AEA Papers and Proceedings 2022, 112: 415–420
https://doi.org/10.1257/pandp.20221090

THE BEHAVIORAL ECONOMICS OF RISKY CHOICE: NEW PERSPECTIVES

Robustness of Rank Independence in Risky Choice†

By B. Douglas Bernheim, Rebecca Royer, and Charles Sprenger*

The famous Allais (1953) paradoxes challenge the validity of the independence axiom that lies at the heart of expected utility theory (EUT). As originally formulated, prospect theory (OPT, due to Kahneman and Tversky 1979) rationalizes such phenomena by allowing decision weights to vary nonlinearly with probabilities but thereby introduces violations of first-order stochastic dominance. Cumulative prospect theory (CPT, due to Tversky and Kahneman 1992) avoids this difficulty by assuming that the decision weight associated with an outcome depends on not only its probability but also its rank.

Bernheim and Sprenger (2020) present a novel experimental test of CPT. Specifically, for a given three-outcome lottery $L = (X, Y, Z; p, q, 1-p-q)$ with $Y > Z$, they measure discrete analogs of marginal rates of substitution between Z and Y (*equalizing reductions*). With rank-dependent probability weighting, the trade-off between Y and Z depends upon whether $X \geq Y > Z$, $Y > X \geq Z$, or $Y > Z > X$. As it turns out, the percentage change in the equalizing reduction as X crosses from one of these regimes to another identifies the degree of rank dependence in probability weights and, consequently, the shape of the CPT probability weighting function, nonparametrically. Bernheim and Sprenger (2020) present two main findings based on variation in equalizing reductions for a single set of tasks.

FINDING 1: *The probability weights for outcomes Y and Z do not vary meaningfully with their ranks (comparing cases with* $X \geq Y > Z$ *to cases with* $Y > X \geq Z$*).*

FINDING 2: *Responses to variation in probabilities with fixed ranks imply probability weights that match standard estimates based on experiments with two-outcome lotteries.*

Unlike OPT, CPT posits a form of rank-dependent probability weighting that makes these findings irreconcilable. Interpreted through the lens of CPT, Finding 1 implies that the probability weighting function must be linear over the pertinent range. But Finding 2 implies that the probability weighting function is highly nonlinear over the very same range.

Since the publication of Bernheim and Sprenger (2020), some proponents of CPT have raised the possibility that Finding 1 may be a consequence of experimental procedures rather than underlying preferences. Possible concerns include the following: (i) comprehension of the three-option lotteries may have been poor; (ii) the number of decision tasks may have been overwhelming; (iii) the stakes may have been too low; (iv) the analysis may have overlooked evidence of rank dependence associated with transitions between the regimes $Y > X \geq Z$ and $Y > Z > X$, which shed light on nonlinearities in probability weighting for probabilities near 1; and (v) the structure of the decision tasks may have triggered a heuristic involving the "cancellation" of a common outcome (X).[1]

*Bernheim: Department of Economics, Stanford University (email: bernheim@stanford.edu); Royer: Department of Economics, UC San Diego. (email: rroyer@ucsd.edu); Sprenger: Division of Humanities and Social Sciences, California Institute of Technology (email: sprenger@caltech.edu).

†Go to https://doi.org/10.1257/pandp.20221090 to visit the article page for additional materials and author disclosure statement(s).

[1] In their critique of Bernheim and Sprenger (2020), Abdellaoui et al. (2020) raise these issues and make a number of other points.

It is unlikely that the aforementioned considerations account for Bernheim and Sprenger's (2020) results. Concerns (i)–(iii)—complexity, fatigue, and low stakes—are not specifically entwined with rank dependence; they would tend to suppress any nuanced feature of decision making. But the setting is not so complex, fatiguing, or inconsequential that it fails to activate conventional probability weighting (Finding 2). Bernheim and Sprenger (2020) also addressed concern (ii) (fatigue) by presenting corroborating cross-subject results based on each subject's first task. Concern (iv) merely raises the possibility that Finding 1 might not hold on an unexamined portion of the probability domain; it cannot resolve the conflict that CPT implies between Findings 1 and 2 on the examined portion of that domain. Finally, Bernheim and Sprenger (2020) address concern (v) through a supplemental experiment with modified tasks that render the cancellation heuristic inapplicable. The supplemental experiment sacrifices some advantages of the original but still yields no evidence of rank-dependent probability weighting.

In this paper, we demonstrate that Finding 1 is indeed robust with respect to alternative experimental procedures that address each of the five concerns articulated above. Naturally, a comprehensive evaluation of CPT must consider other evidence concerning its validity. However, other existing tests suffer from serious confounds that the Bernheim-Sprenger approach avoids; see Bernheim and Sprenger (2020).[2]

I. Review and Extension of Methods

Regardless of whether the applicable theory is CPT, OPT, or EUT, we can write the indifference condition that defines the equalizing reduction for the lottery $L = (X, Y, Z; p, q, 1-p-q)$ as follows:

$$w_X u(X) + w_Y u(Y) + w_Z u(Z) = w_X u(X) + w_Y u(Y+m) + w_Z u(Z-k),$$

where w_s is the decision weight for $s \in \{X, Y, Z\}$. For small m, it follows that $\frac{k}{m} \approx \frac{w_Y u'(Y)}{w_Z u'(Z)}$. For any change from X'' to X', the associated k'' and k' therefore satisfy

$$\log(k') - \log(k'') \approx \log\left(\frac{w'_Y}{w'_Z}\right) - \log\left(\frac{w''_Y}{w''_Z}\right).$$

Thus, the percentage change in the equalizing reduction nonparametrically measures the percentage change in relative decision weights resulting from a change in X.

To determine whether the weights are rank dependent, we choose values X' and X'' so that the ranks of the outcomes differ. For $X'' > Y > Z$ and $Y > X' > Z$, CPT implies

$$\log(k') - \log(k'') \approx \log\left(\frac{\pi(q)}{q}\right) - \log\left(\frac{\pi(p+q) - \pi(p)}{q}\right).$$

In other words, under the maintained hypothesis of CPT, the percentage change in the equalizing reduction nonparametrically measures the percentage change in the slope of the probability weighting function between the intervals $[0, q]$ and $[p, p+q]$.

Bernheim and Sprenger (2020) found essentially no difference in equalizing reductions between the regimes $X'' > Y > Z$ and $Y > X' > Z$ for $p \in \{0.1, 0.4, 0.6\}$ (with $q = 0.3$), which means there is no evidence of rank-dependent probability weighting (Finding 1). Treating CPT as a maintained hypothesis, they concluded that the probability weighting function must be linear over the interval $[0, 0.9]$.

If CPT is valid, then one can also recover the probability weighting function by fixing ranks and varying probabilities. Defining $\phi = \left(\frac{q}{1-p-q}\right)\frac{m}{k}$ and using the approximation $\frac{k}{m} \approx \frac{w_Y u'(Y)}{w_Z u'(Z)}$ along with the definitions of w_Y and w_Z for CPT within the regime $Y > X > Z$, we see that for any change in p, say from p' to p,

$$\log(\phi) - \log(\phi') \approx \log\left(\frac{\pi(1) - \pi(p+q)}{1-p-q}\right) - \log\left(\frac{\pi(1) - \pi(p'+q)}{1-p'-q}\right).$$

[2] This approach improves significantly upon prior tests of comonotonic and noncomonotonic independence (see, e.g., Birnbaum and McIntosh 1996; Wu 1994; Wakker et al. 1994) by neutralizing important confounds and providing quantitative nonparametric measures of nonlinearities in the probability weighting function.

Table 1—Conditions for Measuring Equalizing Reductions

	Elicitation	Cancellation tasks	(X, X', X'')	(Y, Z)	Stakes multiplier	m	Visual display	Probability training	Incentivized	Order	# of subjects
Condition 1	Price list	Yes	(3, 22, 31)	(24, 18)	1x	5	Yes	Yes	1/5	Random	93
Condition 2	BDM	Yes	(3, 22, 31)	(24, 18)	1x	5	Yes	Yes	1/5	Random	109
Condition 3	BDM	Yes	(12, 88, 124)	(96, 72)	4x	5	Yes	Yes	1/20	Random	103
Condition 4	BDM	Yes	(12, 88, 124)	(96, 72)	4x	5	Yes	Yes	None	Random	111
Condition 5	BDM	Yes	(12, 88, 124)	(96, 72)	4x	20	Yes	Yes	1/20	Random	89
Condition 6	BDM	Yes	(48, 352, 496)	(384, 288)	16x	5	Yes	Yes	None	Random	89
Condition 7	BDM	Yes	(48, 352, 496)	(384, 288)	16x	80	Yes	Yes	None	Random	103

In other words, the percentage change in ϕ (which is measurable) is a nonparametric estimate of the percentage change in the average slope of the probability weighting function between the intervals $[p' + q, 1]$ and $[p + q, 1]$.

Bernheim and Sprenger (2020) found large differences in ϕ for $p \in \{0.1, 0.4, 0.6\}$ (with $q = 0.3$), which means there is evidence of substantial nonlinearities in probability weighting. In particular, their estimates imply that the probability weighting function is highly nonlinear throughout the interval $[0.4, 1]$ (Finding 2).

Treating CPT as a maintained hypothesis, Finding 1 and Finding 2 clearly have contradictory implications for the properties of the probability weighting function over the interval $[0.4, 0.9]$. Bernheim and Sprenger (2020) therefore reject CPT. Their findings are instead consistent with nonlinear rank-independent probability weighting.

We extend these methods in two ways. First, we also examine changes from X' to X satisfying $Y > X' > Z$ and $Y > Z > X$. For CPT, we have

$$\log(k) - \log(k') \approx \log\left(\frac{\pi(1) - \pi(p+q)}{1 - p - q}\right) - \log\left(\frac{\pi(1-p) - \pi(q)}{1 - p - q}\right).$$

In other words, under the maintained hypothesis of CPT, the percentage change in the equalizing reduction between the regimes $Y > X' > Z$ and $Y > Z > X$ nonparametrically measures the percentage change in the slope of the probability weighting function between the intervals $[q, 1-p]$ and $[p + q, 1]$. Thus, CPT can account for invariance of equalizing reductions across all three regimes only if the slope of the probability weighting function is constant (i.e., the function is linear) over the entire interval $[0, 1]$.

Second, we modify the method by eliciting k_+ and k_-, defined as follows:

$$w_X u(X-1) + w_Y u(Y) + w_Z u(Z) = w_X u(X) + w_Y u(Y+m) + w_Z u(Z - k_+)$$

$$w_X u(X) + w_Y u(Y) + w_Z u(Z) = w_X u(X-1) + w_Y u(Y+m) + w_Z u(Z - k_-).$$

For small m, we have $\frac{0.5(k_+ + k_-)}{m} \approx \frac{w_Y u'(Y)}{w_Z u'(Z)}$, so k and $k_M \equiv 0.5(k_+ + k_-)$ measure the same theoretical object. Intuitively, substituting k_M for k immunizes our method against the cancellation heuristic because the lotteries that define k_+ and k_- involve no common outcomes. In contrast to the supplemental experiment in Bernheim and Sprenger (2020), this method preserves all quantitative inferences concerning rank dependence.

II. Experimental Design

Our seven conditions all measure equalizing reductions (k) and modified equalizing reductions (k_+ and k_-) for the probability vector $(p, q, 1-p-q) = (0.4, 0.3, 0.3)$. Table 1 summarizes the main features of each condition; the online Appendix includes screenshots. We conducted the experiment in September 2021 on the Prolific platform using Otree software (Chen, Schonger, and Wickens 2016).

Condition 1 follows Bernheim and Sprenger (2020) in fixing $Y = \$24$, $Z = \$18$, and $m = \$5$. We used price lists to elicit equalizing reductions and modified equalizing reductions in random order for $X'' = \$31$, $X' = \$22$, and $X = \$3$. We display lottery distributions visually using the method of Lopes and Oden (1999). Subjects also receive training to facilitate their comprehension of the probabilities: they draw from the distribution 18 times and report their outcomes. This procedure allows the meaning of the probability distribution to "sink in" (Heffetz 2018). Thus, Condition 1 addresses concern (i) (comprehension of probabilities) through visual presentation and training, concern (iv) (limited scope) by encompassing all three regimes, and concern (v) (heuristic cancellation) by eliciting modified equalizing reductions.

Condition 2 is the same as Condition 1 except that it employs the titration Becker–DeGroot–Marschak (BDM) mechanism of Mazar, Koszegi, and Ariely (2014), wherein subjects first state a valuation, then review implications for options just below and just above the provisional point of indifference, then (potentially) revise their initial response. This procedure improves upon the original BDM mechanism by walking subjects through the contingent implications of their choices. It creates the same incentives as the corresponding price lists of Condition 1, but subjects make only 9 decisions rather than 585 component choices. This condition therefore addresses concern (ii) (decision fatigue) in addition to concerns (i), (iv), and (v).

The remaining five conditions follow the same procedures as Condition 2 but inflate X, Y and Z by a factor of 4 (Conditions 3, 4, and 5) or 16 (Conditions 6 and 7). The value of m is \$5 in Conditions 3, 4, and 6; \$20 in Condition 5; and \$80 in Condition 7. In other words, we either inflate m by the same factor as the outcomes or leave it fixed at \$5. These conditions address concern (iii) (small stakes).

Conditions 1, 2, 3, and 5 involve real choices. We paid 1 out of every 5 subjects based on one of their choices in Conditions 1 and 2 and paid 1 out of every 20 subjects in Conditions 3 and 5. Conditions 4, 6, and 7 involve hypothetical choices. This variation provides additional opportunities to test whether incentives induce rank-dependent behavior.

Our procedures prevent subjects' choices from switching back and forth between (X, Y, Z) and $(X, Y+m, Z-k)$ as k increases. This restriction has the advantage of yielding an unambiguous measure of k and (for Condition 1) of reducing each price list to a single choice, thereby minimizing decision fatigue. A disadvantage is that it sacrifices a potential indicator of poor comprehension (multiple switching). An alternate measure is whether the elicitations yield boundary values. Overall, 2.3 percent (13.9 percent) of observations take on the highest (lowest) value. Only 4.6 percent of subjects provide no interior values, and 80.2 percent provide 2 or fewer boundary values. Dropping these responses does not meaningfully change our findings.

III. Results

Table 2 and Figure 1 present our results. Condition 1 replicates the findings of Bernheim and Sprenger (2020). In both cases, the estimated change in decision weights between the regimes $Z < X' < Y$ and $Z < Y < X''$, $\log(k') - \log(k'')$, is close to zero. As before, we fail to reject the null hypothesis of rank independence (i.e., equality between k' and k'').

Condition 1 also extends the prior investigation by examining tasks with $X < Z < Y$. The differences between k, k', and k'' are small (on the order of 1 to 2 percent) and statistically insignificant, indicating the virtual absence of rank dependence. Treating CPT as a maintained hypothesis, one would conclude that the average slope of the probability weighting function is essentially unchanged between the intervals $[0, 0.3]$ and $[0.4, 0.7]$ as well as between the intervals $[0.3, 0.6]$ and $[0.7, 1]$.

Results based on k_+ and k_-, which are immune to the cancellation heuristic, corroborate the (near) rank independence of probability weights. Log differences in equalizing reductions imply that probability weights change only slightly (by 1 to 4 percent) due to a change in ranks. Critically, this finding does not reflect a tendency to cancel *approximately* common outcomes—i.e., to ignore the difference between X and $X-1$. As shown in Figure 1, values of k_+ are generally higher than values of k_-, and the difference is statistically significant ($\chi^2 = 20.94; p = 0.000$). However, the average of k_+ and k_- is statistically indistinguishable from k ($\chi^2 = 1.46; p = 0.228$), which suggests that cancellation is unimportant.

To put the preceding findings in context, Figure 1 also displays predicted values of k derived from the parameterized version of CPT due to Tversky and Kahneman (1992). According to this model, equalizing reductions should exhibit discontinuous increases

Table 2—Mean Equalizing Reductions and Estimated Rank Dependence

	Mean equalizing reduction			Rank dependence		
	Equalizing reduction–k $X < Z < Y$	Equalizing reduction–k' $Z < X' < Y$	Equalizing reduction–k'' $Z < Y < X''$	$\log(k) - \log(k')$	$\log(k) - \log(k'')$	$\log(k') - \log(k'')$
Bernheim Sprenger (2020) $(p,q,1-p-q) = (0.4,0.3,0.3)$, $(Y,Z) = (24,18)$		4.32 (0.12)	4.34 (0.12)			−0.01 (0.02)
Condition 1 (price list, 1x stakes, $m = \$5$)	4.25 (0.29)	4.27 (0.28)	4.35 (0.26)	−0.00 (0.06)	−0.02 (0.05)	−0.02 (0.05)
Cancellation tasks: $0.5(k_+ + k_-)$	4.05 (0.30)	4.21 (0.31)	4.29 (0.30)	−0.04 (0.06)	−0.06 (0.06)	−0.02 (0.07)
Condition 2 (BDM, 1x stakes, $m = \$5$)	3.50 (0.36)	3.57 (0.37)	3.67 (0.39)	−0.02 (0.07)	−0.05 (0.07)	−0.03 (0.06)
Cancellation tasks: $0.5(k_+ + k_-)$	3.53 (0.37)	3.60 (0.36)	3.55 (0.41)	−0.02 (0.08)	−0.00 (0.08)	0.02 (0.08)
Condition 3 (BDM, 4x stakes, $m = \$5$)	5.17 (0.35)	5.44 (0.39)	4.99 (0.33)	−0.05 (0.06)	0.04 (0.05)	0.09 (0.06)
Cancellation tasks: $0.5(k_+ + k_-)$	5.19 (0.39)	5.34 (0.43)	4.93 (0.37)	−0.03 (0.08)	0.05 (0.07)	0.08 (0.07)
Condition 4 (BDM, 4x stakes, $m = \$5$, hyp.)	4.67 (0.39)	4.39 (0.34)	4.32 (0.41)	0.06 (0.07)	0.08 (0.09)	0.02 (0.08)
Cancellation tasks: $0.5(k_+ + k_-)$	4.56 (0.40)	4.47 (0.37)	4.26 (0.44)	0.02 (0.08)	0.07 (0.10)	0.05 (0.09)
Condition 5† (BDM, 4x stakes, $m = \$20$)	3.42 (0.37)	3.52 (0.39)	4.07 (0.40)	−0.03 (0.10)	−0.17 (0.09)	−0.15 (0.07)
Cancellation tasks: $0.5(k_+ + k_-)$	3.25 (0.38)	3.50 (0.42)	4.11 (0.41)	−0.07 (0.13)	−0.23 (0.11)	−0.16 (0.09)
Condition 6 (BDM, 16x stakes, $m = \$5$, hyp.)	5.37 (0.47)	4.50 (0.46)	4.88 (0.43)	0.18 (0.09)	0.09 (0.08)	−0.08 (0.08)
Cancellation tasks: $0.5(k_+ + k_-)$	5.14 (0.53)	4.59 (0.49)	4.75 (0.44)	0.11 (0.11)	0.08 (0.08)	−0.03 (0.10)
Condition 7† (BDM, 16x stakes, $m = \$80$, hyp.)	4.85 (0.49)	4.55 (0.44)	4.44 (0.48)	0.06 (0.06)	0.09 (0.08)	0.03 (0.06)
Cancellation tasks: $0.5(k_+ + k_-)$	4.88 (0.50)	4.60 (0.45)	4.37 (0.47)	0.06 (0.06)	0.11 (0.08)	0.05 (0.06)
Aggregate values†	4.47 (0.15)	4.34 (0.15)	4.39 (0.15)	0.03 (0.03)	0.02 (0.03)	−0.01 (0.03)
Cancellation tasks: $0.5(k_+ + k_-)$	4.38 (0.16)	4.34 (0.15)	4.32 (0.16)	0.01 (0.03)	0.01 (0.03)	0.01 (0.03)

Notes: Mean values of k are estimated from interval regressions of experimental response on indicators for the rank of X, with titration BDM data converted to equivalent price list responses. † indicates normalized values, $k \cdot (5/m)$, for ease of comparison across m values. Rows titled "Cancellation tasks" are based only on modified equalizing reductions. Standard errors, in parentheses, are clustered at individual level and calculated using the delta method.

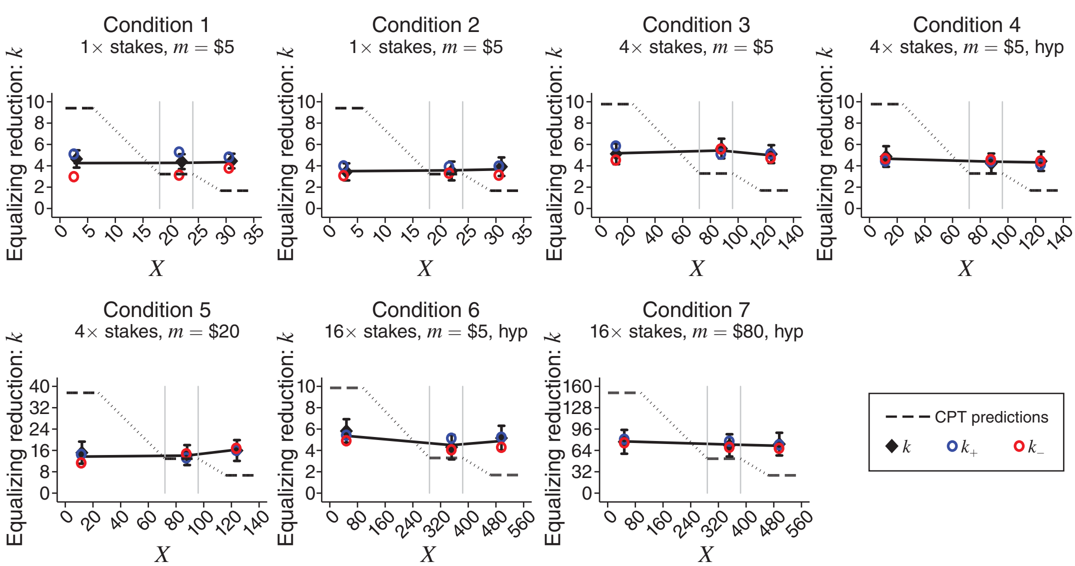

Figure 1. Mean Equalizing Reductions and Standard CPT Predictions

Notes: Black diamonds (with corresponding 95 percent confidence intervals) indicate mean values for k. Blue (red) circles indicate mean values for k_+ (k_-). Grey vertical lines mark values for outcomes Z and Y. Dashed lines indicate predicted values of k in each regime based on standard CPT parametrizations, calculated at X, X', and X''. Dotted lines connect CPT predictions between regimes.

across regimes (moving from $X'' > Y > Z$ to $Y > X' > Z$ to $Y > Z > X$) on the order of 66–176 percent.[3] Subjects display far less sensitivity to ranks than predicted. It is worth emphasizing that Bernheim and Sprenger (2020) found similar patterns of nonlinear probability weighting based on fixed-rank variation in probabilities for equalizing reduction tasks.

We similarly find no evidence of significant rank dependence in Condition 2, which aims to reduce decision fatigue by substituting a BDM mechanism for price lists. For Conditions 3 through 7, which vary the size of stakes and the nature of incentives, we generally find that the actual changes in relative probability weights are less than 10 percent, compared with the CPT predictions of 66–176 percent. Out of the 42 comparisons in the "Rank dependence" portion of Table 2, we reject the null hypothesis of rank independence at the 5 percent level in only 3 cases. In two of these three, the actual change in the equalizing reduction and the CPT prediction have opposite signs. We also provide an omnibus measure of rank dependence aggregating all of our conditions and fail to reject the null hypothesis of rank independence overall. Thus, we confirm that Finding 1 of Bernheim and Sprenger (2020) is robust.

REFERENCES

Abdellaoui, Mohammed, Chen Li, Peter P. Wakker, and George Wu. 2020. "A Defense of Prospect Theory in Bernheim & Sprenger's Experiment." Unpublished.

Allais, Maurice. 1953. "Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Americaine." *Econometrica* 21 (4): 503–46.

Bernheim, B. Douglas, and Charles Sprenger. 2020. "On the Empirical Validity of Cumulative Prospect Theory: Experimental Evidence of Rank-Independent Probability Weighting." *Econometrica* 88 (4): 1363–409.

Chen, Daniel L., Martin Schonger, and Chris Wickens. 2016. "oTree: An Open-Source Platform for Laboratory, Online, and Field Experiments." *Journal of Behavioral and Experimental Finance* 9: 88–97.

Heffetz, Ori. 2018. "Are Reference Points Merely Lagged Beliefs Over Probabilities." NBER Working Paper 24721.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47 (2): 263–92.

Lopes, Lola L., and 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 (2): 286–313.

Mazar, Nina, Botond Koszegi, and Dan Ariely. 2014. "True Context-Dependent Preferences? The Cause of Market-Dependent Valuations." *Journal of Behavioral Decision Making* 27 (3): 200–208.

Tversky, Amos, and Daniel Kahneman. 1992. "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk and Uncertainty* 5 (4): 297–323.

[3] The parameterization assumes that $u(x) = x^{0.88}$ and $\pi(p) = p^{0.61}/\left(p^{0.61} + (1-p)^{0.61}\right)^{1/0.61}$. The changes in equalizing reductions (107 percent, 173 percent, and 66 percent for the 3 comparisons in the "Rank dependence" portion of Table 2) closely approximate the changes in probability weights (110 percent, 176 percent, and 66 percent, respectively).