Online appendix of

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

by
Peter P. Wakker
Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, the Netherlands, Wakker@ese.eur.nl
October 2022

Appendix OA.1 replaces the deterministic analysis in §2.1 of Bernheim & Sprenger (2020, SB) by a probabilistic analysis. It shows that stable and systematic patterns in SB must reflect heuristics and cannot reflect true preferences because SB's test statistics for true preferences have essentially a Cauchy distribution. Appendix OA.2. shows that SB's main experiment tested rank-dependence where it has been known not to be strong. Appendix OA.3 shows that, contrary to SB's suggestion, complexity aversion cannot help in precluding the undesirable violations of stochastic dominance of SB's rank-independent weighting. Appendix OA.4 lists 15 mistakes by SB not discussed in the main text. The lengthy Appendix OA.5 (pp. 11-56) lists the 59 references in Wakker (2022a) that have a keyword "PT falsified," with the annotations kept. Finally, appendix OA.6 gives the references for Appendixes OA.1-OA.4.

OA.1. A statistical problem if noise is considered in SB's §2.1

Bernheim & Sprenger (SB), in their §2.1, reported a deterministic CPT preference analysis of their stimuli that uses ratios $\frac{k}{k'}$. Here k and k' are small in an absolute sense relative to the other numbers in the stimuli. We note here that $18 - k$ and $18 - k'$ were the values actually elicited. Small relative errors in these give large relative errors in k, k'. Hence, ratios $\frac{k}{k'}$ are very vulnerable to noise, as with the

Cauchy distribution. As SB emphasized throughout, it is important to reckon with noise beyond a deterministic analysis. It would indeed have been of great interest to analyze the role of plausible noise *in preferences* in their §2.1. Adding an error term to their CPT values (as in SB's Eq. 5 and Footnote 60) affects the certainty equivalents of the overall lotteries by some dollars. Given the complexity of their three-outcome lotteries and Ramsey's trifle problem, such errors in preference values are plausible. This leads to errors in the measured k, k' that may readily make them approximate 0 (no negative answers were possible). If such errors occur with probabilities exceeding 0.05, then the confidence intervals of the ratios $\frac{k}{k'}$ span the whole $\mathbb{R}^+$. Then SB's analysis will lack the statistical power to reject any hypothesis *about preference*, be it rank dependence or rank independence.

The above statistical analysis was back-of-the-envelope, for illustrative purposes. It shows that a fully elaborated power analysis, based on adding plausible error models *for preferences* to the calibrated CPT models used throughout SB's paper, would have provided useful insights. It would have shown if the claim in their abstract "Conventional calibrations of CPT preferences imply that the percentage change in probability weights should be an order of magnitude larger than we observe," and claimed nonoverlapping confidence intervals (SB p. 1366 middle; p. 1382 l. 12; p. 1388), can hold statistically for realistic (noisy) preference calibration models. It would also show whether the variances found in their data may at all represent preferences rather than heuristics, as I argue.

The data that SB found did not exhibit the volatility just suggested. SB obtained stable patterns giving statistical power and tight confidence intervals. However, this is an extra problem for them, because it further shows that their experiment did not measure preferences, and did not, even in an as-if sense, speak to our Eqs. 2 or 3. Instead, subjects faced with hundreds of choices of complex and nearly-identical lotteries, for a one-time trifle reward received with some probability, develop simple algebraic heuristics, and this gave SB's stable findings. The stability found must reflect coherent arbitrariness. That is, multiple repetitions of complex tasks can lead to stable but invalid patterns, in our case heuristics instead of preferences. This has been shown in many studies. Ariely, Loewenstein, & Prelec (2001) called it coherent arbitrariness, while Loomes, Starmer, & Sugden (2003) called it the shaping hypothesis. See also Baron et al. (2001 p. 3 l. -2), Carlin (1992 p. 219), Dolan & Stalmeier (2003), and Hardisty et al. (2013).

OA.2. Weakness of rank-dependence in longshot effect

In their main experiment, SB considered changes in decision weights only when the rank changes from middle to best. It is well-known that rank dependence is not strong there (DWZ p. 185 ll. 4-6; DWZ p. 197 l. 7). Stronger rank dependence occurs when ranks change from middle to worst, consistent with the certainty effect. As for the change in rank considered by SB's first experiment, quite some studies found that the increase of decision weight is weak or absent there. Even, several studies found an effect oppositge to the prevailing finding, decreasing iso increasing decision weight, consistent with pessimistic probability weighting. See van de Kuilen & Wakker (2011). Their Footnotes 7 & 8 survey many papers that found this opposite effect. In view of this literature, finding neutrality (no effect) of rank dependence in SB's first experiment is no surprise anyhow. It does not provide new insights into rank dependence.

OA.3. Complexity aversion does not preclude violations of stochastic dominance

SB suggested an integration of rank-independent probability weighting and complexity aversion, but did not specify how this would be. This makes it hard to verify (or concretely criticize) their claims. However, the following observation leaves little hope that complexity aversion could remedy against violations of stochastic dominance in any way. The idea is, in words, that every violation of stochastic dominance can be replicated arbitrarily closely using only lotteries with the same number of outcomes, so that complexity aversion is kept fixed. Fudenberg & Puri (2022) did specify an integration of rank-independent probability weighting and the following analysis applies to their model.

OBSERVATION OA.3.1. Complexity aversion cannot help to reduce violations of stochastic dominance.

PROOF. All what follows holds both for separable prospect theory and for 1979 prospect theory, and we write PT for both. We will use continuity of π. Assume a rank-independent probability weighting model that applies only to all N-outcome lotteries. Assume the model would give an "inexplicable" (SB's term) violation of stochastic dominance between an m-outcome lottery and an n-outcome lottery for some $m \leq N$, $n \leq N$ if it could have been applied also to those m-outcome and n-outcome lotteries. Then this violation can be replicated arbitrarily closely using only two N-outcome lotteries, by approximating the m-outcome and n-outcome lotteries by N-outcome lotteries whose PT values also converge. For instance, for the m outcome lottery, we take an outcome $x_j \neq 0$ received with probability $p_j > 0$ and we reduce its probability to $p_j - \varepsilon$. We next bring in $N - m$ new outcomes $x_j(1 + i \times \delta)$ ($i = 1, \dots, N - m$), each with probability $\frac{\varepsilon}{N-m}$. We take $\delta > 0$ and, hence, the non-weighted outcome (the one closest to the reference point) of the lottery remains the same. If δ is sufficiently small then all outcomes are distinct, giving an N outcome lottery. For $0 < \varepsilon < p_j$ and ε and δ tending to 0, the PT value of the N outcome lotteries tends to that of the m outcome lottery, using continuity of v and π. We can similarly approximate the n-outcome lottery. In words: every violation of stochastic dominance between two lotteries with different numbers of outcomes can be replicated using only lotteries with the same number of outcomes. Thus, complexity aversion cannot serve well to preclude violations of stochastic dominance.

A numerical illustration: take the example in our §2, and assume the rank-independent functional there only for lotteries with exactly 100 outcomes. Then we replace the sure outcome 6.90 in the example by the following probability distribution with 100 outcomes:

$$\left(\frac{\varepsilon}{99}: 6.9 \times (1 + 1 \times \delta), \frac{\varepsilon}{99}: 6.9 \times (1 + 2 \times \delta), \ldots, \frac{\varepsilon}{99}: 6.9 \times (1 + 99 \times \delta), 1 - \varepsilon: 6.9\right)$$

For $\varepsilon > 0$ and $\delta > 0$ very small the value of this lottery is extremely close to $v(6.9)$. For sufficiently small $\varepsilon > 0$ and $\delta > 0$, say 10^{-10}, it is exceeded by the value of the following minor improvement of the lottery displayed above:

$$(0.01: 1.1 + 1 \times 10^{-5}, \ldots, 0.01: 1.1 + 99 \times 10^{-5}, 0.01: 0).$$

The maximal outcome of the latter lottery is exceeded by a factor of more than 6 by the minimal outcome of the former lottery, but yet it is preferred. This is, again, an absurd violation of stochastic dominance. We have essentially replicated the violation of §2.2 using only lotteries with exactly 100 outcomes and proper approximations. That is, the absurd violation of §2 remains. Complexity aversion cannot preclude it, and offers no remedy. It cannot save rank-independent probability weighting. □

OA.4. Further inaccuracies in Bernheim & Sprenger (2020)

This online appendix lists 15 inaccuracies by SB not mentioned in the main text because they do not directly affect the main conclusions.

Mathematical mistakes:

OA.4.1. [COMONOTONIC INDEPENDENCE]

SB p. 1376 ℓ. 8: Schmeidler's (1989) comonotonic independence is different than what SB claimed. For instance, it involves a mixture operation.

OA.4.2. [K INDEPENDENT OF X]

P. 1367, Footnote 7: SB in fact needed linear utility. Then, contrary to SB's claims, k does NOT depend on X there under PT and EU, and neither does it under rank dependence as long as ranks are kept fixed (comonotonicity), as follows from Eq. 5 in the main text.

OA.4.3. [$p = 1$ FOR COMMON RATIO]

P. 1390 ℓ. 10: the common ratio effect is only strong, and often only defined, with probability $p = 1$ involved. Nonlinearity of w in [0.9, 1] does accommodate this, contrary to SB's claim.

Next follow incorrect citations by SB. The mistakes always go in the direction of downplaying other contributions.

Incorrect claims on prospect theory:

The next three mistakes, together with footnote 5 in the main text, show that almost every sentence in SB's footnotes 3 and 4 on prospect theory is wrong.

OA.4.4. [ONLY ONE VERSION OF 1979 PROSPECT THEORY]

SB's Footnote 3: "Kahneman and Tversky (1979) actually provided two formulations of Prospect Theory"

Incorrect. There is only one (Wakker 2022b).

OA.4.5. [EXPLICIT!]

SB's Footnote 3: "They implicitly invoked the same assumption [their Equation 1] when examining the Allais common consequence paradox (p. 282)."

Incorrect citation. Kahneman & Tversky (1979) write it explicitly on p. 282 top.

OA.4.6. [AGAIN, ONLY ONE VERSION OF 1979 PROSPECT THEORY]

SB's Footnote 4: "Kahneman and Tversky also provided a formulation for two-outcome lotteries with either all positive or all negative outcomes that does indeed respect dominance (see, e.g., Equation (2) of Kahneman and Tversky (1979))."

Incorrect. Their Eq. 2 is part of the ONLY version of 1979 prospect theory (Wakker 2022b) and, as is well known, this does violate stochastic dominance.

Further Incorrect citations:

OA.4.7. [NO PROPER JUSTICE TO WEBER & KIRSNER]

Weber & Kirsner (1997) found significant rank dependence for the same kind of stimuli as considered by SB, providing straight counterevidence to SB. SB did not make this clear but only cited them ambiguously in Footnote 6.

OA.4.8. [IDENTIFIABILITY IN OTHER STUDIES]

P. 1382: "Tversky and Kahneman (1992) and Tversky and Fox (1995) obtained probability weighting parameters from certainty equivalents by parameterizing both the utility and probability weighting functions and assuming each observation satisfies the indifference condition $u(C) = \pi(p)u(25)$."

Wrong citation. Those papers used essentially richer stimuli. For the stimuli mentioned there and used by SB, the model is not even identifiable (see §2.3).

OA.4.9. [REAL INCENTIVES IN BIRNBAUM]

P. 1401 Footnote 69: "Interestingly, in incentivized tasks, we do not see the failure of coalescing noted by Birnbaum (2008) for hypothetical choice."

Wrong citation: Birnbaum used real incentives. His 2008 paper reviewed Birnbaum (2004), in particular, his Table 3. His §2 there explained that he used real incentives. Probably SB gambled on their incorrect claim to cover up the puzzling point that their finding is opposite to Birnbaum's (as it is, unbeknownst to them, opposite to most of the literature). Also note that Birnbaum (2008) extensively discussed what SB called complexity aversion, but they did not cite him for that, or the many other papers that Birnbaum cited on it.

Weak writings:

OA.4.10. [REFERENCE DEPENDENCE]

SB claimed to also falsify models with reference dependence, but these claims are incorrect for the same reasons as their claims about rank dependence are (incorrect formulas, unidentifiable estimates, bad stimuli, and so on).

OA.4.11. [Statistical analysis lacking for main claims]

P. 1399 last para of §5: "equalizing reductions respond *strongly* to changes in X" [italics added]

No statistical analysis is given to justify this claim. The confidence intervals in Figure 5B overlap, leaving unclear whether what SB called "strongly" is even significant. SB made the same unfounded claim of dependence on X on p. 1396 ℓ. -5/-2 and p. 1398 last sentence of §5.3. Further, their claimed explanation, through utility curvature, is implausible because utility curvature is weak for moderate payoffs as in their experiment.

Besides the above point (SB's third finding end of §5), the first two findings there ((1): nonzero impact of probability; (2): absence of complete randomness) concern trivial strawmen. Their whole claim of genuine effects in their 2nd experiment, needed to claim genuine absence of rank dependence, hinges on the above, unsubstantiated, claim.

OA.4.12. [UNFOUNDED SPECULATION]

P. 1380: "If isolation fails in this context, then our subjects would not exhibit standard patterns of probability weighting in binary tasks. [Then what else? Linear weighting???] Conversely, if our subjects do exhibit standard probability weighting patterns in binary tasks, then one cannot reasonably attribute the absence of implied discontinuities in the equalizing reduction tasks to a failure of isolation."

Unfounded speculations on what happens if isolation fails.

OA.4.13. [ASSUMED PROPERTIES u AND w?]

SB never wrote what properties w and u should have. Strictly increasing? Stochastic dominance? Continuous? Yet they used such properties. This is why I assume them explicitly below Eq. 1 in the main text.

OA.4.14. [APPLES VS. PEARS]

P. 1377: "because the essence of our approach is to measure characteristics of indifference curves (MRSs), all potential confounds associated with unintended variations in "distance to indifference" are eliminated."

A paraphrase: because we measure apples, all problems of pears are eliminated. Their measurements of indifferences do have the analogous problem. See §7. For example, if the errors in their indifference measurements are not constant or are extreme, then their claimed p-values and confidence intervals are not valid either.

OA.4.15. [NO USE REPORTING EXPERIMENT 1][1]

SB claimed that Experiment 2 would show absence of cancellation in Experiment 1, (p. 1367 end of first para: "clearly refuting the cancellation hypothesis.") contradicting the consensus in the field (Weber & Kirsner 1997) and unfounded. SB *only* justified Experiment 1 by referring to Experiment 2. Experiment 1 added nothing. Thus, one small Experiment 2 of 84 subjects (with no statistical analysis to support the main claim, see Mistake OA.4.11) should discard a Nobel-sharing theory used in 1000s of studies. SB's misleading claim was repeated in the last para of §5.3. Mistake OA.4.11 above showed how weak the evidence of their Experiment 2 in fact is.

[1] Given that the measurements of rank-independent weighting are not correct, all that remains of SB's experiments is the test of rank dependence.

OA.5. The 59 references of Wakker (2022a) with the keyword "PT falsified"

The annotated bibliograpy of Wakker (2022a) is instable in the sense of being updated every year. Therefore, this online appendix lists the 59 references in the version of 2022. The annotations, all including the keyword "PT falsified," are left above the references.

01

{% **real incentives/hypothetical choice**: find no difference in patterns, but less error for real incentives.

Do decision under risk both with monetary outcomes and with time as outcome. For time, subjects were told beforehand that the experiment would last approximately 2 hours, where it might be 1 or 3. The time unit designated a time to wait in the lab with no amusing/useful things like computers or mobile phones available. They were anchored to think 2 hours, but then it could become more (gains) or less (losses).

concave utility for gains, convex utility for losses: (§5.1) They find pronounced concavity for gains, and moderate concavity, and not convexity, for losses. For time less concavity for gains than for money. Loss aversion lower for time than for money (end of §5.1).

inverse-S: (§5.2) confirmed for time and money, and for gains and losses.

On average more inverse-S for time than for money, both for gains and for losses. For time, probability weighting has more elevation for both gains (optimism) and losses (pessimism). Which is not very nice for PT. Probability weighting depending on outcomes can be taken as a violation of PT (**PT falsified; probability weighting depends on outcomes**). The symmetry for gains and losses is nice for reflection. Would be interesting to see if at the individual level there is much difference between probability weighting for time and for money, but the paper does not report it. (Statistics may not be easy.)

losses from prior endowment mechanism: this they do. For money there is the usual problem that subjects may integrate the prior endowment with the loss and, hence, not perceive losses, which is why they do money only hypothetically, something that I agree with. For time such integration is less likely because time loss is not so easily integrated with the prior endowment OF MONEY (they are paid for the time loss). This makes this paper the most convincing implementation of real incentives for losses that I have seen in the literature so far. %}

Abdellaoui, Mohammed & Emmanuel Kemel (2014) "Eliciting Prospect Theory when Consequences Are Measured in Time Units: "Time Is not Money," *Management Science* 60, 1844–1859.

02

{% **PT falsified**: this paper shows that a majority prefers, with probabilities 1/4 not written, the lotteryt

(−1000, −800, 1200, 1600) to the lottery (−1000, −800, 800, 2000). The choice is a nice combination of choices considered in several revent papers by Levy & Levy but, contrary to the latter, the authors analyze the choice correctly, and establish a clear violation of PT. %}

Baltussen, Guido, Thierry Post, & Pim van Vliet (2006) "Violations of CPT in Mixed Gambles," *Management Science* 52, 1288–1290.

03

{% **PT falsified**: subjects have to do common-ratio choices, and others, not once, but repeatedly, say 200 times. They don't get any info about probabilities etc., only can push one of two buttons and from experience find out what probability distribution can be. They don't even know that it is one fixed probability distribution. Real incentives: they are paid in points, and in end sum total of points is converted to money. Loss aversion is confirmed. Other than that, all phenomena are opposite to prospect theory, with underweighting of **small probabilities**, anti-certainty effect, more risk seeking with gains than with losses, etc. A remarkable and original finding. The authors' explanation is that the subjects in their experiment experience the gambles rather than get descriptions of the gambles. It is surprising to me that subjects do not get close to expected value maximization.

My explanation (ex post indeed): the subjects put the question "which button would give the best outcome" central, and not "which button would give the best probability distribution over outcomes." They get to see which button gave best outcomes in most of the cases, with recency effect reinforcing it. Thus, subjects experience only the likelihood aspect, whether or not events with good/better outcomes obtain or not. The subjects do not experience the outcomes, because these are just abstract numbers to be experienced only after the experiment. This procedure leads to likelihood-oversensitivity, and S-shaped rather than **inverse-S** -shaped nonlinear measures. Example of recency effect: if subjects, for instance, remember only which option gave the best result on the last trial, then they choose the event that with highest probability gives the best outcome (a heuristic advanced by Blavatskyy). Outcomes will be perceived as ordinal more than as cardinal. The authors themselves may have alluded to this explanation on p. 221 just above Experiments 3a and 3b, when they refer to MacDonald, Kagel, & Battalio (1991, EJ) who found the opposite of what they found in an experiment with animals:

"For example, MacDonald et al. used a within-subject design and allowed the decision makers to immediately consume their rewards." %}

Barron, Greg & Ido Erev (2003) "Small Feedback-Based Decisions and Their Limited Correspondence to Description-Based Decisions," *Journal of Behavioral Decision Making* 16, 215–233.

04

{% experimental testing of, a.o., Ido & I.;

real incentives/hypothetical choice: p. 45 shows that there is a quantitative difference (more risk aversion for real incentives, both for gains and for losses) but the qualitative phenomena are the same. P. 28 also states this.

losses from prior endowment mechanism: seem to do this. Their Table 3 seems to find significant deviation from integration.

Risk averse for gains, risk seeking for losses: find what they call qualified support.

reference-dependence test: test and find it confirmed in §3.1 (p. 31). That is, they find asset integration falsified.

P. 32: less risk seeking for losses than risk aversion for gains.

PT falsified: p. 35: **risk seeking for symmetric fifty-fifty gambles**: they find it for (0.5, 20; 0.5, −20). %}

Battalio, Raymond C., John H. Kagel, & Komain Jiranyakul (1990) "Testing between Alternative Models of Choice under Uncertainty: Some Initial Results," *Journal of Risk and Uncertainty* 3, 25–50.

05

{% Violations of betweenness and also of mixture symmetry of quadratic utility;

RDU better, "Squiggle Hypothesis" for **probability triangle** supports **inverse-S** weighting functions; intersection point, however, seems to be below .16 iso .33. That is, at .16 their observations already suggest convex probability transformation; leads him to question RDU.

Real incentives: the **random incentive system** was used.

second-order probabilities; **backward induction/normal form, descriptive**: shows that **RCLA** is violated more than compound independence and, therefore, gives evidence in favor of backward induction/backward induction.

PT falsified: original prospect theory of 1979 is violated. %}

Bernasconi, Michele (1994) "Nonlinear Preference and Two-stage Lotteries: Theories and Evidence," *Economic Journal* 104, 54–70.

06

{% **PT falsified**: this paper claims to find that, but I disagree.

SECTION 1. INTRODUCTION

This paper, abbreviated BS henceforth, criticizes rank dependence, introduced by Quiggin (1982) for risk, and independently by Schmeidler (1989) for uncertainty. Rank dependence is central in Tversky & Kahneman's (1992) new prospect theory and many of my works. I co-authored a criticism of BS, at http://personal.eur.nl/wakker/refs/pdf/bscritic/abd.li.wak.wu_bernh.sp_linenrs20aug2020.pdf abbreviated AL henceforth. Thus, I am not a neutral commentator here. I think that BS is very weak, and damaging to the field.

As everyone will guess, AL was submitted to Econometrica, and, as will be clear, it was rejected. Given that I still maintain all the criticisms expressed by AL here, it is also clear that I disagree with all of Econometrica's objections to AL: they did not provide any serious counterargument. Now that Econometrica has let Bernheim & Sprenger publish an incorrect formula of the Nobel-awarded 1979 prospect theory, and has refused to correct it, what else can one do than warn people so as to minimize damage? The same holds for Berrnheim & Sprenger's incorrect identification of an unidentifiable functional, their incorrect claim of invality of general statistical counting tests, their attempt to revive separable prospect theory that was properly abandoned in the 1980s because of not just violating stochastic dominance but absurdly violating stochastic dominance, their unfounded pushing of their misnomer complexity aversion that was empirically rejected decades ago, their incorrect priority claims, and their other mistakes below. When Nilsson, Rieskamp, & Wagenmakers discovered that their 2011 paper in *Journal of Mathematical Psychology* used an incorrect formula of 1979 prospect theory, this alone was enough reason for the authors and journal to publish a correction in 2020. In this regard, Econometrica, Bernheim, and Sprenger behaved differently.

AL was written following academic conventions of diplomacy. Here, where I express subjective opinions on works, I can be more explicit and clear.

SECTION 2. ORGANIZATION

References below can be found in this bibliography. I will as much as possible use BS's notation and terminology, often reluctantly:

- "CPT" iso PT
- "rank-independent probability weighting": this term is uninformative, like non-elephant zoology. BS use it to refer to what is often called separable probability weighting ($\Sigma w(p_i)u(x_i)$). It was popular in psychology (Preston & Baratta 1948; Edwards 1962) until the 1980s, when it was abandoned mostly because Fishburn (1978 JPE) and others discovered that it violates stochastic dominance. BS try to revive this old psychological theory.
- "complexity aversion": that subjects supposedly have an aversion to many

outcomes, also for gains. The claim is empirically incorrect (see Mistake 3.8 below). The literature uses the term for phenomena other than dependency on nr. of outcomes, where the term is correct.

Next, three more sections follow.

SECTION 3. LIST OF BS'S MISTAKES DESCRIBED BY AL

BS claim a "*novel" falsification of CPT showing its "stunning failure*."
Mistakes:

3.1. [Ignoring priority of stronger counterevidence]
Even if BS's experiment had been correct, stronger violations of the same kind have been reported long before (and so have many different violations), ignored by BS, and invalidating their novelty claims. (AL §6.4)

3.2. [Ignoring ocean of positive evidence]
Many more positive results for CPT were obtained. One should look at the balance of all evidence ==> CPT most popular today. Even if BS had been correct, it would have been a very marginal contribution to an ocean of preceding evidence, ignored by BS, and invalidating their "failure of CPT" claims. (AL p. 16 ℓ. 6-12)

BS claim that *rank-independent probability weighting is better*. Mistakes:

3.3. [Misleading presentation of rank-independent probability weighting]
BS once acknowledge that rank-independent weighting violates stochastic dominance ("This is a serious flaw", BS p. 1364). But then the rest of their paper still presents it, misleadingly, as a promising alternative, apparently forgetting about the flaw, also prior to bringing in the (irrelevant; see below) complexity aversion (link to misleading citations from BS). BS are apparently not aware of the problematic absurdity, also descriptively, of the stochastic dominance violations (AL p. 4 ℓℓ. 10-16). The following Mistake 3.4 continues on this.

3.4. [Complexity aversion as incorrect remedy for Mistake 3.3]
BS incorrectly suggest complexity aversion as a remedy for the violations of stochastic dominance (BS end of §6). However, it is not; see AL §6.3. A less diplomatic and, hence, clearer, explanation is here (link). Thus, BS's suggested alternative for rank-dependent probability weighting does not work.

BS's *rank-independent probability weighting*, further mistakes:

3.5. [Wrong formula of prospect theory]

BS use an incorrect formula of 1979 prospect theory for rank-independent probability weighting (AL p. 3 ℓ.24 – p. 4 ℓ.7).

3.6. [Models not identifiable from their data]

The models that BS claim to estimate are not even identifiable from their data. (AL p. 4 ℓ. 25 - p. 5 ℓ. 11)

BS claim *novelty/usefulness of complexity aversion*; mistakes:

3.7. [Complexity aversion theoretically discarded long ago]

See AL §6.3 & AL Online Appendix (added at the end of the AL file.) This invalidates BS's novelty/interest claims.

3.8. [Complexity aversion empirically falsified long ago]

See AL Online Appendix p. 3 (added at the end of the AL file). This invalidates BS's empirical claims.

BS claim a *new general nonparametric measurement* of decision weights. Mistakes:

3.9. [Trifle problem]

Their preference measurement does not work because of Ramsey's trifle problem (payoff *differences* too small). (AL p. 7 ℓ. 20 - p. 8 ℓ. 27 & §5)

3.10. [Three-outcome lotteries are too complex]

It has often been pointed out that, in general, three-outcome lotteries, as used by BS, are too complex for subjects. Hence, all cited studies with three-outcome lotteries other than BS did special efforts, with special layouts and visual aids (AL p. 12 ℓℓ. 5-9). BS, unaware, did not do so.

3.11. [Linear utility]

The trifle problem can be avoided, but then linear utility is needed, invalidating BS's claims of generality and nonparametric analysis. (AL Assumption 1, p. 6 & p. 11 ℓ. 19 - p. 12 ℓ. 4)

3.12. [Further incorrect generality claim]

BS footnote 13, claiming validity even for nondifferentiable utility, is incorrect, and is based on a mathematical mistake. (AL p.6 Footnote 5)

3.13. [Invalid priority claim on measurement and test]

BS p. 1376 claims novelty: "However, our use of equalizing reductions has no counterpart in the existing literature." However, Diecidue, Wakker, & Zeelenberg (2007) used the (corrected) method before (AL p. 11 ℓ. 19 - p. 12 ℓ. 4) for uncertainty, which is more interesting than risk as in BS.

BS claim invalidity of statistical *counting tests*, used throughout all empirical sciences. Mistakes:

3.14. [Ignorance of randomness underlying statistical tests 1st]

BS do not know that *every* statistical test is based on an underlying probabilistic ("noise") model. (AL p. 13 ℓℓ. 14 -25)

3.15. [Ignorance of randomness underlying statistical tests 2nd]

BS's claimed first problem for counting tests only shows that *there exists* an error model under which counting tests are invalid. But this trivially holds for *every* statistical test, including all their own tests. (AL p. 13 ℓℓ. 30-33)

3.16. [Invalid no-power counterexample]

BS's claimed second problem for counting tests considers stimuli where EU and CPT make identical predictions. BS criticize counting tests for having no power then. But, again, this then trivially holds for *every* statistical test. (AL p. 14 ℓℓ. 9-19.)

SECTION 4. QUALIFICATIONS AND IMPLICATIONS OF THE PRECEDING BS MISTAKES

Elementary theoretical blunders: Mistakes 3.5 (wrong PT formula), 3.6 (nonidentifiability), 3.14 (randomness in statistics), and 3.16 (no-power)

Elementary experimental blunder: Mistake 3.9

Naive: Mistakes 3.1 and 3.2. Thinking, 30 years after the introduction of CPT, 20 years after its shared Nobel memorial prize, and after 1000s of applications, to be the first to ("properly") test one of its two main nonclassic components, is naive. Thinking that two (in fact only one; see Mistake 5.17 in §5 below) small experiments can speak final verdict, 30 years after, is so too. (Mistakes 3.14-3.16 are also naïve.)

Further: BS's other mistakes are more understandable, though still revealing lack of dedication/understanding & literature search/knowledge.

Damage: One can predict much damage to come from BS, augmented by the prominence of its outlet: use of incorrect formulas/measurement methods, investigations of things done before, wrong and useless separable probability weighting, rejections of papers using the currently best descriptive CPT, and so on.

SECTION 5. MISTAKES BY BS NOT MENTIONED IN AL

AL focused on BS's mistakes that were directly relevant for the main conclusions, and other mistakes whose mention could not be avoided (e.g., incorrect notation). However, having read their paper in detail, I know many more inaccuracies and weaknesses in BS, not mentioned by AL. I list such next.

Mathematical mistakes:

5.1. [Comonotonic independence]

BS p. 1376 ℓ. 8: Schmeidler's (1989) comonotonic independence is different than what BS claim. For instance, it involves a mixture operation.

5.2. [k independent of X]

P. 1367, Footnote 7: BS in fact need linear utility. Then, contrary to BS's claims, k does NOT depend on X there under PT and EU, and neither does it under rank dependence as long as ranks are kept fixed (comonotonicity), as follows from AL Eq. 8).

5.3. [$p = 1$ for common ratio]

P. 1390 ℓ. 10: the common ratio effect is only strong, and often only defined, with probability $p = 1$ involved. Nonlinearity of w in [0,9, 1] does accommodate this, contrary to BS's claim.

5.4. [brackets iso braces]

BS's notation of lotteries violates mathematical rules (AL Footnote 2). Braces denote sets that are not ordered and cannot be used here.

Further:

5.5. [Reference dependence]

BS claim to also falsify models with reference dependence, but these claims are incorrect for the same reasons as their claims about rank dependence are (wrong formulas, unidentifiable estimates, bad stimuli, and so on).

Incorrect citations:

5.6. [No proper justice to Weber & Kirsner]

Weber & Kirsner (1997) find significant rank dependence for the same kind of stimuli as considered by BS, providing straight counterevidence to BS. BS do not make this clear but only cite them ambiguously in Footnote 6.

5.7. [Identifiability in other studies]

P. 1382: "Tversky and Kahneman (1992) and Tversky and Fox (1995) obtained probability weighting parameters from certainty equivalents by parameterizing both the utility and probability weighting functions and assuming each observation satisfies the indifference condition $u(C) = \pi(p)u(25)$."

Wrong citation. Those papers used essentially richer stimuli. For the stimuli mentioned there and used by BS, the model is not even identifiable (see Mistake 1.6).

5.8. [Real incentives in Birnbaum]

P. 1401 Footnote 69: "Interestingly, in incentivized tasks, we do not see the failure of coalescing noted by Birnbaum (2008) for hypothetical choice."

Wrong citation: Birnbaum used real incentives. His 2008 paper reviews Birnbaum (2004), in particular, his Table 3. His §2 there explains that he used real incentives. Probably BS gambled on their incorrect claim to cover up the puzzling point that their finding is opposite to Birnbaum's (as it is, unbeknownst to them, to most of the literature). Also note that Birnbaum (2008) extensively discussed what BS call complexity aversion, but they do not cite him for that, or the many other papers Birnbaum cites on it.

The next four mistakes show that almost every sentence in BS's footnotes 3 and 4, on prospect theory, is wrong.

5.9. [Only one version of 1979 prospect theory]

BS's Footnote 3: "Kahneman and Tversky (1979) actually provided two formulations of Prospect Theory"

Incorrect. There is only one (AL Eq. 3).

5.10. [No wrong prospect theory formula in other papers]

BS's Footnote 3: "extensions which correspond to our three-outcome formulation are provided by, for example, Camerer and Ho (1994) and Fennema and Wakker (1997)":

Incorrect citations. See AL p. 17 $\ell\ell$. 25-32).

5.11. [Explicit!]

BS's Footnote 3: "They implicitly invoked the same assumption [their Equation 1] when examining the Allais common consequence paradox (p. 282)."

Incorrect citation. Kahneman & Tversky (1979) write it explicitly on p. 282 top.

5.12. [Again, only one version of 1979 prospect theory]

BS's Footnote 4: "Kahneman and Tversky also provided a formulation for two-outcome lotteries with either all positive or all negative outcomes that does indeed respect dominance (see, e.g., Equation (2) of Kahneman and Tversky (1979))."

Incorrect. Their Eq. 2 is part of the ONLY version of 1979 prospect theory and, as is well known, this does violate stochastic dominance.

Weak writings:

5.13. [Statistical analysis lacking for main claims]

P. 1399 last para of §5: "equalizing reductions respond *strongly* to changes in X" [italics added]

No statistical analysis is given to justify this claim. The confidence intervals in Figure 5B overlap, leaving unclear whether what BS call "strongly" is even significant. BS make the same unfounded claim of dependence on X on p. 1396 ℓ. -5/-2 and p. 1398 last sentence of §5.3. Further, their claimed explanation, through utility curvature, is implausible because utility curvature is weak for moderate payoffs as in their experiment.

Besides the above point (BS's third finding end of §5), the first two findings there ((1): nonzero impact of probability; (2): absence of complete randomness) concern trivial strawmen. Their whole claim of genuine effects in their 2nd experiment, needed to claim genuine absence of rank dependence, hinges on the above, unsubstantiated, claim.

5.14. [Unfounded speculation]

P. 1380: "If isolation fails in this context, then our subjects would not exhibit standard patterns of probability weighting in binary tasks. [Then what else? Linear weighting???] Conversely, if our subjects do exhibit standard probability weighting patterns in binary tasks, then one cannot reasonably attribute the absence of implied discontinuities in the equalizing reduction tasks to a failure of isolation."

Unfounded speculations on what happens if isolation fails.

5.15. [Assumed properties u and w?]

BS never say what properties w and u have. Strictly increasing? Stoch. dom? Continuous? Yet they use such properties. This is why AL assumes them explicitly below their Eq. 1.

5.16. [apples vs. pears]

P. 1377: "because the essence of our approach is to measure characteristics of indifference curves (MRSs), all potential confounds associated with unintended variations in "distance to indifference" are eliminated."

A paraphrase: because we measure apples, all problems of pears are eliminated. Their measurements of indifferences do have the analogous problem. See Mistake 3.15 in §3. For example, if the errors in their indifference measurements are not constant or are extreme, then their claimed p-values and confidence intervals are not valid either.

5.17. [No use reporting Experiment 1]

BS claim that Experiment 2 would show absence of cancellation in Experiment 1, (p. 1367 end of first para: "clearly refuting the cancellation hypothesis.") contradicting the consensus in the field (Weber & Kirsner 1997) and unfounded. BS *only* justify Experiment 1 by referring to Experiment 2. Experiment 1 adds nothing. Thus, one small Experiment 2 of 84 subjects (with no statistical analysis to support the main claim, see Mistake 3.13) should discard a Nobel-sharing theory used in 1000s of studies. BS's misleading claim is repeated in the last para of §5.3. Mistake 3.13 above showed how weak the evidence of their Experiment 2 in fact is.

5.18. [Complexity aversion is misnomer]

BS's complexity aversion for dependence on number of outcomes is a misnomer as explained at the beginning of these annotations and more in AL Online Appendix p. 2 ℓℓ. 17-32 & p. 1 ℓℓ. 17-21 (added at the end of AL). %}

Bernheim, B. Douglas & Charles Sprenger (2020) "On the Empirical Validity of Cumulative Prospect Theory: Experimental Evidence of Rank-Independent Probability Weighting," *Econometrica* 88, 1363–1409.

07

{% Real incentives: **random incentive system**.

PT falsified: Tables 5 and 6 give some violations of the s.th.pr. Here, after change of the common outcome, also one other outcome of one gamble is increased, whence preference reversals in one direction do not really violate the s.th.pr., but reversals in other direction do so strongly. The stimuli were so constructed that in each case most reversals were in the direction that entails strong violation of s.th.pr. In each case, all gambles could be considered comonotonic and it was also a violation of the comonotonic s.th.pr. The violations could simply be inconsistency were it not that the violations in one direction are significantly more frequent than in the other direction. So, violation of PT. Not violation of **inverse-S.** %}

Birnbaum, Michael H. (1999) "Testing Critical Properties of Decision Making of the Internet," *Psychological Science* 10, 399–407.

08

{% **coalescing**: as much evidence for complexity aversion (if splitting the lowest outcome) as for complexity seeking.

Real incentives: **random incentive system**;

An interesting decomposition of some things going on in the Allais paradox.

Finds violations of the s.th.pr. as in Birnbaum & McIntosh (1996), falsifying the **inverse-S** prob weighting of PT. (**PT falsified**)

P. 98 3rd para explains that splitting the best outcome improves, but splitting the worst worsens. Increasing weights nonnormalized, as in separable OPT, means that splitting gains always improves. Increasing weights normalized means that splitting lowest outcome worsens, also if gain. This is Birnbaum's models.

That salience of common outcome enhances s.th.pr om p. 94: "Event framing would be expected to reduce violations of branch independence in the split forms. Such choices might be termed ''transparent'' tests of branch independence in the framed form, because both gambles would clearly share a common event–consequence branch. In such a framed format, a decision-maker should find it easy to cancel branches that are identical in two choices and to make a choice based strictly on what is left." %}

Birnbaum, Michael H. (2004) "Causes of Allais Common Consequence Paradoxes: An Experimental Dissection," *Journal of Mathematical Psychology* 48, 87–106.

09

{% Branch independence is the sure-thing principle for events for which probability is also given.

PT falsified: evidence against **inverse-S**: finds violations of the s.th.pr. like Birnbaum & McIntosh (1996), falsifying the inverse-S prob weighting of PT;

real incentives: all choices were hypothetical

SEU = SEU: five lines below (1), and in the citation of Edwards in first paragraph of second column of p. 87;

biseparable utility %}

Birnbaum, Michael H. & Darin Beeghley (1997) "Violations of Branch Independence in Judgments of the Value of Gambles," *Psychological Science* 8, 87–94.

10

{% **PT falsified**: evidence against **inverse-S**

real incentives: all choices were hypothetical

Finds violations of the s.th.pr. like Birnbaum & McIntosh (1996), falsifying the inverse-S prob weighting of PT, also for four-outcome gambles distribution-independence is something of that kind, shifting probability mass from one common outcome to the other. Humphrey & Verschoor (2004) independently found the same. %}

Birnbaum, Michael H. & Alfredo Chavez (1997) "Tests of Theories of Decision Making: Violations of Branch Independence and Distribution Independence," *Organizational Behavior and Human Decision Processes* 71, 161–194.

11

{% **PT falsified**: evidence against **inverse-S**

Real incentives: it was all hypothetical choice;

Considers choices (R_1, R_2, C) versus (S_1, S_2, C), $R_1 > S_1 > S_2 > R_2$. PT with inverse-S predicts that there will be fewer risky choices as C increases. (If C increases from worst (< R_2) to intermediate (between S_1 and S_2) then inverse-S would have the decision weight of S_2 and R_2 increase, enhancing safe choice. If C increases from intermediate to highest (> R_1) then inverse-S would have the decision weight of S_1 and R_1 *decrease*, which *again* enhances risk aversion.) It is found, however, that there are more risky choices (in agreement, in fact, with Machina's fanning out). As the lotteries get better because of C increasing, people get more risk seeking rather than risk averse. See Table 1 where the percentage of safe choices decreases rather than increases as we move to the right. So, the extreme outcomes seem to be underweighted rather than overweighted.

The paper gives an extensive theoretical analysis. The most extensive tests are in Birnbaum & Navarrete (1998) (the main topic of which, by the way, is another), which also describes the other preceding evidence. In particular, the B&M paper considers only three equally likely outcomes, B&N considers richer probability triples.

P. 91 gives refs to people who argue that independence-tests are mixed up with other assumptions. %}

Birnbaum, Michael H. & William R. McIntosh (1996) "Violations of Branch Independence in Choices between Gambles," *Organizational Behavior and Human Decision Processes* 67, 91–110.

12

{% **PT: data on probability weighting**; **coalescing**;

PT falsified: evidence against **inverse-S**

Real incentives: it was all hypothetical choice;

evidence against inverse-S probability weighting, especially Table 4, see the comments in Birnbaum & McIntosh (1996).

coalescing: a systematic method for studying event splitting and the violations of stochastic dominance, the effect nicely illustrated by Tversky & Kahneman (1986, p. 178, problem 7). %}

Birnbaum, Michael H. & Juan B. Navarrete (1998) "Testing Descriptive Utility Theories: Violations of Stochastic Dominance and Cumulative Independence," *Journal of Risk and Uncertainty* 17, 49–78.

13

{% real incentives: RIS. **PT falsified**

Obtains systematic examples of reversed common ratio. If to choose between sure outcome and prospect with considerably higher EV, most choose the latter, risky, option. If then the probabilities of nonzero outcomes are scaled down by a common factor, many switch to a safe choice. For example, $60 \prec 100_{3/4}0$ (64.9%) but $60_{1/3}0 \succ 100_{1/4}0$ (67.1%). I wondered if some error theory could account for it, with simply more errors in the latter choice because then the options are more indifferent. But this does not work well because the paradoxical choices are majority choices. The finding $60_{1/3}0 \succ 100_{1/4}0$ (67.1%) is amazing and puzzling.

The paper considers some error theories but they cannot account for the finding. These findings violate every existing theory. %}

Blavatskyy, Pavlo R. (2010) "Reverse Common Ratio Effect," *Journal of Risk and Uncertainty* 40, 219–241.

14

{% **risk seeking for symmetric fifty-fifty gambles**: they don't have fifty-fifty gambles, but do find risk seeking for small amounts.

PT falsified.

Consider gains and losses, and probabilites 0.20 and 0.80 of getting the gain or loss.

Compare $\$80_{0.2}\$0$ and $-\$80_{0.2}\$0$). Can be done in two steps: step 1, translation by subtracting \$80, so that $\$80_{0.2}\$0$ is changed into $\$0_{0.2}-\$80$. Step 2, switching good- and bad-outcome probability, so that $\$0_{0.2}-\$80$ is changed into $\$0_{0.8}-\$80$.

They find that translation from gains to losses always increases risk seeking, both for high-probability and for low-probability for best outcome. They find that switching probability of bad outcome from 0.2 to 0.8 always increases risk seeking, both for gains and for losses.

Testing reflection for high-probability nonzero has translation and switch go in same direction, enhancing risk seeking for losses. Testing reflection for low-probability nonzero has translation and swiches go in opposite directions. In prospect theory, probability weighting and utility curvature have opposite effects for small-probability-nonzero-outcomes, although they both support the reflection effect because they both switch from gains to losses.

Consider also 7 different stakes. People are risk averse for high stakes and risk seeking for small, for high and low probabilities and for gains and losses (**probability weighting depends on outcomes**). Maybe some **Utility of gambling** generating the risk seeking for small amounts!? So that we may want to avoid small-amount prospects, considering this just a bias? %}

Bosch-Domènech, Antoni & Joaquim Silvestre (2006) "Reflections on Gains and Losses: A 2×2×7 Experiment," *Journal of Risk and Uncertainty* 33, 217–235.

15

{% Find risk seeking for small outcomes but risk aversion for large ones. A generalized logarithmic utility (ln (x + a)) fits better than the common log-power or linear-exponential. The authors use hypothetical choices for losses and so as to examine real large stakes. They also find some violations of separability of probability weighting versus utility of outcome. (**PT falsified; probability weighting depends on outcomes**).

decreasing ARA/increasing RRA: they find increasing relative risk aversion!

%}

Bouchouicha, Ranoua & Ferdinand M. Vieider (2017) "Accommodating Stake Effects under Prospect Theory," *Journal of Risk and Uncertainty* 55, 1–28.

16

{% **SPT iso OPT**: pp. 74-75 really uses the right formula for 1979 prospect theory (Eqs. 17 and 18)! This is exceptional. Almost all other authors do this wrong. The domain has only prospects with at most two nonzero outcomes, so, it is possible.

Paper tests gain- and loss prospects, but not mixed ones. For probability weighting, the paper allows for discontinuities at $p = 0$ and $p = 1$, capturing some insensitivity. In the interior, $0 < p < 1$, it only considers convex weighting functions, unfortunately (p. 75).

Risk averse for gains, risk seeking for losses (p. 85, table 5 and p. 89); more subjects are risk averse for gains than risk seeking for losses

real incentives/hypothetical choice: done here (p. 81). Half of of the participants were paid, half were not; no difference was found, neither in consistency, nor in risky choosing, nor in violations of independence. Discussed in §3.3 (p. 82 ff). P. 82 tests isolation of RIS by allowing subjects, after selection the choice to play for real, to change previously stated preference, with 80 subjects. Only 2 out of 80 subjects changed. They show that independence is massively violated, but isolation is not. This is a mild form of deception because experimental choices, announced to be consequential, in fact are not really so (**deception when implementing real incentivescrowding-out**).

losses from prior endowment mechanism: said on p. 81; done for 96 subjects; p. 84/85 suggests that only part of subjects, not all, do isolation/integration of payment, but gives no very clear evidence on how many by using unclear overall tests.

P. 89: **risk averse for gains, risk seeking for losses** is found

P. 85 has nice discussion of within/between subjects and representative agent.

PT falsified: p. 94 describes dependence of probability weighting on outcomes in prospect theory. (**probability weighting depends on outcomes**)

reflection at individual level for risk: unfortunately the paper does not report this (Section 4.2). It only confirms reflection at average level (Section 4.1).

inconsistency in repeated risky choice: this paper has 31.6% %}

Camerer, Colin F. (1989) "An Experimental Test of Several Generalized Utility Theories," *Journal of Risk and Uncertainty* 2, 61–104.

17

{% They show that salience theory can accommodate skewness preference. However, they do not take salience theory in its original form, but a continuous version that in fact is a special case of (generalized) regret theory. Fortunately, they state this explicitly, in §2 (I would have preferred in the intro though).

P. 2063 para below Def. 3 discusses a normalization. But it should be understood that the preference functional is invariant up to multiplication by any positive function g(C) where g can entirely depend on the choice situation C, so that this normalization has no empirical meaning.

In itself it is not surprising that salience theory can accommodate much because of its big generality, also its continuous version. In an experiment they find violations of transitivity. This is a violation of every transitive theory including prospect theory (**PT falsified**). It can be taken as support for salience theory

§2.2 defines certainty equivalents. In the absence of transitivity, these do not mean much.

§7.3 critically discusses regret theory. For one, the authors argue that regret must be anticipated, requiring info about the forgone outome. This info need not ocur in their experiment, for instance if subjects reeive a sure outcome.I see this differently. First, regret theory is only more convincing if info about foregone outcomes, and will still be working, but weaker, if not. But, seond, this holds the same for salience theory. Salience will be weaker if no info about foregone outcome. Further, this is only a difference of interpretation, not of preference functional. %}

Dertwinkel-Kalt, Markus & Mats Köster (2020) "Salience and Skewness Preferences," *Journal of the European Economic Association* 18, 2057–2107.

18
{% **PT falsified**: §III.B lists some.
Describes many empirical studies, oriented towards finance. Does not refer to Tversky & Kahneman (1992).
Risk averse for gains, risk seeking for losses: mentions several studies that find it. %}
Edwards, Kimberley D. (1996) "Prospect Theory: A Literature Review," *International Review of Financial Analysis* 5, 18–38.

19
{% Consider cases where the status-quo health state of people improves and consider health states that originally were above the status quo but are below now. They assume that utility is concave above the status quo and convex below (which, strictly speaking, is not defined for the nonquantitative outcomes considered here; but this problem can be fixed). This aspect of prospect theory, if taken in isolation, would imply that the health states considered have lower utility now than they had before. The authors test this hypothesis for 14 subjects. For 8 subjects they find higher utility now, contrary to the hypothesis, for 6 the same utility, and for 0 lower. They conclude that prospect theory is violated. (**PT falsified**)
It would be interesting to analyze the case considering loss aversion. Loss aversion is stronger than the concavity/convexity effect considered below. If I see things right, loss aversion will decrease the utility of outcomes that originally were closely above the status quo and now are considerably below, but will increase the utility of outcomes that originally were considerably above the status quo but now are closely below. In a complete analysis of prospect theory, also probability weighting would be incorporated. Thus, for a complete analysis of prospect theory it is not clear if the data of this paper confirm or reject it.
There are also intertemporal dependencies different than prospect theory that are effective here. %}

Feeny, David & Ken Eng (2006) "A Test of Prospect Theory," *International Journal of Technology Assessment in Health Care* 21, 511–516.

20

{% **decreasing ARA/increasing RRA**:

inverse-S: confirm it both for gains and for losses, using Goldstein & Einhorn (1987) two-parameter family

Risk averse for gains, risk seeking for losses: find it well confirmed.

reflection at individual level for risk: they have it in their data but do not report it.

Experiment in Bejing 2005 with real incentives for Chinese students (N = 153), and CEs (certainty equivalents) of 56 lotteries, using a finite mixture regression model. Stakes were like 1-hour wage (low-stake) versus 40-hour wages (high-stake). Always choice between sure outcome and 2-outcome prospect in choice lists to get CEs. Use the Goldstein & Einhorn (1987) two-parameter family for probability weighting, and power-utility.

Unfortunately, they implemented two choices for real for each subject, being one for high-stake and one for low-stake (the high-low stake comparison is within-subject), giving an income effect. It will, unfortunately, amplify a contrast effect with subjects simply taking low-stakes not very seriously. Not much can be done about this (other than do between-subject).

P. 154 footnote 5 properly points out that loss aversion does not affect choices between losses under PT; this paper only considers nonmixed prospects.

Point out that measurements of utility and risk aversion, and investigations of whether risk aversion is decreasing or increasing and whether concavity of utility is decreasing or increasing, cannot be settled properly if there is no correction for probability weighting and other things. Find increase in relative risk aversion for gains, but find that this is primarily driven by different probability weighting for high outcomes than for low. The latter entails a violation of prospect theory (**PT falsified; probability weighting depends on outcomes**). No increase or decrease but constant attitude is found for losses.

Losses with real incentives are implemented in an unconventional way: for each gain-choice there was a corresponding loss-choice that consisted of first a (choice-situation-dependent!) prior endowment and then the losses-choice, such that after integration of the endowment with the loss-choice the loss-choice was the same as the gain-choice. So, differences between gains and losses are a matter of framing, and this is how the authors often refer to it. Discussion of it on p. 170.

P. 151 top references several studies showing that heterogenous models can be really off. They find 1/4 subjects doing EV, and 3/4 PT. %}

Fehr-Duda, Helga, Adrian Bruhin, Thomas Epper, & Renate Schubert (2010) "Rationality on the Rise: Why Relative Risk Aversion Increases with Stake Size," *Journal of Risk and Uncertainty* 40, 147–180.

21

{% **survey on nonEU**:

Focuses on decision under risk with a bit on ambiguity.

Not primarily a complete survey but rather a didactical account giving the main ideas, with some nicely written sentences. For example, p. 51, on Rabin's paradox: "by way of analogy, if one could perceive the curvature of the earth by walking the length of a football field, then the earth must be implausibly small."

loss aversion: erroneously thinking it is reflection: this paper of course does NOT make this mistake. It usefully lists it as the first of some misunderstandings (top p. 55): "A few points of common confusion are worth highlighting at this juncture. First, loss aversion is not the same as risk seeking for losses. …Second, decision weights are not generally interpreted as a measure of belief. … Third, the concavity (convexity) of the value function is not the same as risk aversion (risk seeking), and overweighting low-probability gains (losses) is not the same as risk seeking (risk aversion)."

P. 58 brings up the two-stage model of PT for ambiguity, in the spirit of Tversky that I know well, having discussed it so much with him: there is belief and risk-probability weighting in the first para, with no space for the typical Ellsberg source preference. The latter is considered a relatively unimportant phenomenon much driven by contrast effects beyond individual choice, and reluctantly showing up in the 2nd para. Tversky convinced me of this and it underlied my work on ambiguity ever after. Tversky mostly discussed these things with Craig and me.

PT falsified: pp. 59-63 lists violations. The 2nd part of this paper is on external validity from lab to field, giving procedures to work on this.

P. 79 (conclusion) (**PT/RDU most popular for risk**):

"Despite its limitations, we find that prospect theory is the most successful general purpose model currently available for predicting, describing, and interpreting decisions under risk; to our reading alternative models that we reviewed outperform prospect theory only under specific conditions."

%}

Fox, Craig R., Carsten Erner, & Daniel J. Walters (2015) "Decision under Risk: From the Field to the Laboratory and back." *In* Gideon Keren & George Wu (eds.), *The Wiley Blackwell Handbook of Judgment and Decision Making*, 43–88, Blackwell, Oxford, UK.

22

{% Quantal Response Equilibrium (QRE) is explained in my annotations to McKelvey & Palfrey (1995).

It is a highly desirable step forward in game theory that not just expected value, but more general risk attitude models, are used for evaluations of strategies given others' choice probabilities. For the future of prospect theory etc., it is necessary to find applications in other domains such as here in game theory.

The precise working of the models, and the precise estimations of individual risk evaluations from the findings from game theory, are still complex. The only observable from behavior is the choice probabilities. To what extent these can be ascribed to individual evaluation, expected utility, prospect theory, or whatever the considered theory is, or some transformation of such an evaluation, and to what extent they can be ascribed to the noise parameters and other aspects of the strategic situation, depends on the models and parametric families chosen by the experimenters. That the choice probabilities depend on probabilities/utilities only through the EU or prospect theory of a prospect, so that this functional form is separable, is already a heavy assumption. As another example, in the middle of p. 255, the authors write that overbidding by some players will enhance overbidding by the others, in other words, overbidding is a self-reinforcing effect. In the analysis of this paper, however, stronger overbidding leads to higher estimates of individual risk aversion. Thus, estimates of individual risk attitudes are affected by strategic aspects of the game. One observable (choice probability) is used to estimate two or more parameters.

Another difference between these games and usual individual decision theories is that these theories consider decisions that are repeated many times, with repeated payoffs, income effects, etc. We must assume that in each repeated game, a strong isolation effect takes place, where the players forget about all other games. In spite of these difficulties, this is a highly intriguing attempt to apply individual risk theories in other domains.

When they do expected utility with power utility as index of risk aversion, they estimate the coefficient of RRA as 0.52 (so, power 0.48), which is similar to other findings in the literature. (**PT falsified**) When they do rank-dependent utility with linear utility, and Prelec's two-parameter family, they find convex and not **inverse-S** weighting functions. This puts the ball in the court of the inverse-S advocates. To maintain their hypothesis, they have to find other explanations for the strategic behavior of subjects than put forward in this paper. %}

Goeree, Jacob K., Charles A. Holt, & Thomas R. Palfrey (2002) "Quantal Response Equilibrium and Overbidding in Private-Value Auctions," *Journal of Economic Theory* 104, 247–272.

23

{% **PT falsified**: find S-shaped rather than **inverse-S** shaped probability weighting. P. 105 2nd para reports evidence against the procedure of paying in probabilities.

For the risk aversion assessment in the games as in §4, there is only one nonzero outcome, and then the problem is that a common power of utility and probability weighting is unidentifiable without further assumptions. The lottery-choice data in §5 have more variation in outcomes and there the problem does not arise.

The paper assumes that Nash equilibrium is what should/will happen under EU and no probabilistic choice. Many people, including me, do not find this a plausible assumption. %}

Goeree, Jacob K., Charles A. Holt, & Thomas R. Palfrey (2003) "Risk Averse Behavior in Generalized Matching Pennies Games," *Games and Economic Behavior* 45, 97–113.

24

{% **PT falsified**; find deviating kinds of reflection effects and different parameters when fitting. Main point of this work: propensity to show risk aversion/seeking depend on actual lottery pairs and person's proclivity.

Experiment 1 considered hypothetical choice, Experiment 2 real prizes (possibly given to charity). Stimuli were formulated as investments in the stock market (with selling short also).

Risk averse for gains, risk seeking for losses: is found. Further, there is more risk aversion for gains than risk seeking for losses:

- See Fig. 1: above 0.5 on y-axis risk seeking is found. Highest 80% risk seeking for losses, lowest 5% risk seeking (so, 95% risk aversion) for gains. For most gamble pairs in Appendix C (all with d ≠ 0.5) risk aversion is more pronounced than risk seeking.

- Table 2 on p. 948: more risk aversion for gains than risk seeking for losses, because always the loss- and gain percentage sum to less than 100%, so that for gains we are closer to zero (total risk aversion) than for losses we are to 100% (total risk seeking). Average 57% risk seeking for losses, 100–35 = 65% risk aversion for gains.

reflection at individual level for risk: no clear pattern, depending much on particular prospects

- Personal communication (email of Claudia of April 7 '04): in total, 87% of participants have risk aversion for gains, 63% have risk seeking for losses. %}

González-Vallejo, Claudia C., Aaron A. Reid, & Joel Schiltz (2003) "Context Effects: The Proportional Difference Model and the Reflection of Preference," *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29, 942–953.

25

{% **PT falsified**: this paper falsifies any other classical economic theory as well, with its extensive risk seeking, especially for gains.

Choices between one nonzero outcome prospects, and the sure outcome that was always the expectation of the prospect. Did it for children, young adults, and adults, ages 5-8, 9-13, 14-20, and 21-64. Did it for probabilities 0.02, 0.10, 0.80, and 0.98. Find in everything the almost exact opposite of the fourfold pattern predicted by prospect theory: people seem to underweight **small probabilities** and overweight high probabilities, both for gains and for losses, yielding the exact opposite of the fourfold pattern. As people are older they are closer to expected value maximization. People are closer to expected value maximization for gains than for losses. People are more risk averse for gains than for losses.

Real incentives: **random incentive system** where one choice is played for real. Implementation of losses: through **prior endowment mechanism** to ensure no real loss.

P. 59: people who violated monotonicity tended to be more risk averse.

P. 60 bottom: strange is that the majority choices, 56%, were risk seeking, and were so mostly for gains. Maybe the design generated a strong joy of gambling? This is evidence against prospect theory, but against any other current theory as well.

linear utility for small stakes: they make this assumption for pragmatic reasons.

The authors conjecture (p. 72 penultimate paragraph) that their deviating findings may be due to their stimuli of risky versus riskless choices, claiming that this is different to almost all prior work. This is not so, Tversky & Kahneman (1992) and many others also considered such choices (not doing WTP but choice). %}

Harbaugh, William T., Kate Krause, & Lise Vesterlund (2002) "Risk Attitudes of Children and Adults: Choices over Small and Large Probability Gains and Losses," *Experimental Economics* 5, 53–84.

26

{% **equate risk aversion with concave utility under nonEU**: p. 597: unfortunately, they use the term risk neutral for linear utility, also under PT, even though with linear utility there then can still be large deviations from risk neutrality due to probability weighting. They mention that only few studies have tested the fourfold pattern using choices. The following search key words in this bibliography can give related references:

concave utility for gains, convex utility for losses;

Risk averse for gains, risk seeking for losses

PT falsified

risk seeking for small-probability gains

P. 598 last para explains why their 2002 study is so unique.

losses from prior endowment mechanism: subjects received $22 in beginning, well, it was put on a table in front of them and apparently not yet put in their pocket. They might have to pay back from that.

random incentive system: each subject was paid twice, so, there is income effect. When they played their first choice they did not yet know a second would come (p. 601 l. 6), so, this can be taken as without the income effect (but then with a minor deception) (**deception when implementing real incentives**). Second time they were, again, endowed with $22.

Although pricing tasks confirm 4-fold pattern, I find it hard to interpret the stimuli and results. Subjects had to pay their WTP to get a gain prospect, so that losses could be involved and it was not really a gain prospect. The authors point this out in footnote 8 (p. 599) and discuss it more in §5, but nevertheless analyze what they call gain prospects as if gain prospects. Further complication is that, with prior endowment put on table before them, it is not clear to me if subjects integrated or not, took it as house money or not, and so on.

P. 602 writes that loss aversion can explain that for losses the WTP in absolute value was usually found to be larger than for gains. If subjects took the prospects as the authors analyze and describe them (gain-prospects and loss-prospects) then there would be no mixed prospects and loss aversion had no role to play. (**loss aversion: erroneously thinking it is reflection**)

Pp. 602-603 finds relations at individual level between gain- and loss-attitudes, different than Cohen, Jaffray, & Said (1987) who found no relation.

In the choice task where subjects chose between prospects and their expected values, but were endowed with $22, not given but put on the table before them. They found mostly nonsignificant deviation from EV, and the deviations all suggested to go opposite to the 4-fold pattern. I find it hard to assess the effect of the prior endowment mechanism though. Much of this evidence does not only go against PT, but against any theory we know.

In some places the authors put forward the dual self theories when discussing their results. %}

Harbaugh, William T., Kate Krause, & Lise Vesterlund (2010) "The Fourfold Pattern of Risk Attitudes in Choice and Pricing Tasks," *Economic Journal* 120, 595–611.

27

{% **PT falsified;** They ask subjects introspective question about values of positive and small negative amounts. For small amounts they find stronger evaluations of positive amounts, deviating from loss aversion. For large amounts they find loss aversion. Experiment 1: how nice/unnice is it to gain/lose money. Experiment 2 repeats it for money gained/lost against a bookmaker. A control question could have been how happy subjects feel if they neither gain nor lose, so as to determine what the value of the reference point is and if it is really the neutrality point of the scale the authors use.

Another aside is that loss aversion may be due to the overweighting of the loss experience/anticipation and not to the experience itself.

risk seeking for symmetric fifty-fifty gambles: experiment 3 asks for −x such that (−x, p; y) ~ (−a, p; b) (not incentivized).

Problem with small amounts is that distorting factors such as joy of playing and framing decide. %}

Harinck, Fieke, Eric van Dijk, Ilja van Beest, & Paul Mersmann (2007) "When Gains Loom Larger than Losses," *Psychological Science* 18, 1099–1105.

28

{% **equate risk aversion with concave utility under nonEU**: they explicitly state, somewhere in the middle, that risk aversion, risk seeking, and so on, refers only to utility curvature, also under prospect theory. Confusing, because then we do not know how to refer to what is traditionally called risk aversion (preference of EV, involving both utility, probability weighting, and loss aversion)! Unfortunately, the paper, whereas mentioning original 1979 prospect theory, the separable-weighting generalization often used (though not really prospect theory), and the new 1992 version, but leaves it completely unspecified which of these versions is used in the analysis, for instance, by not giving the formula.

PT falsified: they confirm the violations of inverse-S found by Humphrey, & Arjen Verschoor (2004).

They measure probability weighting but use the RIS, something strongly criticized by Harrison & Swarthout (2014). %}

Harrison, Glenn W., Steven J. Humphrey, & Arjen Verschoor (2010) "Choice under Uncertainty: Evidence from Ethiopia, India and Uganda," *Economic Journal* 120, 80–104.

29

{% **PT falsified** & **reflection at individual level for risk**: they present data that violate reflection by measuring risk attitudes for both gains and losses, both between and within subjects. There are no clear patterns and findings, and there are relations in all directions. Unfortunately, they do not report correlations, but only patterns of risk seeking/risk aversion, which is similar to median splits. Tversky & Kahneman (1992, p. 308) will criticize this research for underestimating the unreliability of individual choices.

Table 3 and p. 409: more risk aversion for gains than risk seeking for losses.

Risk averse for gains, risk seeking for losses: Table 3 is nice way to inspect data. Fourfold pattern is confirmed with one exception: for gains with probabilities below .01, down to .001, they do not find risk seeking. For probabilities .1 and .2 they do. For losses they do find the fourfold pattern of risk aversion for small probabilities but risk seeking for moderate and high probabilities.

insurance frame increases risk aversion: seems they have that. %}

Hershey, John C. & Paul J.H. Schoemaker (1980) "Prospect Theory's Reflection Hypothesis: A Critical Examination," *Organizational Behavior and Human Performance* 25, 395–418.

30

{% **PT falsified** & **inverse-S**: they test the common consequence effect and find risk aversion increasing and not decreasing, which is the exact opposite of inverse-S. This independently replicates the same finding as by Birnbaum, for instance in Birnbaum & Chavez (1997).

Use random incentive system. Did it with poor farmers from the countries mentioned in the title.

More elaborate results, with error theories added, are in Humphrey & Verschoor (2004, *Journal of African Economies*). Unfortunately, the papers have no cross references to explain their overlap and priority. %}

Humphrey, Steven J. & Arjan Verschoor (2004) "The Probability Weighting Function: Experimental Evidence from Uganda, India and Ethiopia," *Economics Letters* 84, 419–425.

31
{% **PT falsified** & **inverse-S**: do same as their 2004 Economics Letters paper, but more elaborate, with error theory added. Then still they prefer RDU with error better than EU with error. (e.g. p. 82 & 84) %}

Humphrey, Steven J. & Arjan Verschoor (2004) "Decision-Making under Risk among Small Farmers in East Uganda," *Journal of African Economies* 13, 44–101.

32
{% Experiment plus desire to link individual and group behavior.

PT falsified: **risk seeking for symmetric fifty-fifty gambles**: they seem to find it. %}

Kameda, Tatsuya & James H. Davis (1990) "The Function of the Reference Point in Individual and Group Risk Decision Making," *Organizational Behavior and Human Decision Processes* 46, 55–76.

33
{% https://doi.org/10.1007/s00355-018-1111-y

The authors measure CEs (certainty equivalents) using choice lists, for money and number of human lives, with losses also involved in mixed prospects, and fit prospect theory. Of course, must be hypothetical. For gains, probability weighting is the same for money and lives.

PT falsified; probability weighting depends on outcomes: for losses, probability weighting is less elevated for lives, suggesting more risk seeking there. They find bigger loss aversion for human lives, suggesting more risk aversion for mixed prospects. %}

Kemel, Emmanuel & Corina Paraschiv (2018) "Deciding about Human Lives: An Experimental Measure of Risk Attitudes under Prospect Theory," *Social Choice and Welfare* 51:163–192.

34
{% **PT falsified**: a theory where people choose several reference points, and primarily go by the probability of exceeding those, fits data well. It is like Diecidue & van de Ven (2008) and Payne (2005) although they do not cite those. It is also like Lopes model, which is cited. However, the reference points are simply introduced

here physically as thresholds above which the subjects gain points to participate in a bonus. Thus they are just outcomes rather than psychological thresholds and in this sense the paper does not really show that thresholds lead to deviations from just maximizing outcomes. %}

Koop, Gregory K. & Joseph G. Johnson (2012) "The Use of Multiple Reference Points in Risky Decision Making," *Journal of Behavioral Decision Making* 25: 49–62 (2012).

35

{% **PT falsified; probability weighting depends on outcomes** %}

Krawczyk, Michal W. (2015) "Probability Weighting in Different Domains: The Role of Affect, Fungibility, and Stakes," *Journal of Economic Psychology* 51, 1–15.

36

{% P. 1487 calls prospect theory the leading psychological descriptive theory of "decision making" without there specifying risk. (**PT/RDU most popular for risk**)

This paper uses the term precautionary decision as equivalent to insurance decision, deviating from economic terminology where it means reducing but not entirely removing bad probabilities.

In rank-dependent theories, including PT, one can use two dual ways of using the probability weighting function in the preference functional (top-down or bottom-up), and this paper left me confused on what they do. What is high one way, is low the other way, and what reflects optimism one way, reflects pessimism the other way. (Inverse-S is not really affected by it.) In the early days of RDU, bottom-up was most common, but nowadays top-down is the almost universally agreed upon convention. For PT of Tversky & Kahneman (1992), top-down for gains and bottom-up for losses is the common way. This paper uses PT but, unfortunately, does not specify which way of integration it uses. P. 1491 penultimate para of 1st column claims that the δ parameter of probability weighting represents attractiveness of a lottery, without specifying if this is for gains or losses. The weighting function is given in Eq. 1, p. 1491, and it is the Goldstein-Einhorn (1987) family (they cite Gonzalez & Wu 1999). The authors interpret the parameter δ, the index of elevation, as attractiveness (2nd para below Eq. 1). However, under common PT, for losses, it is the opposite, unattractiveness. And insurance is about losses (although this paper considers both gains and losses). This left me confused. On p. 1501, 2nd column, end of 3rd para, the authors write that overweighting of probabilities means risk aversion for losses, suggesting that they did use the common way of integration. My comments below will, therefore, assume the common way of integration.

It has often been observed that framing a risky choice as an insurance decision increases risk aversion. (**insurance frame increases risk aversion**) The authors mention this on p. 1488. For probability weighting for losses under PT this will increase pessimism; i.e., it will increase the weighting function and the intersection point with the diagonal. This paper confirms this finding in a number of experiments.(**PT falsified; probability weighting depends on outcomes**:) The authors take this as evidence against PT. But I find it so much one of the many known framing effects affecting every theory, that I would not interpret it that way.

A general phenomenon with experiments is that subjects often replace the info given by the experimenter with their own experiences. If the experimenter says "assume that this has probability 1" they may reason: "the experimenter may say so, but I think it is different" and they go by their own ideas. This may explain why subjects in this experiment were not only affected by the probabilities given by the experimenters, but also by accessibility. P. 1497 1st column penultimate para writes:

It was not clear to me what accessibility means in this paper, and to what extent it is anything beyond probability/frequency, although it apparently is assumed to happen only with insurance events and not with just probability-gambles. P. 1495 2nd column will give high-frequency events as an example of accessible events.

P. 1495 text from 1st to 2nd column: "When evaluating risks for insurance, people do not usually use statistical evidence about the probability of risky events. Instead, people may commonly rely on inferences based on what they remember hearing or observing about a particular risk (Hertwig, Pachur, & Kurzenha¨user, 2005; Slovic, Fischhoff, & Lichtenstein, 1979; Tversky & Kahneman, 1973)."

P. 1497: "in the low-frequency insurance risk condition, we attached the highest probability levels to those risks judged as less frequent in the norming procedure" Such hypothetical things may be hard to imagine for subjects, and they may rather substitute their own ideas. For such kinds of questions, real incentives are useful. This paper did everything hypothetical, asking subjects over 100s of hypothetical choices.

They did data-fitting on many choices from which CEs (certainty equivalents) were derived using power utility and the Goldstein & Einhorn family (for which they refer to Gonzalez & Wu 1999). They usually confirm **inverse-S**. %}

Kusev, Petko, Paul van Schaik, Peter Ayton, John Dent, & Nick Chater (2009) "Exaggerated Risk: Prospect Theory and Probability Weighting In Risky Choice," *Journal of Experimental Psychology: Learning, Memory, and Cognition* 35, 1487–1505.

37

{% Asks participants to choose x to optimize (p,x; q,20–x; r,0) and also in (p',x; q',20–x; r',0) with p'/p = q'/q. EU predicts same x. This is not found. EV predicts x = 20 or x = 0, but there were remarkably many deviations.

PT falsified: regarding **inverse-S**: for RDU, his evidence cannot be reconciled with an inverse-S weighting function (p. 104) but it can neither be with a convex (pp. 1-3).

Uses $2p^3 - 3p^2 + 2p$ as inverse-S weighting function. %}

Loomes, Graham (1991) "Evidence of a New Violation of the Independence Axiom," *Journal of Risk and Uncertainty* 4, 92–109.

38

{% **PT falsified**: measure certainty equivalents of prospects, allowing for choice errors. Find violations of PT, and suggest that a similarity theory may fit better. The authors are negative on PT (which they call CPT): "If CPT is to justify its current status as the front runner among alternatives to EUT, it should be able to organise the data from our CREPROBS treatment; but it cannot do so," (p. 209). The main purpose of the paper is to argue for the use of error theories. %}

Loomes, Graham & Ganna Pogrebna (2014) "Testing for Independence while Allowing for Probabilistic Choice," *Journal of Risk and Uncertainty* 49, 189–211.

39

{% Multioutcome lotteries; conclude that PT does not do well (**PT falsified**); seems that "cautiously hopeful" is her term for **inverse-S** %}

Lopes, Lola L. (1990) "Re-Modeling Risk Aversion: A Comparison of Bernoullian and Rank Dependent Value Approaches." *In* George M. von Furstenberg (ed.) *Acting under Uncertainty: Multidisciplinary Conceptions*, 267–299, Kluwer, Dordrecht.

40

{% There is a clear definition of SP/A theory, clearer than Lopes' papers, in Ch. 26 of Shefrin, Hersh M. (2008) "*A Behavioral Approach to Asset Pricing Theory;* 2nd *edn.*"

In SP/A theory, a prospect (lottery over money) depends on

(1): SP. This is a rank-dependent utility, with linear utility, and a weighting function that is a convex combination of a power function p^r and a dual power function $1 - (1-p)^{r'}$, where the first captures pessimism and the second optimism. For the claims about mixed weighting functions in Eqs. 9 and 10 (p. 290), it is important to know that the parameters q_r and q_p are supposed to be positive (I assume), so that the w-weighted curve is convex and the (1−w) weighted curve is concave, and the convex mix gives an **inverse-S** shape.

(2) A: an aspiration level, i.e. an outcome, is chosen, and A is the probability of (weakly!?) exceeding it.

How these two are combined, is explicitly left unspecified. P. 291 end of penultimate para writes that, if these two components prefer a different prospect (so, if the case is not totally trivial), then SP/A predicts "conflict." This gives a revealed-preference oriented economist little hope of being informed about what choice then results. The text then writes that such conflict cannot result from "single-criterion" models such as CPT (p.s.: CPT and all economic models can consider multi-criteria optimization in utility), which further reduces my hope of being informed about the resulting choice in any not-completely-trivial situation. P. 300 2nd para will mention an aggregation of the two components but it is not clear how, apparently through a nmerical Table 5.

The first para on p. 292 confuses monotonicity with absolute risk aversion, and erroneously claims that CPT would have constant absolute risk aversion.

Although in several places the paper writes that it, unlike prospect theory, has no reference point but *instead* an aspiration level, SP/A theory turns out to have a reference point still because it does distinguish between gains and losses, where every parameter in the model (including probability weighting, contrary to what Shefrin, 2008, p. 429 last sentence, claims) can depend on the sign (pp. 290-291 & 299). In particular, the aspiration level can be different for gains than for losses (then how about mixed prospects?), and will later (p. 300 top) be taken to be 0 for losses and, ad hoc, 1 for gains.

P. 302, Eq. 16 suddenly does aggregate SP and A into a decision formula, although it is a probabilistic choice model, with no deterministic model specified. For me, the formula comes out of the blue, seeming to assign the same weight to SP as to A. (I'd expect SP to have more weight.) Does this satisfy stochastic dominance? Some form of transitivity?

P. 310 penultimate para has a nice text on risk aversion being conflated with utility.

Shefrin (2008 p. 431 bottom) writes that the weighting function in prospect theory captures perception, but in SP/A it captures emotions.

In Table 5 it is amazing that the very crude A-criterion alone (just the probability of exceeding aspiration, which is nothing but probabilities related to 0) explains data so well. Then SP/A will do better than PT! Makes me wonder about the stimuli.

PT falsified: not strongly. Mostly, Lopes' SP/A theory fits data better than her implementation of PT (which is questionable given that she, erroneously, thinks that PT satisfies constant absolute risk aversion).

1. **convex utility for** losse**s**: for losses subjects are risk-neutral more than risk-seeking

2. Subjects seem to prefer (0.5: 50, 0.5: 150) to 100 for sure. Seems to agree with Lopes SP/A theory, while violating PT. (Is like {**risk seeking for symmetric fifty-fifty gambles**}, but not symmetric about 0.)

Risk averse for gains, risk seeking for losses: seem to be risk neutral for losses; multioutcome lotteries.

loss aversion without mixed prospects: they claim to estimate loss aversion λ, but they do not consider mixed prospects and, therefore, it is impossible to estimate λ.

linear utility for small stakes: p. 290 footnote 1 %}

Lopes, Lola L. & Gregg C. Oden (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.

41

{% The paper provides two examples of plausible preferences that violate RDU (CEU (Choquet expected utility) as Machina call it) for uncertainty. Baillon, L'Haridon, & Placido (2009) show that the examples also violate most other nonEU models for uncertainty popular today in the Anscombe-Aumann framework; without that, Machina's counterexample only concerns RDU. In particular, the examples violate the comonotonic sure-thing principle and even tail-independence. I find the second example, the reflection example (pp. 389-390), impressive, nay, brilliant. But other than that I prefer different interpretations and explanations than the author gives for almost everything.

The reflection example (with my interpretations): an urn contains 100 balls. 50 balls marked 1 or 2 in unknown proportion, and 50 marked 3 or 4 in unknown proportion. One ball is drawn randomly. E_j: the number drawn is j. Consider (with \$1000 as unit) preferences between f_5 and f_6, and then between f_7 and f_8:

		#50	#50
f_5	=	(E_1:4, E_2:8,	E_3:4, E_4:0),
f_6	=	(E_1:4, E_2:4,	E_3:8, E_4:0),
f_7	=	(E_1:0, E_2:8,	E_3:4, E_4:4),
f_8	=	(E_1:0, E_2:4,	E_3:8, E_4:4),

Ambiguity averse people will have $f_6 > f_5$ because f_6 has one outcome, 4, resulting with known probability ½, whereas f_5 has all outcomes ambiguous. For exactly the same reason, ambiguity averse people will have $f_7 > f_8$. These claims were confirmed empirically by L'Haridon & Placido (2010).

Btw., because of informational symmetry, f_7 is like f_6 and f_8 is like f_5, so that the second preference follows from the first from informational symmetry.

RDU however predicts indifference between the four acts because RDU considers likelihoods of what are known as goodnews ("decumulative;" "ranks") events. For all four acts, the goodnews event of receiving 8 contains one E_j, the goodnews event of receiving 4 or 8 contains three E_js, and the goodnews event of receiving 0, 4, or 8 contains all four E_js. Beause of informational symmetry, each goodnews event has the same weight under each act, implying immediately that the four acts are indifferent by RDU, simply having identical Choquet integals. (Btw: Machina uses a different reasoning, being that the comonotonic sure-thing principle, and even tail independence, require that a strict preference between f_5 and f_6 be the same as between f_7 and f_8, rather than between f_8 and f_7 as informational symmetry has it. Because informational symmetry is unquestionable, RDU hence cannot have strict preference and must have indifferences.)

(Another btw.:

Sarin & Wakker 1992 axiomatized RDU using an axiom that acts are equivalent whenever all goodnews events have the same likelihood, in an axiom called cumulative dominance.)

I like Machina's reflection example much because it addresses a fundamental issue of RDU (with similar things for most other nonEU theories as demonstrated by Baillon, L'Haridon, & Placido), being that RDU focuses on likelihoods of goodnews events, but Machina's example shows that subjects are also partially driven by likelihoods of separate-outcome events, as considered in old pre-rank-dependent nonadditive probability models. (**PT falsified**)

I regret that Machina does not refer to the role of separate-outcome events and the unambiguity of one outcome in his reasoning against indifference. Instead he uses a complex riding-on reasoning (f_5 has two small ambiguities and f_6 one big; if one had something like aversion to mean-preserving spreads one would prefer f_5; as Baillon, l'Haridon, & Placido rightfully point out, ambiguity is more cognitive than motivational, is more subject to diminishing sensitivity, and it is more categorical ambiguous versus unambiguous than more versus less, so that the two ambiguities of f_5 will count more negatively than the one ambiguity or f_6) that can only be understood by specialists, and then after some effort, but that will enter the mind of no natural subject that I can think of. He thus does not choose side for one strict preference or the other even though it is clear enough I think, and he further refers to an unclear tradeoff between objective and subjective uncertainty.

Machina's 50-51 example, while equally valid as the reflection example, is way less clear. Now unambiguity must be traded against an objective-likelihood argument in a first choice problem (between f_1 and f_2) and also in a second choice problem (between f_3 and f_4). In the second choice problem the ambiguity degree of all goodnews events is the same as in the first and it can be proved under RDU that the preference in the second choice problem should be the same as in the first. In the second choice problem the ambiguity degree of all separate-outcome events is not the same as in the first, and therefore choices can be different. Because of the tradeoff with objective probability this example is less clear, and will work less well empirically than the reflection example. Machina's explanation on pp. 388-389 again (as in the reflection example) does not raise the argument of a separate-outcome event, unfortunately. Instead if raises an unclear correlation argument. One problem is that correlation is not defined as he discusses it. (You need numbers to correlate, so, how should this be with events? Indicator functions will not help. He could formalize the first point in terms of stochastic-like or sigma-algebra-like independence. Btw., p. 388 last line "corrected" should be "correlated" and this is a typo.) He proceeds claiming that in the second choice problem some correlations are less, and this is not clear either.

He also overstates implications. P. 389 4th para suggests that models like RDU, which maintain comonotonic separability, kep the Ellsberg problem. He tries to suggest there that his example is as strong and fundamental as Ellsberg's. This is not so; it is different, and less strong, albeit surely interesting.

P. 390 writes: "If there is a general lesson to be learned from Ellsberg's examples and the examples here, it is that the phenomenon of ambiguity aversion is intrinsically one of nonseparable preferences across mutually exclusive events, and that models that exhibit full—or even partial—event-separability cannot capture all aspects of this phenomenon.."

This text suggests that all models of nonEU for ambiguity should consider interactions and violations of separabililty of events. I in fact agree but I disagree that Machina's examples, which are only two examples, (nor the Ellsberg examples which Machina puts on the same footing there), could prove this in general, as Machina is suggesting. Even worse, Machina claims that *every* partial form of event-separability will fail. This claim is completely unfounded. Machina has done no more than show a problem for comonotonic separability (sure-thing principle) and even for tail-separability (independence). Theories that completely give up any event-separability will probably be too general to be tractable. For the same reason, the general Machina (1982) nonexpected utility, while useful to bring some theoretical points, is too general for most purposes.

Something else I found amazing is that on several occasions (p. 390 2nd para "the issue is not how individuals *ought* to choose …" and the closing sentence on p. 391) Machina treats ambiguity purely descriptively, and nothing normatively. I as Bayesian like to have ambiguity only descriptively, but still would not explicitly exclude any normatively-based discussion of it. %}

Machina, Mark J. (2009) "Risk, Ambiguity, and the Rank-Dependence Axioms," *American Economic Review* 99, 385–392.

42

{% N = 74. Hypothetical (footnote 11, p. 447: because BDM (Becker-DeGroot-Marschak) needs (according to the authors) EU. Btw, although EU, implemented the natural way in dynamic choice, is sufficient for BDM, it is not necessary! A common confusion.

PT falsified: when they tried to refine EU by CEU (Choquet expected utility), they actually got worse results. So, CEU picks up more noise than essential things (overfitting). To elicit CEU, they first assume EU for given probabilities so as to get utility and then elicit capacities from that. Or they equate the capacity of an event with the probability of a matched known-probability event, which also requires EU for risk. Martin Weber (personal communication) conjectured that the poor performance of CEU may be due to participants first getting many known-probability questions preceding the ambiguity questions which may have distorted their ambiguity perception.

ambiguity seeking for losses: they find ambiguity aversion for gains but, on average, ambiguity neutrality for losses. P. 448 2nd para: significant difference between gains and losses. Capacities for losses are significantly different than for gains.

reflection at individual level for ambiguity: although they have the data, within-subject, they do not report it.

P. 442 ll. 4-5: they apparently assume EU for risk. %}

Mangelsdorff, Lukas & Martin Weber (1994) "Testing Choquet Expected Utility," *Journal of Economic Behavior and Organization* 25, 437–457.

43

{% **PT falsified**: a useful paper putting PT to new tests and demonstrating that we need better parametric families.

The defenses of PT demonstrating that it accommodates the Allais paradox, gambling, insurance, etc., have usually focused on only one of these phenomena. Parametric fittings of PT have not been checked yet for what they say about these known phenomena. This paper is the first, to my knowledge, to see if the parameters found for PT can do more and explain known patterns of choices jointly, and if the parameters found give plausible behavior outside the immediate paradoxes. The current parametric families don't perform well. For example, the T&K families, if explaining the Allais paradox, must be very risk averse, too much to give much gambling for low probabilities. Similar observations apply to coexistence of gambling and insurance. Risk premia are calculated and often are not very plausible. %}

Neilson, William S. & C. Jill Stowe (2001) "A Further Examination of Cumulative Prospect Theory Parameterizations," *Journal of Risk and Uncertainty* 24, 31–46.

44

{% **PT falsified; probability weighting depends on outcomes**: they investigate this. Several studies have shown that affectrich outcomes can affect probability weighting, the electric shocks versus moviestar kisses of Rottenstreich & Hsee (2001) being most well known. This paper shows the effect very thoroughly, also within-subject, and is the first to do so. The main finding is that affect-rich outcomes make people less, or even completely, insensitive to probabilities. Process data with eye tracking support this claim. The authors interpret disregarding probabilities as something fundamentally different than bigger insensitivity (p. 75 last para of 1st column and p. 76 2nd column 2nd para), and follow that same interpretation in other papers. I disagree. It is an extreme case of insensitivity. Thus, what the authors take as evidence against inverse-S, in my opinion is strong support. %}

Pachur, Thorsten, Ralph Hertwig, & Roland Wolkewitz (2014) "The Affect Gap in Risky Choice: Affect-Rich Outcomes Attenuate Attention to Probability Information," *Decision* 1, 64–78.

45

{% **PT falsified** %}

Payne, John W. (2005) "It Is whether You Win or Lose: The Importance of the Overall Probabilities of Winning or Losing in Risky Choice," *Journal of Risk and Uncertainty* 30, 5–19.

46

{% Use hypothetical choice. Study relation between inverse-S and cognitive ability (**cognitive ability related to likelihood insensitivity (= inverse-S) & inverse-S (= likelihood insensitivity) related to emotions**).

With affect-rich outcomes (voucher for romantic dinner) there is more likelihood insensitivity than with affect-poor outcomes (reduction of electricity bill). (**PT falsified**: see also **probability weighting depends on outcomes;**) Numerosity (Berlin number task) also seems to reduce likelihood insensitivity (in re-appraisal task.). These results, however, seem to hold only for small probabilities, and not for large.

To calculate probability weighting, they assume linear utility, which for moderate stakes is fine. Data-fitting is by minimizing quadratic distance. They confirm inverse S. %}

Petrova, Dafina G., Joop van der Pligt, & Rocio Garcia-Retamero (2014) "Feeling the Numbers: On the Interplay between Risk, Affect, and Numeracy," *Journal of Behavioral Decision Making* 27, 191–199.

47

{% **PT falsified**: a detailed study finding many violations of gain-loss separability in PT, using both CE measurements and choice. They use randomly generated stimuli. %}

Por, Han-Hui & David V. Budescu (2013) "Revisiting the Gain–Loss Separability Assumption in Prospect Theory," *Journal of Behavioral Decision Making* 26, 385–396.

48

{% **PT falsified**: background risk can "destroy" most of rank dependence, because the background risk mostly determines the ranking position of outcomes that can be all over the place. I learned this from Quiggin (personal communication, end of 1990s). This paper resulted from the insight but, unfortunately, it its final version only has a weaker result, being that background risk can reduce the risk premium under constant relative and constant absolute risk aversion. A related result is in Barberis, Huang, & Thaler (2006). %}

Quiggin, John (2003) "Background Risk in Generalized Expected Utility Theory," *Economic Theory* 22, 607–611.

49

{% **Utility of gambling**: a low-affect outcome was preferred to a high-affect outcome if received with certainty, but not if received with low probability.

PT falsified; probability weighting depends on outcomes: probability weighting more curved for more affective outcomes (**inverse-S (= likelihood insensitivity) related to emotions**) %}

Rottenstreich, Yuval & Christopher K. Hsee (2001) "Money, Kisses, and Electric Shocks: On the Affective Psychology of Risk," *Psychological Science* 12, 185–190.

50

{% N = 60; essentially hypothetical; gain- and loss questions were separated by a week. P. 541 1st column explains some of data analysis but I do not understand. The authors claim that for examining risk aversion, a value function must be specified, and they take 2/3 power for gains and ¾ power for losses. This leaves me in the blue what their concept of risk aversion is. Some lines below it is written that they analyze risk aversion "if we ignore for the moment effects due to probability weighting" and again I have no clue what they are doing.

PT falsified: **risk averse for gains, risk seeking for losses**: seem to be risk neutral for losses; multioutcome lotteries; conclude that OPT does not do well. %}

Schneider, Sandra L. & Lola L. Lopes (1986) "Reflection in Preferences under Risk: Who and when May Suggest why," *Journal of Experimental Psychology: Human Perception and Performance* 12, 535–548.

51

{% **real incentives/hypothetical choice**: uses random incentive system;

PT falsified: when OPT (1979-prospect theory) predicted particular violations of transitivity and monotonicity (if no editing), the theory was widely criticized for it. This paper, however, tests such violations of transitivity (or monotonicity) and finds them confirmed. It, thus, gives empirical support to OPT.

Details:

Prospect A = $14_{0.20}0$; Prospect B = $8_{0.30}0$; Prospect C = (0.15:8, 0.15:7.75, 0.70:0). By monotonicity, B ≻ C, but by subadditivity of probability weighting under OPT (which does not amount to event-splitting here because lotteries are always collapsed) we can have C ≻ B. OPT predicts C ≻ A ≻ B (including C ≻ B) because the evaluating function implies these prefs. It, however, predicts B ≻ C because of monotonicity and editing, and thus intransitivity results.

Testing number of cycles C ≻ A ≻ B ≻ C versus number of reversed cycles C ≺ A ≺ B ≺ C would not be very satisfactory because simple error theories could predict fewer errors in B ≻ C because of salience of monotonicity, and thus predominance of former cycles, without genuine intransitivity underlying it. This paper, therefore, tests only frequency of A ≻ C versus A ≻ B, and finds the former dominating. This is enough, under any plausible error theory, to ensure that either monotonicity or transitivity must be violated. Data find few violations of monotonicity and, hence, transitivity must be violated. These data were found for many stimuli A,B,C similar to the above ones. %}

Starmer, Chris (1999) "Cycling with Rules of Thumb: An Experimental Test for a New Form of Non-Transitive Behavior," *Theory and Decision* 46, 141–158.

52

{% **real incentives/hypothetical choice**: **random incentive system**, explained on p. 93; this is same experiment as their 1989 JRU paper, so see there for further explanation.

PT falsified: they find a necessary condition of PT and RDU violated. The necessary condition, explained on pp. 86-90, was found by accident (explained on p. 95 bottom), but actually is really clever.

Define the cumulative prospect theory functional (so, rank- and sign-dependent utility) for decision under risk, in the appendix. Preceded Tversky & Kahneman (1992) and Luce & Fishburn (1991). Well, they don't take a general probability transformation for losses but the dual of the one for gains (as reflection would have it), but still it is clear that the rank- and sign-dependent idea is there. This paper was, in turn, preceded by Šipoš (Sipos) (1979) who also defines the symmetrical integral. %}

Starmer, Chris & Robert Sugden (1989) "Violations of the Independence Axiom in Common Ratio Problems: An Experimental Test of Some Competing Hypotheses," *Annals of Operations Research* 19, 79–102.

53

{% **PT falsified**: propose a theory that is a kind of mix of CBDT of Gilboa & Schmeidler, Arducci's range-frequency theory, and Erev's Decision-from-Experience-theory. Choice alternatives are evaluated by comparison to related alternatives stored in memory, and binary comparisons with those. It leads to alternative explanations for some of the main empirical findings, such as concave utility, **inverse-S** probability weighting, loss aversion, and hyperbolic discounting. %}

Stewart, Neil, Nick Chater, & Gordon D.A. Brown (2006) "Decision by Sampling," *Cognitive Psychology* 53, 1–26.

54

{% **SEU = SEU**: p. 688 2nd para lists Savage (1954) as one of the nonEU theories for risk.

PT falsified: this paper gives further evidence on the theories of Stewart et al, that decisions, utility, and so on are influenced by stimuli seen before. The authors use pessimistic words such as "there is no stable mapping between attribute values and their subjective equivalents." I have a different DESCRIPTIVE opinion coming from the NORMATIVE view (not central among psychologists) that such subjective equivalents should exist for rational decisions, and then the descriptive goal to find them as good as possible despite the big biases and noise that exist. %}

Stewart, Neil, Stian Reimers, & Adam J.L. Harris (2015) "On the Origin of Utility, Weighting, and Discounting Functions: How They Get Their Shapes and how to Change Their Shapes," *Management Science* 61, 687–705.

55

{% **PT falsified; probability weighting depends on outcomes**: they investigate this. They confirm that affect-rich outcomes give more pronounced insensitivity (inverse-S). On one point my interpretation is different than the authors'. I think that probability neglect is an extreme form of insensitivity, and not something different as the authors think, expressed in their title ("versus"), and what they have as a central theme throughout their paper. Figure 7.1.1, p. 205, of Wakker (2010) shows the point, with to the left perfect sensitivity, in the middle partial sensitivity, and to the right extreme insensitivity which means probability neglect. Thus, what the authors take as evidence against inverse-S, in my opinion is strong support.

They also find higher elevation of probability weighting for affect-rich outcomes. It was not clear to me from the text and the formulas if higher elevation was coupled with more or with less risk aversion. Also, with only one nonzero outcome, elevation may be determined only up to one joint power for utility and probability weighting. This need not affect inverse-S but it does affect elevation. Adding assumptions about (the power of utility makes the power of probability weighting also indentifiable. %}

Suter, Renata, Thorsten Pachur, & Ralph Hertwig (2016) "How Affect Shapes Risky Choice: Distorted Probability Weighting Versus Probability Neglect,"*Journal of Behavioral Decision Making* 29, 437–449.

56

{% real incentives: not used; instead, flat payment

PT falsified through **coalescing**;

inverse-S: taking PT violations as they are, probability weighting seems to be inverse-S.

Finds violations of PT (= 1992 prospect theory; the author writes CPT) due to cancelling of common outcomes, which original 1979 prospect theory (OPT) can account for. I did not find definitions of the theories in the paper, and am not sure which version of OPT the author uses. P. 57 writes "whether or not the editing stage is frormalized"

Structure on p. 42, with r = q′-q, and s remaining probability.

R				S				
p	q	r	s	p	q	r	s	
$\underline{x}$	y	0	0	$\underline{x}$	y′	y′	0	A question
$\underline{y}$	y	0	0	$\underline{y}$	y′	y′	0	B question

The A question concerns choosing between

(p:$\underline{x}$, q:y, r:0, s:0) and (p:$\underline{x}$, q:y′, r:y′, s:0). In the B question, the underlined common outcome $\underline{x}$ has been replaced by a common outcome $\underline{y}$.

Cancellation here does not work to enhance the sure-thing principle, but differently: Consider, with majority preferences indicated in percentages

0.32	0.01	0.01	0.66		0.32	0.01	0.01	0.66		
3600	3500	0	0	[60%]	**3600**	2000	2000	0		Question A
3500	3500	0	0		**3500**	2000	2000	0	[78%]	Question A′

This violates the comonotonic sure-thing principle, and even Green & Jullien's ordinal independence. Explanation: in Question A, the common 3600 is ignored, and then the longshot effect gives overweighting of the best (of what remains) outcome 3500. In reality, the prospects are presented in collapsed form with outcome 0 not written. Then Question A′ becomes

(0:33: 3500) versus (0.32: 3500, 0.2: 2000) and there is no longshot perception for the best outcome 3500.

P. 42, ll. 7-8: "we believe that subjects are using this editing operation to simplify the gamble, thus reducing the complexity of the decision-making task."

P. 56, §3.2, discusses between versus within prospect heuristics.

P. 56 has nice balanced writing: "Although the results are not completely clean" %}

Wu, George (1994) "An Empirical Test of Ordinal Independence," *Journal of Risk and Uncertainty* 9, 39–60.

57

{% **PT falsified**: the authors claim that the weighting function for mixed prospects is less sensitive than that for pure gains or pure losses (**probability weighting depends on outcomes**). However, they don't have enough data to separate curvature from elevation (they assume only one weighting parameter that captures both) and also cannot separate it from loss aversion.

P. 1332 nicely writes on sign dependence: "Losses are not merely the opposite of gains, but gains and losses appear to be processed in different parts of the brain … and seem to be distinct psychologically, and not just to ends of a continuum"

%}

Wu, George & Alex B. Markle (2008) "An Empirical Test of Gain-Loss Separability in Prospect Theory," *Management Science* 54, 1322–1335.

58

{% https://doi.org/10.1007/s00426-018-1013-8

PT falsified: this paper re-analyzes classical evidence favoring loss aversion, such as Fishburn & Kochenberger (1979), showing many weak points in that evidence. It argues that loss aversion was found for high stakes, but not for small ones.

I imagine that for high stakes, concavity of utility for gains and fear of ruin for losses, rather than loss aversion, can be doing much. For small stakes, joy of gambling and peanut effect can distort. For intermediate outcomes, loss aversion is more manifest. The distinction between what is small and what is moderate in the author's terminology and in mine plays a big role here. I am more positive about loss aversion than the author. I think that loss aversion is strong and frequent, but, it is very volatile and can double or entirely disappear just by small changes in the stimuli. As components of decision attitudes become more volatile as they are more irrational. Loss aversion in the strict sense as I take it (only what results from reframing effects on reference point, and not "genuine" utility) is very volatile.

In the penultimate para, p. 1337, the author seems to argue that increased attention for losses is not loss aversion, and is not cognitive. I do not understand this para, and disagree. It can still be cognitive, and is as much part of loss aversion as strengthened feelings. Peeters & Czapinski (1990) give a good discussion of these two together comprising loss aversion. %}

Yechiam, Eldad (2019) “Acceptable Losses: The Debatable Origins of Loss Aversion,” *Psychological Research* 83, 1327–1339.

59

{% **losses give more/less noise**: several studies have found that choices under losses are more difficult and, hence, noisier than choices under gains (de Lara Resende, Guilherme, & Wu 2010 p. 129; Gonzalez, Dana, Koshino, & Just 2005 JEΨ; Lopes 1987). Somewhat different in spirit but not contradictory is that rewarding in terms of imposing losses to punish mistakes can work more effectively than imposing gains for good acts in making people make right choices. The presence of losses can make people pay more attention, improving decision quality.

PT falsified: this paper has an interesting experiment: people can choose between safe 35 and risky $200_{0.5}1$, and also between safe 35 and risky $200_{0.5}(-1)$. (Unit of outcome is points converted into small money amounts at the end of the experiment, with repeated payments, so income effects.) They more often choose risky in the second case, amounting to a violation of transitivity or stochastic dominance! The explanation is that the loss makes people pay more attention and, thus, they more rationally choose the highest expected value. This goes against the spirit of loss aversion. Interesting finding. They show that it is increased attention rather than contrast effect, because if the risky option has lower expected value then the loss makes people more often choose against the, now inferior, risky prospect. (**cognitive ability related to risk/ambiguity aversion**) Note that, in general, loss aversion can be generated by increased attention for losses (rather than losses having lower utility), but the above increased attention is of a different kind.

They also find Slovic/Birnbaum-type paradoxes where changing a zero outcome into a loss increases evaluation, which is one of these weird zero-outcome paradoxes.

The conclusion writes: “losses may be treated as signals of attention and not only as signals of avoidance. … Our findings demonstrate that the attentional effect of losses is indeed distinct from loss aversion,” %}

Yechiam, Eldad & Guy Hochman (2013) “Loss-Aversion or Loss-Attention: The Impact of Losses on Cognitive Performance,” *Cognitive Psychology* 66, 212–231.

OA.6. References for online appendices OA.1-OA.4

Ariely, Dan, George F. Loewenstein, & Drazen Prelec (2001) " 'Coherent Arbitrariness': Stable Demand Curves without Stable Preferences," *Quarterly Journal of Economics* 118, 73–106.

Baron, Jonathan, Zhijun Wu, Dallas J. Brennan, Christine Weeks, & Peter A. Ubel (2001) "Analog Scale, Magnitude Estimation, and Person Trade-Off as Measures of Health Utility: Biases and Their Correction," *Journal of Behavioral Decision Making* 14, 17–34.

Bernheim, B. Douglas & Charles Sprenger (2020) "On the Empirical Validity of Cumulative Prospect Theory: Experimental Evidence of Rank-Independent Probability Weighting," *Econometrica* 88, 1363–1409.

Birnbaum, Michael H. (2004) "Causes of Allais Common Consequence Paradoxes: An Experimental Dissection," *Journal of Mathematical Psychology* 48, 87–106.

Birnbaum, Michael H. (2008) "New Paradoxes of Risky Decision Making," *Psychological Review* 115, 463–501.

Carlin, Paul S. (1992) "Violations of the Reduction and Independence Axioms in Allais-Type and Common-Ratio Experiments," *Journal of Economic Behavior and Organization* 19, 213–235.

Dolan, Paul & Peep Stalmeier (2003) "The Validity of Time Trade-Off Values in Calculating QALYs: Constant Proportional Time Trade-Off versus the Proportional Heuristic," *Journal of Health Economics* 22, 445–458.

Fudenberg, Drew & Indira Puri (2022) "Simplicity and Probability Weighting in Choice under Risk," *AEA Papers and Proceedings* 112, 421–425.

Hardisty, David J., Katherine F. Thompson, David H. Krantz, & Elke U. Weber (2013) "How to Measure Time Preferences: An Experimental Comparison of Three Methods," *Judgment and Decision Making* 8, 214–235.

Kahneman, Daniel & Amos Tversky (1979) "Prospect Theory: An Analysis of Decision under Risk," *Econometrica* 47, 263–291.

Loomes, Graham, Chris Starmer, & Robert Sugden (2003) "Do Anomalies Disappear in Repeated Markets?," *Economic Journal* 113, C153–C166.

Schmeidler, David (1989) "Subjective Probability and Expected Utility without Additivity," *Econometrica* 57, 571–587.

Tversky, Amos & Craig R. Fox (1995) "Weighing Risk and Uncertainty," *Psychological Review* 102, 269–283.

Tversky, Amos & Daniel Kahneman (1992) "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty* 5, 297–323.

van de Kuilen, Gijs & Peter P. Wakker (2011) "The Midweight Method to Measure Attitudes toward Risk and Ambiguity," *Management Science* 57, 582–598.

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

Wakker, Peter P. (2022b) "The Correct Formula of 1979 Prospect Theory for Multiple Outcomes," *Theory and Decision*, forthcoming.

Weber, Elke U. & Britt Kirsner (1997) "Reasons for Rank-Dependent Utility Evaluation," *Journal of Risk and Uncertainty* 14, 41–61.