights sometimes changed even though ranks did not. SB's finding on rank dependence was reversed in the sense that their decision weights did not change even if ranks did. This absence of rank dependence can be taken as a special case of CPT (expected utility), whereas DWZ's finding really falsifies CPT. Wu (1994), using different stimuli, found more extensive violations of rank dependence. The violations of rank dependence by DWZ, Wu, and others (§8) are more serious than SB's absence of rank dependence in the same way that the Allais paradox (inconsistent utility curvature) is a more serious violation of expected utility than risk neutrality (absence of utility curvature).

Once it is understood that SB's equalizing reductions need linear utility, measuring indifferences and then calculating decision weights using linear utility, as SB (and DWZ) do, is not very original (l'Haridon & Vieider 2019 Eq. (3)). Although linear utility can be defended for the moderate stakes used (DWZ p. 196; l'Haridon & Vieider 2019 p. 189), this method of DWZ, SB, and others never became very popular in economics. Most measurements of prospect theory allow for nonlinear utility.

6. Complexity aversion

SB (p. 1364) wrote:

> "probability weighting [rank-independent, as in separable prospect theory] … implies violations of first-order stochastic dominance … This is a serious flaw …"

Nevertheless, the rest of the paper was invariably positive about rank-independent probability weighting, returning to the displayed problem only at the end. SB then suggested that "complexity aversion" (their term[12]) could resolve the problem. However, the central question should have been whether complexity aversion can help *reduce* the problematic violations of stochastic dominance. SB did not address this question. Instead, they suggested that complexity aversion could *add* violations of stochastic dominance that they considered desirable. Appendix OA.3 (Supplementary Material) does address the central question, with a negative answer: complexity aversion cannot help reduce violations of stochastic dominance. Thus, SB's suggested alternative for rank-dependent probability weighting does not solve the problem.

There are several other problems with SB's analysis of complexity aversion as a dependence on the number of outcomes. Many authors, not cited by SB, have investigated such dependence. Neilson (1992) proposed a formal model, but Humphrey (2001a) tested it unsuccessfully. Related models received some attention in psychology (Birnbaum, 2008 p. 481; Krantz, Luce, Suppes, & Tversky, 1971 Ch. 8; Luce, 2000). Tversky and Kahneman (1992 p. 317) discussed the idea but were pessimistic:

> Despite its greater generality, the cumulative functional is unlikely to be accurate in detail. We suspect that decision weights may be sensitive to the formulation of the prospects, as well as to the *number*, the spacing and the level of outcomes. … The present theory can be generalized to accommodate such effects but it is questionable whether the gain in descriptive validity, achieved by giving up the separability of values and weights, would justify the loss of predictive power and the cost of increased complexity. … The heuristics of choice do not readily lend themselves to formal analysis because their application depends on the formulation of the problem, the method of elicitation, and the context of choice. [italics added]

I share their pessimism.

Another problem is that the prevailing empirical finding for gains, away from the certainty effect, is complexity seeking rather than aversion as claimed by SB, in studies not cited by SB. These studies usually considered a pure case: certainty equivalents are measured for different framings of identical lotteries, for instance $(0.4:30,\ 0.6:20)$ versus $(0.4:30,\ \ 0.3:20,\ \ 0.3:20)$. Although all rational theories require identical certainty equivalents, experiments find systematic differences. Here a pure effect of the perceived number of outcomes occurs, clearer than the discontinuous changes of certainty equivalents considered by SB. Such effects have been known as *event-splitting effects* (or boundary effects or violations of coalescing/collapsing).

Event-splitting effects are special cases of attribute-splitting effects (Weber, Eisenführ, & von Winterfeldt, 1988), or the part-whole bias (Bateman, Munro, Rhodes, Starmer, & Sugden, 1997), or, for uncertainty, the unpacking effect (Tversky & Koehler 1994). That is, splitting something up usually increases the total decision weight. This underlies several theories by Birnbaum cited below, who provided detailed analyses explaining why we mostly find complexity seeking but sometimes complexity aversion. A literature search gave:

- The following three papers report prevailing complexity aversion: Bernheim and Sprenger (2020), Huck and Weizsäcker (1999), Moffatt, Sitzia, and Zizzo (2015).
- The following nine papers report prevailing complexity seeking: Birnbaum (2005), Birnbaum (2007), Erev, Ert, Plonsky, Cohen, and Cohen (2017), Fennema and Wakker (1996), Humphrey (1995), Humphrey (2000), Humphrey (2001a), Humphrey (2006), Starmer and Sugden (1993).
- The following five papers report about as much aversion as seeking: Birnbaum (2004), Birnbaum, Schmidt, and Schneider (2017), Humphrey (2001b), Schmidt and Seidl (2014), Weber (2007).

Several of these references were given by Birnbaum (2008, p. 473), a paper cited by SB (Footnote 69) but not for these findings opposite to their complexity aversion. We conclude that the findings on complexity aversion are volatile, but the literature has documented more complexity seeking than aversion for gains.

The above discussion shows that dependency on the number of outcomes is primarily driven by factors other than complexity perception. SB's term complexity aversion is a misnomer. Aversion to more comprehensive and fitting forms of complexity has been studied by Armantier and Treich (2016), Bruce and Johnson (1996), Kovarik, Levin, and Wang (2016), Mador, Sonsino, and Benzion (2000), and

[11] In the beginning of §3, DWZ explained that three-outcome lotteries are too difficult to evaluate in general; see also DWZ (p. 181, 3rd & 4th para). Hence, they used a visual design (their Fig. 1) to facilitate these choices. This is commonly done for lotteries with three or more outcomes, by Weber & Kirsner (1997) and most others. Lola Lopes, specialized in multi-outcome lotteries, developed special visual designs (Lopes & Oden 1999; Fennema & Wakker 1997). DWZ developed their stimuli in extensive pilot studies with debriefings to identify and then avoid the major heuristics used by subjects (see their p. 188 last two paras; p. 194 last para; p. 195 last para). They used filler questions and more variations in outcomes to further reduce heuristics. Following Weber & Kirsner (1997), DWZ emphasized that avoiding heuristics is desirable to increase statistical power (p. 188 last line; p. 195 3rd para). There are two further differences between DWZ and SB. First, improving one lottery was not compensated by worsening that same lottery elsewhere, but instead by improving the other lottery. This avoids confounds due to differences between improvements and worsenings. Second, DWZ considered the more interesting context of uncertainty (ambiguity) with unknown probabilities instead of risk with known probabilities,. While this does not affect the theoretical working of rank dependence (DWZ, p. 185 ll. 7-8 and Wakker 2010 Fig. 7.4.1 versus 10.4.1), it may, of course, impact differences in empirical findings.

[12] SB (p. 1367 & p. 1402) wrote "a form of complexity aversion: people may prefer lotteries with fewer outcomes because they are easier to understand" and used this definition thoughout their paper.

Sonsino et al. (2002)). SB cited five papers on complexity aversion in their footnote 70, but they were all taken out of context.[13]

SB's reports of rank-independent nonlinear probability weighting are confounded by their incorrect measurements of probability weighting and utility. But their findings, if re-analyzed correctly, probably reflect event-splitting effects rather than complexity aversion. Framing and preference reversal effects of this kind are known to be strong. However, they violate not only prospect theory but every transitive and monotonic economic theory. Fennema and Wakker (1996) and Humphrey (1995) pointed out that event-splitting effects can be modeled using separable prospect theory with subadditive weighting. However, Sonsino et al. (2002) p. 946) found that this way of modeling does not work well.

7. SB's criticism of common statistical tests

SB (§2.3) criticized counting tests, i.e., statistical tests that compare numbers of violations of predictions, preference axioms in our case. Such statistics are commonly used throughout decision theory and all empirical sciences. SB seeked to avoid crediting priority to much preceding literature this way. However, a fundamental problem in SB's discussion is that they did not know that *all* statistical tests, not only their own tests but also counting tests, are based on assumptions about probabilistic errors ("noise").[14] See:

> These types of frequency comparisons raise two difficulties, both stemming from the fact that the results are difficult to interpret *without a parametric model of noisy choice*. First, the premise of the approach—that violation frequencies are *necessarily*[15] higher for invalid axioms—is flawed. [italics and footnote added] [SB p. 1376]

SB's lack of understanding of this point is also apparent in SB (p. 1367 l −3 and p. 1376 2nd para). Their claimed first problem (p. 1376) only shows the *existence* of an error model under which counting tests are incorrect. However, deviating error models exist for every statistical test (Greenland et al. (2016) p. 338 2nd column 1st para), including those used by SB. For example, if the errors in SB's indifference measurements are not constant or not stochastically independent or extreme, then their claimed p-values and confidence intervals are not valid either. Extreme errors are, in fact, plausible given that SB's test statistics were ratios of numbers close to 0 with big variances (Eqs. (5) and 7; see Appendix OA.1 (Supplementary Material). In general, the relevant question is not the existence, but the plausibility, of an error model deviating from the one required. Unaware of the latter concept, SB did not address this relevant question.

SB's second claimed difficulty concerns the example at the end of their Online Appendix B. It assumes stimuli for which CPT and expected utility have identical predictions. Then counting tests indeed have no power. But then, no statistical test does. This trivial example cannot serve as a criticism of counting tests or any other test. I failed to understand SB's description of this example in their main text (p. 1376).[16] SB's claim to have refuted all counting tests, widely used in all empirical sciences, is entirely unfounded. Therefore, SB should not have escaped from crediting priority to the many preceding papers that used counting statistics.

8. Further preceding falsifications of prospect theory

Given that prospect theory is the most tested theory of decision under risk, besides being most confirmed, it is also most violated. Wakker's (2022) keyword "PT falsified" gives 59 references (reproduced in Appendix OA.5 (Supplementary Material)). Further, Fennema and Wakker (1996) replicated Wakker et al. (1994) for ambiguity, and again found no evidence of rank dependence.

We call special attention to a finding of Birnbaum and McIntosh (1996), not cited by SB. It was confirmed in several follow-up studies by Birnbaum and colleagues, surveyed by Birnbaum (2008 pp. 484–487), and found independently by Humphrey and Verschoor (2004). This finding is of special interest because it concerns lotteries of the same format as in Eq. (4), i.e., as in SB's first experiment, with the common outcome X moved to test rank dependence. Prospect theory predicts that weights increase if ranks change from middle to best or worst. SB quantitatively found no change in decision weights. The aforementioned studies avoided heuristics and found a stronger deviation: a decrease, rather than an increase, in weight. These violations (inconsistent rank dependence) are, again, more serious than SB's violations in the same way as the Allais paradox is a more serious violation of expected utility than risk neutrality.

The strongest counterexample to rank dependence that I am aware of is Machina's (2009) reflection example, confirmed empirically by l'Haridon and Placido (2010) and informally qualified as "brilliant" by Wakker (2022).

9. A criticism of Bernheim Et Al. (2022)

Bernheim et al. (2022), BRS henceforth, redid part of SB's first experiment. They avoided cancellation and fatigue and improved stimulus explanations, as was common in preceding studies (DWZ; Weber & Kirsner 1997). However, contrary to BRS's claims, they worsened rather than improved incentives. As explained above, *differences* in outcomes (m, k in BRS's notation), rather than outcomes themselves, provide incentives. Unfortunately, BRS mostly increased outcomes but not their differences, even though Abdellaoui et al. (2020) p. 218, p.819) had warned against this. BRS's subjects still had to bother about trifles. In only one incentivized choice pair ("Condition 5″), difference was increased ($m = 20$ iso $m = 5$), though still below DWZ's minimal difference. Incentives were worsened because only few subjects were paid, 1/20th. Remarkably, BRS's Fig. 1 did suggest possible rank dependence for this choice pair, but BRS reported no statistical analysis of it.

Apart from a trivial test of stochastic dominance (k^+ vs. k^-), all statistical conclusions in BRS were based on accepted null hypotheses. Every statistics textbook warns against this (Greenland et al., 2016, Misinterpretation 4). Power analyses would have been warranted.[17]

BRS ignored and gave no counterarguments to any of Abdellaoui et al. (2020) warnings other than those concerning experimental stimuli. Thus, BRS (Finding 2) continued to use the invalid measurements of probability weighting and utility, to claim novelty on equalizing reductions while not citing the preceding DWZ, to use an incorrect formula of original prospect theory, and to declare all counting statistics invalid (BRS end of §1). BRS did not address the inviability of their suggested alternatives of rank-independent weighting and complexity aversion. BRS's null hypotheses and Finding 1 did not falsify rank dependence but only showed neutrality. For real falsification, BRS should also have solved the many problems of their Finding 2 (which they did not reconsider).

[13] The first four, Iyengar & Kamenica (2010), Iyengar & Lepper (2000), Iyengar, Jiang, & Huberman (2003), and Sonsino & Mandelbaum (2001), considered a different topic, preference against flexibility (number of available choice options to choose one from). The fifth, Stodder (1997), is on confusions of averages versus marginals and complexity of multiple stage lotteries, which, again, are different topics. It is also only theoretical, with no data.

[14] Without it, p-values could not even be defined.

[15] Counting tests, as all statistics, allow for deviating samples, be it with small probabilities not exceeding a significance threshold.

[16] SB wrote: "For any given degree of rank dependence, one can construct simple examples (with constant "distance to indifference") in which the differential between violation frequencies falls anywhere between zero and unity."

[17] Confidence intervals of within-subject differences, rather than the "between"-subject confidence intervals of BRS's Fig. 1, would also have been useful.

10. Conclusion

Bernheim and Sprenger (2020, SB) and Bernheim et al. (2022) reported flawed experiments to falsify rank dependence and prospect theory. Had the experiments been carried out correctly and found the falsifications claimed, these findings still would not have been new. More serious violations of the same kind, taking away novelty (in particular, regarding equalizing reductions), have been reported before, as well as many other negative findings. Nevertheless, rank dependence and prospect theory remain the most popular risk theory today because of many more positive results and, importantly, the absence of a viable alternative. In particular, SB's suggested alternatives of rank-independent weighting, better known as separable prospect theory, and "complexity aversion" (a misnomer), based on incorrect theoretical and empirical claims, have been argued not to work for decades.

It would have been surprising if SB had been the first to "properly" test rank dependence, 40 years after its introduction (Quiggin, 1982; Schmeidler, 1982), 30 years after its incorporation into prospect theory (Tversky & Kahneman 1992), 20 years after its shared prize in memory of Nobel in 2002, and after thousands of applications.

Data availability

No data was used for the research described in the article.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.socec.2022.101950.

References

Abdellaoui, M., Li, C., Wakker, P. P., & Wu, G. (2020). A defense of prospect theory in Bernheim & Sprenger's experiment. *Mimeo*. https://personal.eur.nl/wakker/pdf/abd.li.wak.wu_bernh.sp.pdf.

Ariely, D., Loewenstein, G. F., & Prelec, D. (2001). 'Coherent arbitrariness': Stable demand curves without stable preferences. *Quarterly Journal of Economics, 118*, 73–106. https://doi.org/10.1162/00335530360535153

Armantier, O., & Treich, N. (2016). The rich domain of risk. *Management Science, 62*, 1954–1969. https://doi.org/10.1287/mnsc.2015.2215

Barberis, N. C. (2013). Thirty years of prospect theory in economics: A review and assessment. *Journal of Economic Perspectives, 27*, 173–195. https://doi.org/10.1257/jep.27.1.173

Bateman, I. J., Munro, A., Rhodes, B., Starmer, C., & Sugden, R. (1997). Does part-whole bias exist? An experimental investigation. *Economic Journal, 107*, 322–332. https://doi.org/10.1111/j.0013-0133.1997.160.x

Bernasconi, M. (1994). Nonlinear preference and two-stage lotteries: Theories and evidence. *Economic Journal, 104*, 54–70. https://doi.org/10.2307/2234674

Bernheim, B. D., Royer, R., & Sprenger, C. (2022). Robustness of rank independence in risky choice. *AEA Papers and Proceedings, 112*, 415–420. https://doi.org/10.1257/pandp.20221090

Bernheim, B. D., & Sprenger, C. (2020). On the empirical validity of cumulative prospect theory: Experimental evidence of rank-independent probability weighting. *Econometrica, 88*, 1363–1409. https://doi.org/10.1257/pandp.20221090

Birnbaum, M. H. (2004). Causes of Allais common consequence paradoxes: An experimental dissection. *Journal of Mathematical Psychology, 48*, 87–106. https://doi.org/10.1016/j.jmp.2004.01.001

Birnbaum, M. H. (2005). Three new tests of independence that differentiate models of risky decision making. *Management Science, 51*, 1346–1358. https://doi.org/10.1287/mnsc.1050.0404

Birnbaum, M. H. (2007). Tests of branch splitting and branch-splitting independence in Allais paradoxes with positive and mixed consequences. *Organizational Behavior and Human Decision Processes, 102*, 154–173. https://doi.org/10.1016/j.obhdp.2006.04.004

Birnbaum, M. H. (2008). New paradoxes of risky decision making. *Psychological Review, 115*, 463–501. https://doi.org/10.1037/0033-295X.115.2.463

Birnbaum, M. H., & McIntosh, W. R. (1996). Violations of branch independence in choices between gambles. *Organizational Behavior and Human Decision Processes, 67*, 91–110. https://doi.org/10.1006/obhd.1996.0067

Birnbaum, M. H., Schmidt, U., & Schneider, M. D. (2017). Testing independence conditions in the presence of errors and splitting effects. *Journal of Risk and Uncertainty, 54*, 61–85. https://doi.org/10.1007/s11166-017-9251-5

Blondel, S. (2002). Testing theories of choice under risk: Estimation of individual functionals. *Journal of Risk and Uncertainty, 24*, 251–265. https://doi.org/10.1023/A:1015687502895

Bruce, A. C., & Johnson, J. E. V. (1996). Decision-making under risk: The effect of complexity on performance. *Psychological Reports, 79*, 67–76. https://doi.org/10.2466/pr0.1996.79.1.67

Camerer, C. F., & Ho, T. H. (1994). Violations of the betweenness axiom and nonlinearity in probability. *Journal of Risk and Uncertainty, 8*, 167–196. https://doi.org/10.1007/BF01065371

Cohen, M., & Jaffray, J. Y. (1988). Certainty effect versus probability distortion: An experimental analysis of decision making under risk. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 554–560. https://doi.org/10.1037/0096-1523.14.4.554

Dean, M., & Ortoleva, P. (2017). Allais, Ellsberg, and preferences for hedging. *Theoretical Economics, 12*, 377–424. https://doi.org/10.3982/TE1960Ci

Diecidue, E., Wakker, P. P., & Zeelenberg, M. (2007). Eliciting decision weights by adapting de Finetti's betting-odds method to prospect theory. *Journal of Risk and Uncertainty, 34*, 179–199. https://doi.org/10.1007/s11166-007-9011-z

Edwards, K. D. (1996). Prospect theory: A literature review. *International Review of Financial Analysis, 5*, 18–38. https://doi.org/10.1016/S1057-5219(96)90004-6

Erev, I., Ert, E., Plonsky, O., Cohen, D., & Cohen, O. (2017). From anomalies to forecasts: Toward a descriptive model of decisions under risk, under ambiguity, and from experience. *Psychological Review, 124*, 369–409. https://doi.org/10.1037/rev0000062

Fehr-Duda, H., & Epper, T. (2012). Probability and risk: Foundations and economic implications of probability-dependent risk preferences. *Annual Review of Economics, 4*, 567–593.

Fennema, H., & Wakker, P. P. (1996). A test of rank-dependent utility in the context of ambiguity. *Journal of Risk and Uncertainty, 13*, 19–35.

Fennema, H., & Wakker, P. P. (1997). Original and cumulative prospect theory: A discussion of empirical differences. *Journal of Behavioral Decision Making, 10*, 53–64.

Fishburn, P. C. (1978). On Handa's "new theory of cardinal utility" and the maximization of expected return. *Journal of Political Economy, 86*, 321–324. https://doi.org/10.1086/260670

Fudenberg, D., & Puri, I. (2022). Simplicity and probability weighting in choice under risk. *AEA Papers and Proceedings, 112*, 421–425. https://doi.org/10.1257/pandp.20221091

Gonzalez, R., & Wu, G. (2022). Composition rules in original and cumulative prospect theory. *Theory and Decision, 92*, 647–675. https://doi.org/10.1007/s11238-022-09873-0

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., et al. (2016). Statistical tests, p values, confidence intervals, and power: A guide to misinterpretations. *European Journal of Epidemiology, 31*, 337–350. https://doi.org/10.1007/s10654-016-0149-3

Hirshman, S.D., & Wu, G. (2022). Tests of rank-dependent probability weighting in risky choice. in preparation.

Huck, S., & Weizsäcker, G. (1999). Risk, complexity, and deviations from expected-value maximization: Results of a lottery choice experiment. *Journal of Economic Psychology, 20*, 699–715. https://doi.org/10.1016/S0167-4870(99)00031-8

Humphrey, S. J. (1995). Regret aversion or event-splitting effects? More evidence under risk and uncertainty. *Journal of Risk and Uncertainty, 11*, 263–274. https://doi.org/10.1007/BF01207789

Humphrey, S. J. (2000). The common consequence effect: Testing a unified explanation of recent mixed evidence. *Journal of Economic Behavior and Organization, 41*, 239–263. https://doi.org/10.1016/S0167-2681(99)00075-X

Humphrey, S. J. (2001a). Are event-splitting effects actually boundary effects? *Journal of Risk and Uncertainty, 22*, 79–93. https://doi.org/10.1023/A:1011105607742

Humphrey, S. J. (2001b). Non-transitive choice: Event-splitting effects or framing effects? *Economica, 68*, 77–96. https://doi.org/10.1111/1468-0335.00234

Humphrey, S. J. (2006). Does learning diminish violations of independence, coalescing and monotonicity? *Theory and Decision, 61*, 93–128. https://doi.org/10.1007/s11238-006-8047-x

Humphrey, S. J., & Verschoor, A. (2004). Decision-making under risk among small farmers in East Uganda. *Journal of African Economies, 13*, 44–101. https://doi.org/10.1016/j.econlet.2004.02.015

Iyengar, S.S., Jiang, W., & Huberman, G. (2003). How much choice is too much: Contributions to 401(k) Retirement plans. Pension Research Council Working Paper 2003-10.

Iyengar, S. S., & Kamenica, E. (2010). Choice proliferation, simplicity seeking, and asset allocation. *Journal of Public Economics, 94*, 530–539. https://doi.org/10.1016/j.jpubeco.2010.03.006

Iyengar, S. S., & Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing?". *Journal of Personality and Social Psychology, 79*, 995–1006. https://doi.org/10.1037/0022-3514.79.6.995

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*, 263–291. https://doi.org/10.1007/BF00122574

Kovarik, J., Levin, D., & Wang, T. (2016). Ellsberg paradox: Ambiguity and complexity aversions compared. *Journal of Risk and Uncertainty, 52*, 47–64. https://doi.org/10.1007/s11166-016-9232-0

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *"Foundations of measurement, vol. i (Additive and polynomial representations)*. New York: Academic Press (2nd edn. 2007, New York, Dover Publications.).

Levy, H. (2008). First degree stochastic dominance violations: Decision weights and bounded rationality. *Economic Journal, 118*, 759–774. https://doi.org/10.1111/j.1468-0297.2008.02141.x

L'Haridon, O. (2009). Behavior in the loss domain: An experiment using the probability trade-off consistency condition. *Journal of Economic Psychology, 30*, 540–551. https://doi.org/10.1016/j.joep.2009.03.007

l'Haridon, O., & Placido, L. (2010). Betting on Machina's reflection example: An experiment on ambiguity. *Theory and Decision, 69*, 375–393. https://doi.org/10.1007/s11238-008-9128-9

l'Haridon, O., & Vieider, F. (2019). All over the map: A worldwide comparison of risk preferences. *Quantitative Economics, 10*, 185–215. https://doi.org/10.3982/QE898

Loehman, E. (1998). Testing risk aversion and nonexpected utility theories. *Journal of Economic Behavior and Organization, 33*, 285–302. https://doi.org/10.1016/S0167-2681(97)00097-8

Loomes, G., Starmer, C., & Sugden, R. (2003). Do anomalies disappear in repeated markets? *Economic Journal, 113*, C153–C166. https://doi.org/10.1111/1468-0297.00108

Lopes, L. L., & Oden, G. C. (1999). The role of aspiration level in risky choice: A comparison of cumulative prospect theory and SP/A theory. *Journal of Mathematical Psychology, 43*, 286–313. https://doi.org/10.1006/jmps.1999.1259

Luce, R. D. (2000). *Utility of gains and losses: measurement-theoretical and experimental approaches*. London: Lawrence Erlbaum Publishers.

Luce, R.D., & Suppes, P. (1965). Preference, utility, and subjective probability. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), Handbook of mathematical psychology, Vol. III, 249–410. New York, Wiley.

Machina, M. J. (2009). Risk, ambiguity, and the rank-dependence axioms. *American Economic Review, 99*, 385–392. https://doi.org/10.1257/aer.99.1.385

Mador, G., Sonsino, D., & Benzion, U. (2000). On complexity and lotteries' evaluations — Three experimental observations. *Journal of Economic Psychology, 21*, 625–637. https://doi.org/10.1016/S0167-4870(00)00023-4

Moffatt, P. G., Sitzia, S., & Zizzo, D. J. (2015). Heterogeneity in preferences towards complexity. *Journal of Risk and Uncertainty, 51*, 147–170. https://doi.org/10.1007/s11166-015-9226-3

Murphy, R. O., & ten Brincke, R. H. W. (2018). Hierarchical maximum likelihood parameter estimation for cumulative prospect theory: Improving the reliability of individual risk parameter estimates. *Management Science, 64*, 308–326. https://doi.org/10.1287/mnsc.2016.2591

Neilson, W. S. (1992). Some mixed results on boundary effects. *Economics Letters, 39*, 275–278. https://doi.org/10.1016/0165-1765(92)90260-6

Nilsson, H., Rieskamp, J., & Wagenmakers, E. J. (2020). Commentary: Hierarchical Bayesian parameter estimation for cumulative prospect theory. *Journal of Mathematical Psychology, 98*, Article 102429. https://doi.org/10.1016/j.jmp.2020.102429

Pachur, T., Schulte-Mecklenbeck, M., Murphy, R. O., & Hertwig, R. (2018). Prospect theory reflects selective allocation of attention. *Journal of Experimental Psychology: General, 147*, 147–169. https://doi.org/10.1037/xge0000406

Paradís, J., Viader, P., & Bibiloni, L. (2001). The derivative of Minkowski's ?(x) function. *Journal of Mathematical Analysis and Applications, 253*, 107–125.

Peterson, J. C., Bourgin, D. D., Agrawal, M., Reichman, D., & Griffiths, T. L. (2021). Using large-scale experiments and machine learning to discover theories of human decision-making. *Science, 372*, 1209–1214. https://www.science.org/doi/10.1126/science.abe2629.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behaviour and Organization, 3*, 323–343. https://doi.org/10.1016/0167-2681(82)90008-7

Ramsey, F.P. (1931). Truth and probability. In R. B. Braithwaite (ed.), The foundations of mathematics and other logical essays, 156–198. London, Routledge and Kegan Paul. Reprinted in H.E. Kyburg Jr. & H.E. Smokler (1964, eds.) Studies in subjective probability, 61–92. New York, Wiley. (2nd edn. 1980, New York, Krieger.).

Rieger, M. O., & Wang, M. (2008). Prospect theory for continuous distributions. *Journal of Risk and Uncertainty, 36*, 83–102. https://doi.org/10.1007/s11166-007-9029-2

Ruggeri, K., Alí, S., Berge, M. L., et al. (2020). Replicating patterns of prospect theory for decision under risk. *Nature Human Behavior, 4*, 622–633. https://doi.org/10.1038/s41562-020-0886-x

Samuelson, P. A. (1960). The St. Petersburg paradox as a divergent double limit. *International Economic Review, 1*, 31–37. https://doi.org/10.2307/2525406

Schmeidler, D. (1982). *Subjective probability without additivity. foerder institut of economic research*. Israel: Tel Aviv University, Tel Aviv.

Schmeidler, D. (1989). Subjective probability and expected utility without additivity. *Econometrica, 57*, 571–587. https://doi.org/10.2307/1911053

Schmidt, U., & Seidl, C. (2014). Reconsidering the common ratio effect: The roles of compound independence, reduction, and coalescing. *Theory and Decision, 77*, 323–339. https://doi.org/10.1007/s11238-014-9456-x

Schneider, S. L., & Lopes, L. L. (1986). Reflection in preferences under risk: Who and when may suggest why. *Journal of Experimental Psychology: Human Perception and Performance, 12*, 535–548. https://doi.org/10.1007/s11238-014-9456-x

Smith, V. L. (1982). Microeconomic systems as an experimental science. *American Economic Review, 72*, 923–955. https://www.jstor.org/stable/1812014.

Sonsino, D., Benzion, U., & Mador, G. (2002). The complexity effects on choice with uncertainty—Experimental evidence. *Economic Journal, 112*, 936–965. https://doi.org/10.1111/1468-0297.00073

Sonsino, D., & Mandelbaum, M. (2001). On preference for flexibility and complexity aversion: Experimental evidence. *Theory and Decision, 51*, 197–216. https://doi.org/10.1023/A:1015555026870

Starmer, C. (1999). Cycling with rules of thumb: An experimental test for a new form of non-transitive behavior. *Theory and Decision, 46*, 141–158. https://doi.org/10.1023/A:1004930205037

Starmer, C. (2000). Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature, 38*, 332–382. https://doi.org/10.1257/jel.38.2.332

Starmer, C., & Sugden, R. (1993). Testing for juxtaposition and event-splitting effects. *Journal of Risk and Uncertainty, 6*, 235–254. https://doi.org/10.1007/BF01072613

Stodder, J. (1997). Complexity aversion: Simplification in the Herrnstein and Allais behaviors. *Eastern Economic Journal, 23*, 1–15. https://www.jstor.org/stable/40325750.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*, 297–323. https://doi.org/10.1007/BF00122574

Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review, 101*, 547–567. https://doi.org/10.1037/0033-295X.101.4.547

Wakker, P. P. (2010). *Prospect theory: For risk and ambiguity*. Cambridge, UK: Cambridge University Press.

Wakker, P.P. (2022)., "Annotated bibliography. http://personal.eur.nl/wakker/refs/webrfrncs.docx.

Wakker, P. P. (2023). The correct formula of 1979 prospect theory for multiple outcomes. *Theory and Decision, 94*, 183–187. https://doi.org/10.1007/s11238-022-09885-w

Wakker, P. P., Erev, I., & Weber, E. U. (1994). Comonotonic independence: The critical test between classical and rank-dependent utility theories. *Journal of Risk and Uncertainty, 9*, 195–230. https://doi.org/10.1007/BF01064200

Weber, B. J. (2007). The effects of losses and event splitting on the Allais paradox. Judgement in decision making. *Judgment and Decision Making, 2*, 115–125.

Weber, E. U., & Kirsner, B. (1997). Reasons for rank-dependent utility evaluation. *Journal of Risk and Uncertainty, 14*, 41–61. https://doi.org/10.1023/A:1007769703493

Weber, M., Eisenführ, F., & von Winterfeldt, D. (1988). The effects of splitting attributes on weights in multiattribute utility measurement. *Management Science, 34*, 431–445. https://doi.org/10.1287/mnsc.34.4.431

Wilcox, N. T. (1993). Lottery choice: Incentives, complexity and decision time. *Economic Journal, 103*, 1397–1417. https://doi.org/10.2307/2234473

Wu, G. (1994). An empirical test of ordinal independence. *Journal of Risk and Uncertainty, 9*, 39–60. https://doi.org/10.1007/BF01073402

Wu, G., Zhang, J., & Abdellaoui, M. (2005). Testing prospect theories using tradeoff consistency. *Journal of Risk and Uncertainty, 30*, 107–131. https://doi.org/10.1007/s11166-005-6561-9