ON THE EMPIRICAL VALIDITY OF CUMULATIVE PROSPECT THEORY: EXPERIMENTAL EVIDENCE OF RANK-INDEPENDENT PROBABILITY WEIGHTING

B. DOUGLAS BERNHEIM
Department of Economics, Stanford University and NBER

CHARLES SPRENGER
Rady School of Management and Department of Economics, University of California San Diego

Cumulative Prospect Theory (CPT), the leading behavioral account of decisionmaking under uncertainty, avoids the dominance violations implicit in Prospect Theory (PT) by assuming that the probability weight applied to a given outcome depends on its ranking. We devise a simple and direct nonparametric method for measuring the change in relative probability weights resulting from a change in payoff ranks. We find no evidence that these weights are even modestly sensitive to ranks. Conventional calibrations of CPT preferences imply that the percentage change in probability weights should be an order of magnitude larger than we observe. It follows either that probability weighting is not rank-dependent, or that the weighting function is nearly linear. Nonparametric measurement of the change in relative probability weights resulting from changes in probabilities rules out the second possibility. Additional tests nevertheless indicate that the dominance patterns predicted by PT do not arise. We reconcile these findings by positing a form of complexity aversion that generalizes the well-known certainty effect.

KEYWORDS: Prospect Theory, Cumulative Prospect Theory, rank dependence, certainty equivalents.

1. INTRODUCTION

PROSPECT THEORY (PT), as formulated by Kahneman and Tversky (1979), provides a flexible account of decision making under uncertainty that accommodates a wide variety of departures from the expected utility (EU) paradigm. As a result, it has been enormously influential throughout the social sciences. In contrast to the EU formulations of von Neumann and Morgenstern (1944), Savage (1954), and Samuelson (1952), a central premise of PT holds that attitudes toward objective probabilities display nonlinearities, with highly unlikely events receiving greater proportionate weight than nearly certain ones. This feature reconciles PT with important behavioral puzzles such as the famous Allais (1953) paradoxes, as well as the simultaneous purchase of lottery tickets and insurance, as in Friedman and Savage (1948). Probability weighting is also well-supported by simple and widely-replicated laboratory experiments.[1]

[1]For example, when graphing the empirical certainty equivalent, C, for a lottery that pays X with probability p and 0 with probability $1 - p$, one typically finds an inverse S-shaped pattern, with pX exceeding C for moderate-to-large values of p (as risk aversion would imply), but with the opposite relation for small p (see, e.g., Tversky and Kahneman (1992), Tversky and Fox (1995)).

Unfortunately, the formulation of probability weighting embedded in PT leads to conceptual difficulties because it implies violations of first-order stochastic dominance even in relatively simple settings. This is a serious flaw given the broad consensus that this property renders a model of decisionmaking unappealing on both positive and normative grounds.[2] To understand the problem, consider a lottery that pays X with probability p; for our current purpose, we will leave other events and payoffs unspecified. Now imagine a second lottery, identical to the first, except that it splits the aforementioned event, paying X and $X - \varepsilon$ each with probability $p/2$.[3] Given the S-shape of the probability weighting function, we can choose p so that the total weight assigned to two events occurring with probability $p/2$ discretely exceeds the weight assigned to a single event occurring with probability p. Consequently, if X is large and/or ε is small, the first lottery will yield lower PT utility than the second even though it is clearly preferable based on first-order stochastic dominance.[4]

Ultimately, "rank-dependent" probability weighting was offered as a solution to the stochastic dominance problem (Quiggin (1982)), and was incorporated into a new version of PT known as Cumulative Prospect Theory, henceforth CPT (Tversky and Kahneman (1992)). To understand intuitively how CPT resolves the issue, consider a lottery L with three possible payoffs, $X > Y > Z$, occurring with probabilities p, q, and $1 - p - q$. Another description of the same lottery involves cumulative probabilities: it pays Z with probability 1, adds $Y - Z$ with probability $p + q$, and then incrementally adds $X - Y$ with probability p. Accordingly, within the EU framework, one could write its expected utility as follows:

$$\text{Expected Utility} = u(Z) + (p + q)\big(u(Y) - u(Z)\big) + p\big(u(X) - u(Y)\big).$$

CPT involves an analogous calculation, except that a reference-dependent utility function, $u(\cdot|r)$ (where r is the reference point), is applied to the payoffs, while a weighting function, $\pi(\cdot)$, is applied to the cumulative probabilities:[5]

$$U(L) = \pi(1)u(Z|r) + \pi(p + q)\big[u(Y|r) - u(Z|r)\big] + \pi(p)\big[u(X|r) - u(Y|r)\big].$$

---

[2]As noted by Quiggin (1982), "Transitivity and dominance rules command virtually unanimous assent... even from those who sometimes violate them in practice... If a theory of decision under uncertainty is to be consistent with any of the large body of economic theory which has already been developed... it must satisfy these rules." (p. 325).

[3]Kahneman and Tversky (1979) described their theory as being concerned with lotteries that have at most two nonzero outcomes. Hence, to apply Prospect Theory strictly in accordance with their original intent, one would have to assume that this lottery pays zero with probability $1 - p$. Kahneman and Tversky (1979, p. 288) noted that the model extends naturally to more than two nonzero outcomes, and extensions which correspond to our three-outcome formulation are provided by, for example, Camerer and Ho (1994) and Fennema and Wakker (1997). Kahneman and Tversky (1979) actually provided two formulations of Prospect Theory; we assume their Equation (1) for 'regular prospects.' They implicitly invoked the same assumption when examining the Allais common consequence paradox (p. 282).

[4]Kahneman and Tversky appreciated this problematic implication of PT and attempted to address it through an "editing" assumption: "Direct violations of dominance are prevented, in the present theory, by the assumption that dominated alternatives are detected and eliminated prior to the evaluation of prospects" (p. 284). Most economists have found this ad hoc "fix" conceptually unsatisfactory, and it is rarely invoked in applications. 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)). One can see in that formulation the roots of Cumulative Prospect Theory.

[5]For simplicity, we assume here that the reference point, r, is below the other payoffs.

Normally this expression is rewritten in a form that attaches a weight to each outcome:

$$U(L) = \pi(p)u(X|r) + \big[\pi(p+q) - \pi(p)\big]u(Y|r) + \big[\pi(1) - \pi(p+q)\big]u(Z|r). \quad (1)$$

Now imagine, as before, a second lottery, identical to the first, except that it splits the event yielding the payoff X into two events paying X and $X - \varepsilon$, each with probability $p/2$. In that case, the term $\pi(p/2)u(X|r) + [\pi(p) - \pi(p/2)]u(X - \varepsilon|r)$ replaces the term $\pi(p)u(X|r)$. Notice that the *total* weight assigned to the two events is still $\pi(p)$, the same as for the original lottery. Consequently, the stochastic dominance problem noted above does not arise (Quiggin (1982), Tversky and Kahneman (1992)). CPT nevertheless accommodates the same assortment of EU violations as PT. For these reasons, CPT has replaced PT as the leading behavioral model of decisionmaking under uncertainty.

To understand the sense in which CPT involves *rank-dependent* probability weighting, consider the weight applied to the event that generates the payoff Y as we change the value of X. Initially X exceeds Y, and the weight on Y is $\pi(p+q) - \pi(p)$. As we reduce the value of X, this weight remains unchanged until X passes below Y, at which point it changes discontinuously to $\pi(q)$. Thus, the weight assigned to the event depends not only on probabilities, but also on the *ranking* of the event according to the size of the payoff.

The current paper devises and implements a simple and direct approach to measuring the change in probability weights resulting from a change in payoff ranks. Our method is entirely nonparametric in the sense that it requires no maintained assumptions concerning functional forms, either for utility and risk aversion, or for probability weighting.

An essential feature of our method is that it involves lotteries with three outcomes. To understand why the presence of a third outcome facilitates a sharp and powerful test of the premise, consider equation (1). For any small increase (m) in the value of Y, there is a small *equalizing reduction* (k) in the value of Z that leaves the decisionmaker indifferent. This equalizing reduction measures the marginal rate of substitution between Y and Z, capturing relative probability weights. Both EU theory and PT imply that *the magnitude of the equalizing reduction is entirely independent of the value of X, regardless of functional forms*. The same is true for CPT, provided X remains within two ranges that we empirically examine, $X > Y + m$ and $Y > X > Z$ (as well as within the range $Z > X$). The reason is that, under CPT, the marginal rate of substitution between Y and Z, written MRS_{YZ}, depends only on the values, probabilities, and ranks of Y and Z, none of which change. However, as the value of X crosses from one of these ranges into the other, the ranking of Y changes, which causes the probability weight on Y to change, while the other factors that determine MRS_{YZ} (the probabilities and values of Y and Z, and the rank of Z) remain fixed. As a result, MRS_{YZ} changes discontinuously, producing a discontinuous change in the equalizing reduction. Critically, we show (for small changes) that *the percentage change in the equalizing reduction when X passes through Y precisely measures the percentage change in the probability weight applied to Y resulting from the change in Y's rank.* Thus, our strategy is to quantify the extent of rank dependence in probability weights by eliciting equalizing reductions for $X > Y + m$ and $X \in (Z, Y)$.

Subjects in our experiment perform decision tasks that reveal their equalizing reductions for three-outcome lotteries of the type described above. We find no evidence that probability weights are even modestly sensitive to the ranking of outcomes. The actual percentage change in the equalizing reductions, and hence probability weights, ranges from $+3\%$ to -3%, and in no case can we reject the hypothesis of rank independence. Our estimates rule out changes in probability weights larger than 7% as ranks change with 95% confidence.

It follows from these results either that probability weighting is not rank-dependent, in which case CPT is predicated on a false assumption, or that (contrary to other estimates) the weighting function is nearly linear, in which case CPT does not differ from PT. To distinguish between these possibilities, we devise and implement a nonparametric test of the hypothesis that the probability weighting function is linear. The test exploits responses in the same tasks to variations in probabilities, holding ranks fixed. Using this alternative source of variation, we find evidence of substantial nonlinearities. For example, responses to changes in probabilities (with fixed ranks) imply that the average slope of the probability weighting function is roughly 19% lower on the subinterval of probabilities $[0.4, 0.7]$ than on the subinterval $[0.7, 0.9]$, despite the fact that—under the maintained hypothesis of rank dependence—the absence of responses to change in ranks (with fixed probabilities) implies a constant slope for the probability weighting function over the subinterval $[0.4, 0.9]$. Adopting standard functional assumptions, we then show that the estimated degree of curvature differs sharply depending on whether one draws inferences from responses to variations in payoff ranks or variations in probabilities. Using parametric models estimated based on responses to variations in probabilities with fixed ranks, we predict the degree to which equalizing reductions should change in response to rank changes under the assumption of rank dependence. We also perform these calculations using data from conventional CPT elicitation tasks. In all cases, the predicted changes are an order of magnitude greater than the observed changes, and the confidence intervals are non-overlapping. For example, the conventional CPT calibration implies that increases in X that change the rankings of X and Y in our experiment should change the equalizing reductions by -22% to -46%, even though the actual change is negligible. Thus, the degree of curvature in the probability weighting function implied by responses to variations in probabilities is highly inconsistent with the virtual invariance of equalizing reductions with respect to ranks except under the hypothesis that probability weights are rank-independent. Indeed, parametric estimates show that the PT formulation of probability weighting accounts for the data on equalizing reductions more successfully than the CPT or EU (linear) formulations.

Similar patterns are also apparent at the individual level, with a preponderance of subjects exhibiting virtually no rank dependence for their probability weights, despite responding to changes in probabilities in ways that imply substantial curvature of their probability weighting functions, and hence substantial rank dependence within the CPT framework. The results are robust with respect to a variety of alternative analytic procedures, such as using only between-subject variation and eliminating potentially confused subjects. We also demonstrate that our methods are robust with respect to alternative assumptions about reference points. Endogenizing reference points (as in Bell (1985), Loomes and Sugden (1986)) changes nothing of substance. Significantly, even with linear probability weighting, models with reference distributions (Koszegi and Rabin (2006, 2007)) have similar predictions for equalizing reductions, and hence we falsify them as well.

Our experimental design elicits equalizing reductions through choices over lotteries with a single common outcome, X. A pair of early papers in this area raised the possibility that subjects may employ a heuristic that involves the cancellation of common outcomes (Wu (1994), Weber and Kirsner (1997)).[6] Under that ancillary hypothesis, our

---

[6]Weber and Kirsner (1997) provided evidence from certainty equivalent tasks where no cancellation is possible. They found more support for models of rank dependence when comparing certainty equivalents for lotteries than when comparing choices between the lotteries themselves. We thank an anonymous referee for

method would produce spurious evidence of rank independence. We address this possibility by examining a similar decision setting in which no cancellation is possible: we add m to X instead of to Y, and reduce *both* Y and Z by k. CPT rank dependence predicts discontinuities in k of *opposite signs* as X passes from $X > Y > Z$ to $Y > X' > Z$ to $Y > Z > X''$.[7] For this modified decision task, we again find no evidence of CPT rank dependence, clearly refuting the cancellation hypothesis as a rationale for our results.

It is worth emphasizing that the stunning failure of CPT to account for our data is not a mere technical shortcoming. Our test focuses on a first-order implication of the theory; indeed, it isolates the critical feature that distinguishes CPT from PT. To put the matter starkly, if equalizing reductions in three-outcome lotteries are not rank-dependent, then neither are probability weights, and the CPT agenda is potentially on the wrong track.

What type of model should behavioral economists consider in place of CPT? One possibility is that PT is correct, and that people actually exhibit the implied violations of first-order stochastic dominance. We test this possibility with a third experiment eliciting certainty equivalents for three-outcome lotteries that pay $X + \varepsilon$ with probability $p/2$, $X - \varepsilon$ with probability $p/2$, and Y with probability $1 - p$. We include the case of $\varepsilon = 0$, which reduces to a two-outcome lottery. We choose the parameters so that standard formulations of PT predict a sizable and discontinuous drop in the certainty equivalent at $\varepsilon = 0$. In contrast, CPT implies continuity. Contrary to both predictions, we find a discontinuous *increase* in the certainty equivalent at $\varepsilon = 0$. This behavior implies violations of dominance, but not the type PT predicts.

A good theory of choice under uncertainty would therefore have to account for three patterns: (1) robust evidence of probability weighting based on behavioral responses to variations in probabilities, (2) the absence of rank dependence in equalizing reductions, and (3) the sharp drop in certainty equivalents that results from splitting an event. EU is inconsistent with (1) and (3). In light of (1), CPT is inconsistent with (2) and (3), and PT is inconsistent with (3). We hypothesize that the observed behavior results from a combination of standard PT and a form of complexity aversion: people may prefer lotteries with fewer outcomes because they are easier to understand. One can think of the well-known *certainty effect* as a special case of this more general phenomenon.

Readers sympathetic to the hypothesis of rank dependence may wonder whether an enhanced CPT model with additional degrees of freedom might account for our data. Our analysis rules out the possibility of achieving that objective through alternative assumptions concerning the location of reference points, or by invoking the heuristic cancellation of common outcomes. However, other novel hypotheses may bear investigation.

The current paper is most closely related to a handful of studies that aim to test the axiomatic foundations of rank-dependent models (Wu (1994), Wakker, Erev, and Weber (1994), Fennema and Wakker (1996), Weber and Kirsner (1997), Birnbaum (2008)). Unlike our approach, the methods used in these papers do not yield estimates of the degree to which probability weights depend on payoff ranks (nonparametric or otherwise), and the conclusions the authors draw from them do not necessarily follow in settings with noisy choices; see Section 2.3 and Appendix B of the Supplemental Material (Bernheim and Sprenger (2020)) for details.[8]

---

drawing our attention to this work and inspiring this modification. Our 'split-event' experiments discussed in Section 5 also explore the forces of rank dependence without the potential confound of cancellation.

[7]We did not design the main portion of our investigation around these types of decision tasks because EU, PT, and CPT all imply that the associated value of k should vary with X even when ranks do not change. This variation complicates the task of reliably measuring the change in k when ranks do change.

[8]An alternate strand of literature in psychology tests other CPT implications apart from rank dependence, such as adherence to stochastic dominance, consistency of behavior across different ranges of probabilities,

Aside from the aforementioned studies, the assumption of rank-dependent probability weighting has been the subject of surprisingly little formal scrutiny given its central role in the leading behavioral theory of decisionmaking under uncertainty, as well as in recent applications of the theory.[9] The literature has focused instead on identifying the shapes of CPT functions and associated parameter values based on choices involving binary lotteries (Tversky and Kahneman (1992), Tversky and Fox (1995), Wu and Gonzalez (1996), Gonzalez and Wu (1999), Abdellaoui (2000), Bleichrodt and Pinto (2000), Booij and van de Kuilen (2009), Booij, van Praag, and van de Kuilen (2010), Tanaka, Camerer, and Nguyen (2010)). In cases where the experimental tasks encompass an appropriate range of binary lotteries, one can devise and implement tests of rank dependence, conditional on maintained assumptions about functional forms. Unfortunately, an incorrect functional specification can manifest as spurious rank dependence. To our knowledge, in cases where such data are available, no formal test of rank dependence has been performed.[10] Defenses of rank dependence, such as the discussion in Diecidue and Wakker (2001), are instead typically based on intuitive arguments and/or point to findings concerning the psychology of decisionmaking that arguably resonate with the premise (Lopes (1984), Lopes and Oden (1999), Weber (1994)).

The paper proceeds as follows. Section 2 outlines the pertinent implications of CPT and related theories. Section 3 describes our experimental design, while Section 4 presents our main results, and Section 5 describes various robustness checks. Section 6 discusses implications, including alternative theories and tests thereof. Section 7 concludes.

2. THEORETICAL CONSIDERATIONS

Let $L = (\{p, q, 1 - p - q\}, \{X, Y, Z\})$ represent a lottery with three potential outcomes, X, Y, and Z, played with corresponding probabilities p, q, and $1 - p - q$, with $p, q \geq 0$ and $1 - p - q \leq 1$. EU, PT, and CPT all assume that preferences over such lotteries have

---

separate weighting of gains and losses, and invariance to lottery description, also showing deviations (for a broad review of these exercises, see Birnbaum (2008)).

[9]Barseghyan, Molinari, O'Donoghue, and Teitelbaum (2015) investigated choices involving a range of insurance products. They demonstrated that the bracketing of risks—for example, whether people consider home and automobile insurance together or separately—affects the implications of probability weighting because it changes the ranking of outcomes. Epper and Fehr-Duda (2015) examined the data from Andreoni and Sprenger (2012) on intertemporal decisionmaking under various risk conditions, which exhibits deviations from discounted expected utility. They argued that CPT can rationalize an apparent choice anomaly if one frames two independent binary intertemporal lotteries as a single lottery with four possible outcomes. This alternative framing delivers the desired prediction because it alters the rankings of the four outcomes. Barberis, Mukherjee, and Wang (2016) examined historical monthly returns at the stock level for a five-year window and linked the CPT value of the stock's history to future returns, demonstrating a significant negative correlation. The interpretation for the negative relation is that investors overvalue positively skewed, lottery-like stocks. Given 60 equi-probable monthly return events, PT would equally overweight all outcomes, giving no disproportionate value for skewness. CPT, on the other hand, allows the highest ranked outcomes to receive higher proportionate weight. Barberis, Mukherjee, and Wang (2016) showed that CPT substantially outperforms EU in predicting future returns. Given that the PT formulation (ignoring the reference point) would be collinear with the EU formulation, rank dependence would seem critical for delivering this result.

[10]As we explain in Appendix A of the Supplemental Material (Bernheim and Sprenger (2020)), the data in Tversky and Kahneman (1992) lend themselves to such tests. We show that the data from Tversky and Kahneman (1992) could be interpreted as consistent with rank dependence. However, as noted in Appendix A of the Supplemental Material (Bernheim and Sprenger (2020)), that finding hinges on the validity of their functional form assumptions. We show that, depending on the assumptions for the shape of utility, probability weighting for a given chance of receiving an outcome can either appear to be rank-dependent or not.

the following separable form:

$$U(L) = w_X u(X) + w_Y u(Y) + w_Z u(Z),$$

where w_i represents the decision weight for outcome i.[11] Under EU and PT, w_i is a fixed number that depends only on the probability of event i, and not on the ranks of X, Y, or Z. Under CPT, w_i depends on the probabilities of the three events and the *ranks* of the payoffs. For PT and CPT, our notation suppresses the dependence of $u(\cdot)$ on the reference point, which for simplicity we take as fixed and assume for the moment to be less than X, Y, and Z. We address alternative assumptions about reference dependence in Section 2.2.

Our analysis employs the concept of an *equalizing reduction*, defined as the value of k that delivers indifference between the lottery L and a modified lottery $L_e = (\{p, q, 1 - p - q\}, \{X, Y + m, Z - k\})$, where m is a (small) fixed number. Intuitively, the equalizing reduction approximates the marginal rate of substitution between the payoffs Y and Z (MRS_{YZ}). EU and PT imply that w_Y and w_Z, and therefore MRS_{YZ} for fixed values of Y and Z, are completely independent of X. CPT shares this implication as long as variations in X do not change the payoff ranks. However, as X crosses Y, the rank of Y changes, while all the other factors that determine w_Y and w_Z under CPT (the probabilities of Y and Z, and the rank of Z) remain fixed. To the extent rank dependence is quantitatively important, CPT therefore implies that MRS_{YZ} changes discontinuously at $X = Y$.

The preceding intuition suggests an empirical strategy for evaluating the importance of rank dependence. Assuming m and k are small, so that the payoff ranks are the same for L and L_e, then under all three theories we have

$$w_Y u(Y) + w_Z u(Z) = w_Y u(Y + m) + w_Z u(Z - k), \tag{2}$$

or alternatively,

$$k = Z - u^{-1}\left[u(Z) + \frac{w_Y}{w_Z}\big(u(Y) - U(Y + m)\big) \right]. \tag{3}$$

Suppose we assess k for various values of X. Under EU and PT, $\frac{w_Y}{w_Z}$ is a fixed number (as are Z, Y, and m), so a graph of k against X should be a flat line. Under CPT, $\frac{w_Y}{w_Z}$ is fixed as long as the payoff ranking is preserved, but it changes discontinuously when the value of X passes through the value of Y or Z. Therefore, a graph of k against X should exhibit three flat line segments with discontinuities at $X = Y$ and $X = Z$.[12]

The strength of our approach is that it yields more than a *qualitative* way to gauge the importance of rank dependence—it also provides a *quantitative*, nonparametric estimate of the change in relative decision weights that results from a change in payoff ranks. Say we obtain $\bar{k}$ using the value $\bar{X}$, and $\underline{k}$ using the value $\underline{X}$. Defining $\Delta \log(a) = \log(\bar{a}) - \log(\underline{a})$ (for the generic variable a), we have the following:[13]

---

[11]Our application of PT to these three-outcome lotteries corresponds to the extension of PT provided by, for example, Camerer and Ho (1994) and Fennema and Wakker (1997).

[12]Technically, the discontinuities occur at Y and Z in the limit as m goes to zero.

[13]As noted by one of our referees, it is relatively straightforward to dispense with the assumption that u is differentiable. Continuity and monotonicity of $u(\cdot)$ are sufficient for the existence of positive right-derivatives, which cancel out in the limit.

PROPOSITION 1: *Suppose the reference point is fixed, that decision weights are fixed for a given payoff ranking, and that u is continuously differentiable at Y and Z.*[14] *Consider any $\overline{X}$ and $\underline{X}$ distinct from Y and Z. Then*

$$\lim_{m \to 0} \Delta \log(k) = \Delta \log\left(\frac{w_Y}{w_Z}\right).$$

PROOF: See the Appendix.   *Q.E.D.*

Proposition 1 tells us that the percentage change in k (from $\overline{k}$ to $\underline{k}$) provides a quantitative estimate of the percentage change in the relative decision weights, $\frac{w_Y}{w_Z}$ (from $\frac{\overline{w}_Y}{\overline{w}_Z}$ to $\frac{\underline{w}_Y}{\underline{w}_Z}$), resulting from the change in X (from $\overline{X}$ to $\underline{X}$).

To drive the implications of this point home, suppose in particular that we choose $\overline{X}$ and $\underline{X}$ such that $\overline{X} > Y + m > Z$ and $Y > \underline{X} > Z$. Then, under CPT, we have $\overline{w}_Y = \pi(p + q) - \pi(p)$ and $\overline{w}_Z = 1 - \pi(p + q)$, while $\underline{w}_Y = \pi(q)$ and $\underline{w}_Z = 1 - \pi(p + q)$. It follows that

$$\Delta \log(k) \approx \log\big(\pi(p + q) - \pi(p)\big) - \log\big(\pi(q)\big).$$

Thus, for the maintained hypothesis of CPT, in this special case the percentage change in k provides a quantitative estimate of the percentage change in the probability weight assigned to payoff Y when the value of X passes from above Y to below Y.[15]

Accordingly, the first step in our analysis is to measure changes in the relative decision weights, $\frac{w_Y}{w_Z}$, associated with reversals in the ranks of X and Y. If we find that these changes are robustly close to zero, we can conclude that either the weights are not rank-dependent, in which case CPT is predicated on a false assumption, or the probability weighting function exhibits no meaningful nonlinearities, in which case CPT does not differ from PT.

To be precise, suppose that for specified values of the payoffs, Y and Z, and of the probabilities, p, q, and $1 - p - q$, we find $\Delta \log(k) \approx 0$. Treating rank dependence as a

---

[14]The continuous differentiability requirement rules out cases in which Y or Z coincides with the reference point. The proof extends to these cases but requires attention to some additional technical details.

[15]It is natural to wonder whether our central insight would apply to models in which probability weighting functions include linear segments, so that $\log(\pi(p + q) - \pi(p)) - \log(\pi(q)) = 0$ over a given range. One prominent example is the neo-additive model of Chateauneuf, Eichberger, and Grant (2007). Under the neo-additive model with objective probabilities, decision weights for cumulative probabilities away from 0 and 1 are linear as in expected utility, but extra weight is given to the best and worst outcome in a lottery. For $Z < \underline{X} < Y$, the neo-additive utility is

$$U = \gamma u(Z) + (1 - \gamma - \lambda)\big[pu(\underline{X}) + qu(Y) + (1 - p - q)u(Z)\big] + \lambda u(Y),$$

where γ and λ represent the additional weight on the worst and best outcomes, respectively. In contrast, for $Z < Y < \overline{X}$,

$$U = \gamma u(Z) + (1 - \gamma - \lambda)\big[pu(\overline{X}) + qu(Y) + (1 - p - q)u(Z)\big] + \lambda u(\overline{X}).$$

Because outcome Y is no longer the best outcome, its weight changes discontinuously. It is straightforward to show that

$$\Delta \log(k) \approx \log\left(\frac{(1 - \gamma - \lambda)q}{\lambda + (1 - \gamma - \lambda)q}\right),$$

so a discontinuity in equalizing reductions is predicted, and the log change in equalizing reduction again closely approximates the change in decision weight applied to outcome Y.

maintained hypothesis, we have $\Delta\omega_Z = 0$ by construction, so if we find no change in k, we must also have $\Delta\omega_Y = [\pi(p+q) - \pi(p)] - \pi(q) = 0$. Using this equation along with the assumption that $\pi(0) = 0$ (impossible events are ignored), we obtain

$$\frac{\pi(p+q) - \pi(p)}{q} = \frac{\pi(q) - \pi(0)}{q}. \tag{4}$$

In other words, $\Delta\log(k) = 0$ implies that the average slope of $\pi(\cdot)$ is the same over the intervals $[0, q]$ and $[p, p+q]$. Taking q small, we see that, if this condition holds for all $p \in [0, 1-q]$, then $\pi(\cdot)$ must be linear. In our experiment, we focus on values of p and q that allow us to target the portions of the unit interval for which previous analyses of probability weighting have found pronounced nonlinearities. Using our methods, one could obviously consider additional values of p and q, effectively blanketing the interval with these tests, thereby ruling out nonlinearities more comprehensively under the maintained hypothesis of rank dependence.[16]

Upon determining that $\Delta\log(k)$ is in fact robustly close to zero, we proceed to the second step of our analysis, asking whether this result follows from the absence of rank dependence, or from the absence of meaningful nonlinearities in the probability weighting function. Our strategy is to draw inferences about the shape of the probability weighting function from responses *in the same tasks* to variations in probabilities, holding ranks fixed. We reject rank-dependent probability weighting, and hence CPT, if Steps 1 and 2 yield inconsistent conclusions concerning the shape of $\pi(\cdot)$. For example, if we rule out meaningful nonlinearities in Step 1 (conditional on maintaining rank dependence) but find them in Step 2, then CPT cannot account for the pattern of equalizing reductions. However, the same pattern is consistent with PT, because the hypothesized finding for Step 1 rules out nonlinearities only if one assumes rank dependence.

For Step 2, we proceed as follows. Using the measured equalizing reductions for the same tasks, we can compute the following quantities for multiple values of p and/or q:

$$\overline{\phi}(p, q) \equiv \left(\frac{q}{1-p-q}\right)\frac{m}{\overline{k}} \quad \text{and} \quad \underline{\phi}(p, q) \equiv \left(\frac{q}{1-p-q}\right)\frac{m}{\underline{k}}.$$

Noting that $\frac{k}{m} \approx (\frac{\omega_Y}{\omega_Z})(\frac{u'(Y)}{u'(Z)})$, we have

$$\overline{\phi}(p, q) \approx \left[\left(\frac{\overline{\omega}_Z}{1-p-q}\right)\left(\frac{q}{\overline{\omega}_Y}\right)\right]C \quad \text{and} \quad \underline{\phi}(p, q) \approx \left[\left(\frac{\underline{\omega}_Z}{1-p-q}\right)\left(\frac{q}{\underline{\omega}_Y}\right)\right]C,$$

where $C = \frac{u'(Z)}{u'(Y)}$ is a constant in our experiment because we hold Y and Z fixed. If the decision weights are proportional to the probabilities, then the bracketed terms are identically unity. Thus, if $\overline{\phi}(p, q)$ or $\underline{\phi}(p, q)$ varies with p or q, the probability weighting function cannot be linear.

---

[16]A possible objection is that, in three-outcome lotteries, it is always the case that $p < 1-q$, which means we cannot rule out discontinuities at $\pi(1)$ using equalizing reductions for $\overline{X} > Y > Z$ and $Y > \underline{X} > Z$. However, we can easily solve that problem by also evaluating the change in k when X falls below Z. Although we do not perform such tests as part of our main analyses, we include them along with other robustness checks in Section 5.3.

More specifically, under CPT, we can rewrite these approximations as follows:

$$\overline{\phi}(p,q) \approx \left(\frac{\pi(1) - \pi(p+q)}{1-p-q} \right) \left(\frac{q}{\pi(p+q) - \pi(p)} \right) C \quad \text{and}$$

$$\underline{\phi}(p,q) \approx \left(\frac{\pi(1) - \pi(p+q)}{1-p-q} \right) \left(\frac{q}{\pi(q)} \right) C.$$

Suppose we observe two values of p, call them $p' > p''$, for which $\underline{\phi}(p',q) \neq \underline{\phi}(p'',q)$. Then we can conclude that $\pi(\cdot)$ is not linear throughout the interval $[1 - p' - q, 1]$. Alternatively, suppose we observe $\overline{\phi}(p',q) \neq \overline{\phi}(p'',q)$. Then we can conclude that $\pi(\cdot)$ is not linear throughout the interval $[\min\{1 - p' - q, p''\}, 1]$. Moreover, from the magnitude of the difference, we can obtain a non-parametric estimate of the degree to which the slope of the probability weighting function varies over an identifiable interval; see Section 4.2.2. Analogous statements hold for the probability q.

2.1. *Simulated Equalizing Reductions Under CPT Decisionmaking*

In this section, we examine the particular lotteries studied in our experiment and show that changes in payoff ranks yield large changes in probability weights under standard parameterizations of CPT. We also demonstrate that the percentage change in the equalizing reduction approximates the percentage change in the probability weights to a high degree of accuracy even when m represents a discrete payoff increment of nontrivial magnitude.

We focus on the parametric specification used in the original formulation of CPT (Tversky and Kahneman (1992)),[17] which posited a probability weighting function, $\pi(p) = p^\gamma/(p^\gamma + (1-p)^\gamma)^{1/\gamma}$, a reference point of $r = 0$, and a utility function $u(x) = x^\alpha$ for $x > r = 0$. The parameters identified by Tversky and Kahneman (1992) were $\gamma = 0.61$ and $\alpha = 0.88$.

Consider the lottery, L, with $\{\overline{X}, Y, Z\} = \{\$30, \$24, \$18\}$ and $\{p, q, 1-p-q\} = \{0.4, 0.3, 0.3\}$. Increase Y by $m = \$5$, from \$24 to \$29. For the parameters $\gamma = 0.61$ and $\alpha = 0.88$, the equalizing reduction is $\overline{k} = 1.67$.[18] Now consider the lottery L' with $\{\underline{X}, Y, Z\} = \{\$23, \$24, \$18\}$ and $\{p, q, 1-p-q\} = \{0.4, 0.3, 0.3\}$. For the same CPT parameters as above, the equalizing reduction for $m = 5$ is $\underline{k} = \$3.22$.[19] Thus, a standard parameterization of CPT implies a sharp discontinuity in equalizing reductions: moving from $\underline{X} < Y$ to $\overline{X} > Y + m$ cuts the equalizing reduction roughly in half. The log difference in equalizing reductions, $\Delta \log(k) = -0.66$, closely approximates the change in probability weight associated with the outcome Y, as $\log(\pi(0.7) - \pi(0.3)) - \log(\pi(0.3)) = -0.66$ as well. Indeed, the approximation remains quite close even when the utility function has much greater curvature. For example, with $\alpha = 0.5$, $\Delta \log(k) = -0.65$, and for $\alpha = 0.25$, $\Delta \log(k) = -0.64$.

In Table I, we provide additional simulations with the same values of X, Y, Z, and m as above, but using three different values of γ, 0.4, 0.61, 0.8, as well as three different probability vectors, $\{p, q, 1-p-q\} = \{0.6, 0.3, 0.1\}$, $\{0.4, 0.3, 0.3\}$, and $\{0.1, 0.3, 0.6\}$.[20]

---

[17]Tversky and Fox (1995) and Gonzalez and Wu (1999) employed a similar two-parameter $\pi(\cdot)$ function. See Prelec (1998) for alternative S-shaped specifications.

[18]Note that Y and Z are received with equal probability, so an expected-value decisionmaker would exhibit an equalizing reduction of $\overline{k} = \$5$.

[19]Once again, note that an expected-value decisionmaker would exhibit an equalizing reduction of $\underline{k} = \$5$.

[20]To demonstrate the dependence of discontinuities in equalizing reduction on the extent of probability weighting, we hold α fixed at 0.88 throughout.

TABLE I

CUMULATIVE PROSPECT THEORY SIMULATED EQUALIZING REDUCTIONS[a]

$\{p, q, 1-p-q\}$	$\gamma = 0.4$			$\gamma = 0.61$			$\gamma = 0.8$		
	$\underline{k}$	$\overline{k}$	$\Delta\log(\frac{wY}{wZ}) = \Delta\log(k)$ $\Delta\log(\frac{wY}{wZ}) = \log(\frac{\pi(p+q)-\pi(p)}{\pi(q)})$	$\underline{k}$	$\overline{k}$	$\Delta\log(\frac{wY}{wZ}) = \Delta\log(k)$ $\Delta\log(\frac{wY}{wZ}) = \log(\frac{\pi(p+q)-\pi(p)}{\pi(q)})$	$\underline{k}$	$\overline{k}$	$\Delta\log(\frac{wY}{wZ}) = \Delta\log(k)$ $\Delta\log(\frac{wY}{wZ}) = \log(\frac{\pi(p+q)-\pi(p)}{\pi(q)})$
$\{0.6, 0.3, 0.1\}$	1.97	1.33	−0.39 −0.39	5.17	3.88	−0.29 −0.29	9.21	7.84	−0.16 −0.17
$\{0.4, 0.3, 0.3\}$	1.61	0.53	−1.12 −1.12	3.22	1.67	−0.66 −0.66	4.29	3.13	−0.31 −0.32
$\{0.1, 0.3, 0.6\}$	1.45	0.40	−1.30 −1.30	2.39	1.39	−0.55 −0.55	2.60	2.08	−0.22 −0.22

[a] Dollar values for equalizing reductions in Z for increase in Y to $Y + m$. $\overline{k}$ calculated with $(\overline{X}, Y, Z) = (\$30, \$24, \$18\}$, $m = \$5$. $\underline{k}$ calculated with $(\underline{X}, Y, Z) = (\$23, \$24, \$18\}$, $m = \$5$. CPT calculations with $u(x) = x^{\alpha}$, $\alpha = 0.88$; and $\pi(p) = p^{\gamma}/(p^{\gamma} + (1-p)^{\gamma})^{1/\gamma}$ with γ varying by column.

For the CPT parameter values of Tversky and Kahneman (1992), the probability weight on payoff Y changes by 29 to 66 percent as X passes through Y. Even with more modest curvature of the probability weighting function ($\gamma = 0.8$), the change in probability weight remains sizable. Critically, in all cases, the percentage change in the equalizing reduction approximates the percentage change in the probability weight associated with payoff Y to a high degree of accuracy.

2.2. *Reference Point Formulation and Alternative Models of Reference Dependence*

Throughout the previous discussion, we assumed that the reference point is fixed and below all potential payoffs. While this assumption is a reasonable starting point, one naturally wonders whether our conclusions are robust with respect to other possibilities.

First, consider the possibility that the reference point is exogenous but falls either (i) above all payoffs, or (ii) between the lottery's payoffs, which it segregates into gains and losses. Case (ii) may seem particularly concerning because CPT applies probability weighting to gains and losses separately. Notice, however, that Proposition 1 subsumes these possibilities because it is proved for a specification with general decision weights. Because CPT still implies that the weights change when the value of X passes through Y, precisely the same implications follow.[21] Notably, additional discontinuities in equalizing reductions emerge (for similar reasons) at the point where the outcomes Z and Y pass the reference point (see Appendix C.1 of the Supplemental Material, Bernheim and Sprenger (2020)). Thus, the equalizing reduction approach not only offers a novel test of rank dependence, but could also be used to test the hypothesis that gains and losses relative to an exogenous reference point are weighted differently.

Next, consider the possibility that the reference point depends on the lottery's payoffs, as in Bell (1985), Loomes and Sugden (1986), and others. We will use r to denote a generic reference point and $R(X, Y, Z)$ to denote the reference point for a lottery that yields payoffs $\{X, Y, Z\}$ with probabilities $\{p, q, 1 - p - q\}$. (The reference point may also depend on the probabilities, but we hold them constant, and consequently suppress those arguments for notational simplicity.) Here we will focus on cases in which the reference point r coincides with neither Y nor Z.[22]

---

[21]An additional complication arises for non-infinitesimal values of m, in that an increase from Y to $Y + m$ could cross the reference point, or cause $Z - k$ to cross the reference point. As shown in Appendix C.1 of the Supplemental Material (Bernheim and Sprenger (2020)), the implications of rank dependence are, nevertheless, unchanged.

[22]As the literature has emphasized, there are multiple ways to endogenize the reference point. The proof of Proposition 2 assumes that reference points are "choice-acclimating" (Koszegi and Rabin (2006, 2007)), in the sense that the decisionmaker evaluates each option according to the reference point she would have if she selected that option. For example, according to the theory of disappointment aversion as articulated by Bell (1985) and Loomes and Sugden (1986), the individual chooses $\{X, Y, Z\}$ over $\{X, Y + m, Z - k\}$ when $U(X, Y, Z|R(X, Y, Z)) > U(X, Y + m, Z - k|R(X, Y + m, Z - k))$. An alternative approach is to assume that the reference point resolves based on "first focus" (Koszegi and Rabin (2006)), in the sense that the decisionmaker evaluates all the options in a task according to a single reference point that depends on the task's parameters. In our setting, a natural implementation of this approach would allow the reference point to depend on the static task elements, X, Y, Z, and m, but would treat it as fixed within each equalizing reduction elicitation task. In that case, the individual chooses $\{X, Y, Z\}$ over $\{X, Y + m, Z - k\}$ when $U(X, Y, Z|R(X, Y, Z, m)) > U(X, Y + m, Z - k|R(X, Y, Z, m))$. Proposition 2 also holds for this first-focus approach. Yet another approach involves the analysis of "personal equilibria" (Koszegi and Rabin (2006, 2007)), in which the individual evaluates all options relative to the reference point associated with a potential selection: there is a personal equilibrium of selecting $\{X, Y, Z\}$ over $\{X, Y + m, Z - k\}$ when $U(X, Y, Z|R(X, Y, Z)) \geq U(X, Y + m, Z - k|R(X, Y, Z))$, and of selecting $(X, Y + m, Z - k)$ over (X, Y, Z)

PROPOSITION 2: *Suppose decision weights are fixed for a given payoff ranking, that $u(x, r)$ is continuously differentiable in neighborhoods of $(Y, R(Y, Y, Z))$ and $(Z, R(Y, Y, Z))$, and that R is continuously differentiable in a neighborhood of (Y, Y, Z). Consider any sequence $(\overline{X}^n, \underline{X}^n) \to (Y, Y)$ such that $\overline{X}^n > Y > \underline{X}^n > Z$. Then*

$$\lim_{n \to \infty} \lim_{m \to 0} \Delta \log(k) = \Delta \log\left(\frac{w_Y}{w_Z}\right).$$

PROOF: See the Appendix.   *Q.E.D.*

Thus, even with an endogenous reference point, the discontinuity in the X-k schedule at $X = Y$ still measures the percentage change in the relative decision weights on Y and Z. That said, if the reference point depends on the payoff X, then the X-k schedule may no longer be flat within intervals with fixed decision weights, a possibility that could in principle make the size of any discontinuity more difficult to measure. However, as we will see, there is no indication that this potential issue materializes in practice. On the contrary, the flatness of the empirical X-k schedule eliminates any complications arising from the potential endogeneity of the reference point. An additional discontinuity in the equalizing reduction should emerge where X crosses an endogenous reference point, even in the absence of probability weighting (see Appendix C.2 of the Supplemental Material, Bernheim and Sprenger (2020)). Intuitively, with a "choice-acclimating" (Koszegi and Rabin (2006, 2007)) reference point (as in disappointment aversion), variations in m and k affect the utility derived from X by changing its distance from the reference point. The size of this effect differs discretely according to whether X is above the reference point (and hence a gain), or below the reference point (and hence a loss).[23] Thus, our approach offers a novel test of theories, such as disappointment aversion, for which $X <$ (resp. $>$) Y and Z implies that X is below (resp. above) the reference point.

A final possibility worth considering is that subjects have reference point distributions, as in Koszegi and Rabin (2006, 2007). We show in Appendix C.3 of the Supplemental Material (Bernheim and Sprenger (2020)) that this preference formulation also yields discontinuities in the equalizing reduction as X passes through Y and Z, even without probability weighting.[24] Furthermore, when we apply a calibrated model to our experimental tasks, we find that the implied discontinuities are substantial, and that their signs are opposite those implied by the CPT calibrations. Thus, our approach also offers a novel and discerning test of the Koszegi and Rabin (2006, 2007) framework.

2.3. *Relation to Existing Evidence on Rank Dependence*

As mentioned in the Introduction, our work is most closely related to a handful of studies that aim to test the axiomatic foundations of rank-dependent models (Wu (1994),

---

when $U(X, Y, Z | R(X, Y + m, Z - k)) \leq U(X, Y + m, Z - k | R(X, Y + m, Z - k))$. This approach can give rise to multiple personal equilibria, and hence indeterminacy, absent some refinement (such as "preferred personal equilibria," as in Koszegi and Rabin (2006)).

[23]Note that as the reference point becomes decreasingly sensitive to the payoffs, the size of this discontinuity shrinks (to zero in the case of a fixed reference point). The additional discontinuity is also eliminated in cases where the reference point is sensitive to variation in payoffs across tasks, but is fixed within a task, as in the "first-focus" approach discussed above.

[24]Masatlioglu and Raymond (2016) and Barseghyan et al. (2015) noted a tight connection between rank-dependent theories with specific functional forms associated with pessimism and the Koszegi and Rabin (2006, 2007) model. This work anticipates the results outlined in Appendix C.3 of the Supplemental Material (Bernheim and Sprenger (2020)).

Wakker, Erev, and Weber (1994), Fennema and Wakker (1996), Weber and Kirsner (1997), Birnbaum (2008)). A defining feature of those models is that they assume the independence axiom holds on a limited domain. In particular, we say that two lotteries are *comonotonic* if they induce the same payoff ranking over states of nature. Under EU and CPT, if two comonotonic lotteries yield the same payoff, x_j, in some state j, then a change in x_j that leaves the ranking intact should have no effect on preferences between the lotteries. This property reflects an axiom known as *comonotonic independence* (CI) (Schmeidler (1989)), which EU and CPT both satisfy. Naturally, one can also ask whether preferences between the lotteries are invariant with respect to changes in x_j that alter the payoff ranking. This type of invariance follows from a property known as *non-comonotonic independence* (NCI), which EU satisfies but CPT does not. Thus, evidence validating both CI and NCI would point to EU, and evidence favoring CI while challenging NCI would point to CPT. The experimental tests of CI and NCI conducted generally implement changes to a common state payoff either preserving or altering the ranking of outcomes as described above.[25]

When interpreting laboratory evidence concerning conformance with choice axioms, it is important to allow for the possibility that observed choices are somewhat noisy. As a result, even if a theory captures the essence of decisionmaking, one would expect to observe violations of the axioms that characterize it. What then can one conclude from the frequency of violations? Existing tests of rank dependence involve comparisons between the prevalence of violations for different axioms. For instance, Wakker, Erev, and Weber (1994) attributed the differential between the frequency of NCI violations and CI violations to rank dependence. Because they found little difference in these frequencies, they concluded that rank dependence is not supported.

These types of frequency comparisons raise two difficulties, both stemming from the fact that the results are difficult to interpret without a parametric model of noisy choice. First, the premise of the approach—that violation frequencies are necessarily higher for invalid axioms—is flawed. For reasonable models of noisy choice, noise-induced violations of choice axioms are more likely to occur when the parameters of the tasks place the decisionmaker closer to the point of indifference. Existing approaches provide no way to ensure that the "distance to indifference" is held constant when comparing CI and NCI violations. It is therefore easy to construct examples in which a "noisy" CPT decisionmaker violates CI just as frequently, or even more frequently, than NCI. Second, even if one could control for "distance to indifference," this approach offers no basis for judging whether a given discrepancy between the frequencies of CI and NCI violations is large or small relative to the implications of a reasonably parameterized CPT model. For any given degree of rank dependence, one can construct simple examples (with constant "distance to indifference") in which the differential between violation frequencies falls anywhere between zero and unity. See Appendix B of the Supplemental Material (Bernheim and Sprenger (2020)) for a description of the aforementioned examples.

Because rank dependence is characterized by the restriction of the independence axiom to comonotonic lotteries (Wakker, Erev, and Weber (1994)), any valid test of the hypothesis is necessarily related to the existing studies, and ours is no exception. However, our use of equalizing reductions has no counterpart in the existing literature. Instead of counting violations of CI and NCI, we measure equalizing reductions separately for each lottery, and then compare them across lotteries. Certainly, the constancy of the

---

[25]Note that because both PT and EU feature rank-independent treatment of probabilities, their predictions in these specific CI and NCI tests coincide.

equalizing reduction over values of X that preserve the payoff ordering is an implication of CI, and the discontinuity in the equalizing reduction at the point where $X = Y$ or $X = Z$ reflects a failure of NCI. However, 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. Critically, in addition, magnitudes are interpretable in our framework: as we have shown, the percentage change in the equalizing reduction resulting from a rank-changing variation in X provides a nonparametric estimate of the percentage change in decision weights. Finally, the interpretation of the change in the equalizing reduction remains the same regardless of whether choices are noisy; only the precision of the estimates is affected.

3. DESIGN OF THE MAIN EXPERIMENT

Our experimental design follows closely the theoretical discussion of Section 2. Conditional on various probability vectors, $\{p, q, 1 - p - q\}$, we test for differences in equalizing reductions between lotteries with ranks $\overline{X} > Y > Z$ and those with ranks $Y > \underline{X} > Z$, and examine changes in equalizing reductions associated with variations in the probability p. Subjects also complete a battery of certainty equivalent tasks involving binary lotteries; these tasks are commonly used to derive risk-preference parameters within the CPT framework. This strategy allows us to verify that our subjects exhibit the choice patterns commonly associated with CPT preferences. We divide our discussion of design into three subsections. First, we describe the elicitation of equalizing reductions; second, we detail the conventional elicitation of certainty equivalents; third, we discuss other design details including task orders and payment procedures.

3.1. *Elicitation of Equalizing Reductions*

We elicited equalizing reductions using the method of price lists. In each task, subjects made a series of decisions between 'Option A' and 'Option B', both three-outcome lotteries. Option A was fixed throughout the task as either a lottery with $\overline{X} > Y > Z$ or a lottery with $Y > \underline{X} > Z$. Option B was constructed by adding $5 to Y and reducing Z by k. The value of k varied throughout the task. The point at which an individual switched from choosing Option A to choosing Option B places tight bounds on the equalizing reduction, either $\overline{k}$ or $\underline{k}$. Panels A and B of Figure 1 provide two tasks eliciting $\overline{k}$ and $\underline{k}$. The Supplemental Material (Bernheim and Sprenger (2020)) provides the full instructions given to subjects along with all tasks.

As in the simulations of Section 2, our design fixes $Y = \$24$, $Z = \$18$, and $m = \$5$. We use three values of $\overline{X} > Y + m$, $\{\$34, \$32, \$30\}$, and three values of $\underline{X} < Y$, $\{\$23, \$21, \$19\}$. Our objective in using multiple values of X on either side of Y is to allow for the possibility that the change in the equalizing reduction might be gradual in a neighborhood of the change-in-ranks threshold rather than sudden, perhaps because payoff rankings are non-salient for nearby outcomes. Additionally, as explained in Section 2.2, we can also examine the sensitivity of equalizing reductions to the value of X within the two regimes to draw inferences about the endogeneity of reference points. We use one additional value, $X = \$25$, to check robustness; see the discussion in Appendix F of the Supplemental Material (Bernheim and Sprenger (2020)).[26] This set of values allows us

---

[26]Note that for $X = \$25$, adding $5 to Y induces a change of ranks. These tasks allow us to investigate the possibility that explicit rank changes influence choice. See Appendix F of the Supplemental Material (Bernheim and Sprenger (2020)) for further detail.

Panel A: $\overline{X} > Y > Z$ over $\overline{k}$

	Option A			or	Option B		
	40 in 100 Chance	30 in 100 Chance	30 in 100 Chance		40 in 100 Chance	30 in 100 Chance	30 in 100 Chance
1)	$30	$24	$18	or	$30	$29	$18.00
2)	$30	$24	$18	or	$30	$29	$17.75
3)	$30	$24	$18	or	$30	$29	$17.50
4)	$30	$24	$18	or	$30	$29	$17.00
5)	$30	$24	$18	or	$30	$29	$16.75
6)	$30	$24	$18	or	$30	$29	$16.50
7)	$30	$24	$18	or	$30	$29	$16.25
8)	$30	$24	$18	or	$30	$29	$16.00
9)	$30	$24	$18	or	$30	$29	$15.75
10)	$30	$24	$18	or	$30	$29	$15.50
11)	$30	$24	$18	or	$30	$29	$15.25
12)	$30	$24	$18	or	$30	$29	$15.00
13)	$30	$24	$18	or	$30	$29	$14.50
14)	$30	$24	$18	or	$30	$29	$14.00
15)	$30	$24	$18	or	$30	$29	$13.50
16)	$30	$24	$18	or	$30	$29	$13.00
17)	$30	$24	$18	or	$30	$29	$12.50
18)	$30	$24	$18	or	$30	$29	$12.00
19)	$30	$24	$18	or	$30	$29	$11.50
20)	$30	$24	$18	or	$30	$29	$11.00
21)	$30	$24	$18	or	$30	$29	$10.50
22)	$30	$24	$18	or	$30	$29	$10.00
23)	$30	$24	$18	or	$30	$29	$9.50
24)	$30	$24	$18	or	$30	$29	$9.00
25)	$30	$24	$18	or	$30	$29	$8.50
26)	$30	$24	$18	or	$30	$29	$8.00
27)	$30	$24	$18	or	$30	$29	$7.50
28)	$30	$24	$18	or	$30	$29	$7.00
29)	$30	$24	$18	or	$30	$29	$6.50
30)	$30	$24	$18	or	$30	$29	$6.00
31)	$30	$24	$18	or	$30	$29	$5.50
32)	$30	$24	$18	or	$30	$29	$5.00
33)	$30	$24	$18	or	$30	$29	$4.50
34)	$30	$24	$18	or	$30	$29	$4.00
35)	$30	$24	$18	or	$30	$29	$3.50
36)	$30	$24	$18	or	$30	$29	$3.00
37)	$30	$24	$18	or	$30	$29	$2.50
38)	$30	$24	$18	or	$30	$29	$2.00

Panel B: $Y > \underline{X} > Z$ over $\underline{k}$

	Option A			or	Option B		
	40 in 100 Chance	30 in 100 Chance	30 in 100 Chance		40 in 100 Chance	30 in 100 Chance	30 in 100 Chance
1)	$23	$24	$18	or	$23	$29	$18.00
2)	$23	$24	$18	or	$23	$29	$17.75
3)	$23	$24	$18	or	$23	$29	$17.50
4)	$23	$24	$18	or	$23	$29	$17.00
5)	$23	$24	$18	or	$23	$29	$16.75
6)	$23	$24	$18	or	$23	$29	$16.50
7)	$23	$24	$18	or	$23	$29	$16.25
8)	$23	$24	$18	or	$23	$29	$16.00
9)	$23	$24	$18	or	$23	$29	$15.75
10)	$23	$24	$18	or	$23	$29	$15.50
11)	$23	$24	$18	or	$23	$29	$15.25
12)	$23	$24	$18	or	$23	$29	$15.00
13)	$23	$24	$18	or	$23	$29	$14.50
14)	$23	$24	$18	or	$23	$29	$14.00
15)	$23	$24	$18	or	$23	$29	$13.50
16)	$23	$24	$18	or	$23	$29	$13.00
17)	$23	$24	$18	or	$23	$29	$12.50
18)	$23	$24	$18	or	$23	$29	$12.00
19)	$23	$24	$18	or	$23	$29	$11.50
20)	$23	$24	$18	or	$23	$29	$11.00
21)	$23	$24	$18	or	$23	$29	$10.50
22)	$23	$24	$18	or	$23	$29	$10.00
23)	$23	$24	$18	or	$23	$29	$9.50
24)	$23	$24	$18	or	$23	$29	$9.00
25)	$23	$24	$18	or	$23	$29	$8.50
26)	$23	$24	$18	or	$23	$29	$8.00
27)	$23	$24	$18	or	$23	$29	$7.50
28)	$23	$24	$18	or	$23	$29	$7.00
29)	$23	$24	$18	or	$23	$29	$6.50
30)	$23	$24	$18	or	$23	$29	$6.00
31)	$23	$24	$18	or	$23	$29	$5.50
32)	$23	$24	$18	or	$23	$29	$5.00
33)	$23	$24	$18	or	$23	$29	$4.50
34)	$23	$24	$18	or	$23	$29	$4.00
35)	$23	$24	$18	or	$23	$29	$3.50
36)	$23	$24	$18	or	$23	$29	$3.00
37)	$23	$24	$18	or	$23	$29	$2.50
38)	$23	$24	$18	or	$23	$29	$2.00

FIGURE 1.—Equalizing reduction tasks.

to investigate both rank dependence and the prediction that equalizing reductions are constant within ranks.

As in the simulations of Section 2, we examine three probability vectors, $\{p, q, 1 - p - q\} = \{0.6, 0.3, 0.1\}$, $\{0.4, 0.3, 0.3\}$, and $\{0.1, 0.3, 0.6\}$. The cumulative probability of receiving at least Y ranges from 0.3 to $0.6 + 0.3 = 0.9$, which provides broad scope for detecting the predicted discontinuities. Note that the design varies the relative probabilities of Y and Z from $0.3/0.1 = 3$ to $0.3/0.6 = 1/2$, generating a wide range of equalizing reductions.[27]

With seven values of X and three probability vectors, we have 21 equalizing reduction tasks in total. We organize these tasks into seven blocks, each of which presents the three probability vectors for a single value of X. Hence, the tasks within each block are differentiated by the probability vector, $\{p, q, 1 - p - q\}$. We distributed task blocks to subjects one at a time, and collected responses before moving on to the next block. This feature of our design was intended to limit any tendency to respond mechanically with the same answer as X varies, a possibility that could artificially obscure discontinuities. As we will see, subjects responded strongly to variations in probability within each block before moving on to the next value of X.

3.2. *Prospect Theory Elicitation Tasks*

We also elicit our subjects' risk-preference parameters using the same experimental techniques employed by Tversky and Kahneman (1992). The approach employs seven tasks, each of which elicits the certainty equivalents for a two-outcome lottery, $(p, \$25; 1 - p, 0)$, $p \in \{0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95\}$. We grouped these tasks in a single block. Figure 2 illustrates one of these tasks, and the Supplemental Material (Bernheim and Sprenger (2020)) describes the full set. Although these tasks provide no information concerning rank dependence in probability weighting, they allow us to determine whether our sample exhibits representative risk preferences.

3.3. *Design Details*

One hundred fifty three subjects were recruited from the Stanford Economics Research Laboratory subject pool in September, October, and November of 2014. A total of 20 sessions were conducted with the number of subjects varying between two and sixteen. We varied the order of the six main equalizing reduction blocks systematically across sessions. Subjects completed three equalizing reduction blocks, then the CPT elicitation block, then three more equalizing reduction blocks. The $X = \$25$ equalizing reduction tasks were always presented last. Table II lists all sessions, dates, numbers of subjects, and block orders.

To induce truthful revelation of equalizing reductions and certainty equivalents, we incentivized subjects by paying them based on one randomly selected question in one randomly selected task. On average, subjects earned $26.87. This random-lottery incentive mechanism is widely used in experimental economics, but note that it transforms the experiment into a single compound lottery. The literature on choice under risk, dating to Holt (1986) and Karni and Safra (1987), suggests that random mechanisms need

---

[27]For example, an expected-value decisionmaker would exhibit values of $\bar{k}$, $\underline{k}$ between $2.50 and $15 across these tasks.

	Option A			or	Option B	
	50 in 100 Chance	50 in 100 Chance			100 in 100 Chance	
1)	$25	$0	☐	or	$25.00	☐
2)	$25	$0	☐	or	$24.00	☐
3)	$25	$0	☐	or	$23.00	☐
4)	$25	$0	☐	or	$22.00	☐
5)	$25	$0	☐	or	$21.00	☐
6)	$25	$0	☐	or	$20.00	☐
7)	$25	$0	☐	or	$19.00	☐
8)	$25	$0	☐	or	$18.00	☐
9)	$25	$0	☐	or	$17.00	☐
10)	$25	$0	☐	or	$16.00	☐
11)	$25	$0	☐	or	$15.00	☐
12)	$25	$0	☐	or	$14.00	☐
13)	$25	$0	☐	or	$13.00	☐
14)	$25	$0	☐	or	$12.00	☐
15)	$25	$0	☐	or	$11.00	☐
16)	$25	$0	☐	or	$10.00	☐
17)	$25	$0	☐	or	$9.00	☐
18)	$25	$0	☐	or	$8.00	☐
19)	$25	$0	☐	or	$7.00	☐
20)	$25	$0	☐	or	$6.00	☐
21)	$25	$0	☐	or	$5.00	☐
22)	$25	$0	☐	or	$4.00	☐
23)	$25	$0	☐	or	$3.00	☐
24)	$25	$0	☐	or	$2.00	☐
25)	$25	$0	☐	or	$1.00	☐
26)	$25	$0	☐	or	$0.00	☐

FIGURE 2.—Prospect Theory Elicitation Task.

not be incentive compatible in such contexts if either the Independence or Reduction of Compound Lotteries axioms are violated. As CPT violates independence, this limitation is a potential concern. Importantly, however, Starmer and Sugden (1991) and Cubitt, Starmer, and Sugden (1998) demonstrated that this mechanism can be used even when subjects deviate from expected utility because, in practice, they narrowly frame each lottery, making each decision in isolation. Whether their findings apply to our setting is, of course, an empirical question. Our Prospect Theory elicitation tasks are especially important in this regard because they allow us to assess the validity of the method we use. If isolation fails in this context, then our subjects would not exhibit standard patterns of probability weighting in binary tasks. 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.

Ten of 153 (6.5%) subjects exhibited at least one instance of multiple switching within a single experimental task. This figure compares favorably to other experiments employing

TABLE II

EXPERIMENTAL SESSIONS[a]

Number	Date	Order	# Obs
1	09/24/14	{34, 32, 30, CE, 23, 21, 19, 25}	16
2	09/24/14	{34, 21, 30, CE, 23, 32, 19, 25}	11
3	09/30/14	{23, 32, 19, CE, 34, 21, 30, 25}	9
4	09/30/14	{19, 32, 23, CE, 30, 21, 34, 25}	12
5	10/01/14	{30, 21, 34, CE, 19, 32, 23, 25}	12
6	10/02/14	{21, 30, 34, CE, 32, 19, 23, 25}	14
7	10/07/14	{32, 19, 23, CE, 21, 30, 34, 25}	10
8	10/07/14	{23, 19, 32, CE, 34, 30, 21, 25}	5
9	10/08/14	{34, 30, 21, CE, 23, 19, 32, 25}	13
10	10/09/14	{30, 34, 21, CE, 19, 23, 32, 25}	5
11	10/14/14	{19, 23, 32, CE, 30, 34, 21, 25}	4
12	10/16/14	{32, 23, 19, CE, 21, 34, 30, 25}	6
13	10/26/14	{21, 34, 30, CE, 32, 23, 19, 25}	7
14	10/28/14	{21, 34, 30, CE, 32, 23, 19, 25}	3
15	10/29/14	{32, 23, 19, CE, 21, 34, 30, 25}	2
16	11/05/14	{23, 19, 32, CE, 34, 30, 21, 25}	2
17	11/07/14	{19, 23, 32, CE, 30, 34, 21, 25}	6
18	11/10/14	{23, 19, 32, CE, 34, 30, 21, 25}	6
19	11/14/14	{30, 34, 21, CE, 19, 23, 32, 25}	6
20	11/18/14	{32, 23, 19, CE, 21, 34, 30, 25}	4
Total			153

[a]Session number, date, order, and number of observations. Order of tasks refers to the value of $\overline{X}/\underline{X}$ in each task block. CE corresponds to the block of tasks with certainty equivalent questions.

price lists.[28] Because multiple switch points are difficult to rationalize and may indicate subject confusion, researchers often exclude such observations or mechanically enforce single switch points.[29] We begin by excluding all subjects exhibiting multiple switch points in any task, leaving a sample of 143 subjects. In Appendix F of the Supplemental Material (Bernheim and Sprenger (2020)), we include the ten subjects with multiple switch points, taking each subject's first switch point as their relevant choice, and demonstrate that the results are qualitatively unchanged.

4. MAIN RESULTS

In presenting our main results, we begin with the conventional Prospect Theory elicitation tasks, demonstrating that our subjects exhibit the classic manifestations of probability weighting found in previous studies. Then we explore equalizing reductions. As a first step, we show that there is no relationship between payoff ranks and equalizing reductions. Our nonparametric estimates of the change in the relative probability weights resulting from a change in ranks cluster around zero, and in all cases we can reject even modest changes with 95% confidence. It follows either that probability weighting is not rank-dependent, or that $\pi(\cdot)$ is nearly linear. As a second step, we examine responses in

---

[28]Around 10 percent of subjects feature multiple switch points in similar price-list experiments (Holt and Laury (2002), Meier and Sprenger (2010)), and as many as 50 percent in some cases (Jacobson and Petrie (2009)).

[29]See Harrison, Lau, Rutstrom, and Williams (2005) for discussion.

the same tasks to variations in probabilities, holding ranks fixed. A nonparametric analysis of these responses reveals that $\pi(\cdot)$ is highly nonlinear over a probability range for which the absence of responses to changes in ranks rules out meaningful nonlinearities under the maintained hypothesis of rank dependence. Adopting standard functional assumptions, we then show that the estimated degree of curvature differs sharply depending on whether one draws inferences from responses to variations in payoff ranks or variations in probabilities. Using models estimated based on responses to variations in probabilities with fixed ranks, we predict the degree to which equalizing reductions should change in response to rank reversals under the assumption of rank dependence. We also perform these calculations using data from conventional CPT elicitation tasks. In all cases, the predicted changes are an order of magnitude greater than the observed changes, and the confidence intervals are non-overlapping. The assumption of nonlinear rank-independent probability weighting reconciles the otherwise conflicting results from our first and second steps. Indeed, parametric estimates show that the PT formulation of probability weighting accounts for the data on equalizing reductions more successfully than the CPT or EU (linear) formulations.

4.1. *Certainty Equivalents of Binary Lotteries*

As in the original experiments of Tversky and Kahneman (1992), we administered seven certainty equivalent tasks involving lotteries over payoffs of $25 and $0, with the governing probability $p \in \{0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95\}$. Panel A of Figure 3 summarizes these data. To capture average behavior, we first estimated an interval regression (Stewart (1983)) describing the certainty equivalent, C, as a function of indicators for the experimental probabilities, p.[30] Panel A thus presents the estimated mean certainty equivalent for each value of p along with its 95% confidence interval.[31] Following Tversky and Kahneman (1992), we present the data in Panel A relative to a benchmark of risk neutrality so that the curve would directly reveal the probability weighting function, $\pi(\cdot)$, if the utility function were linear.

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 $u(C) = \pi(p) \cdot u(25)$. We follow Tversky and Kahneman (1992) by assuming power utility, $u(x) = x^\alpha$, and a weighting function $\pi(p) = p^\gamma/(p^\gamma + (1-p)^\gamma)^{1/\gamma}$. We then estimate the parameters γ and α by minimizing the sum of squared residuals for the nonlinear regression equation

$$C = \left[p^\gamma/\left(p^\gamma + (1-p)^\gamma\right)^{1/\gamma} \times 25^\alpha \right]^{\frac{1}{\alpha}} + \epsilon, \tag{5}$$

where C is the midpoint of the certainty equivalent interval defined by experimental choice.

When conducting this analysis for our aggregate data with standard errors clustered on the subject level, we obtain $\alpha = 0.941$ (s.e. $= 0.019$) and $\gamma = 0.715$ (0.015). The benchmark model of linear weighting, $\gamma = 1$, is rejected at all conventional confidence levels

---

[30]Virtually identical results are obtained when using OLS and the midpoint of the interval.

[31]Standard errors are estimated clustered at the individual level. Appendix Table AII, column (1) of the Supplemental Material (Bernheim and Sprenger (2020)) provides corresponding estimates. Column (2) provides estimates of risk premia, demonstrating statistically significant risk tolerance at low probabilities and statistically significant risk aversion at high probabilities.
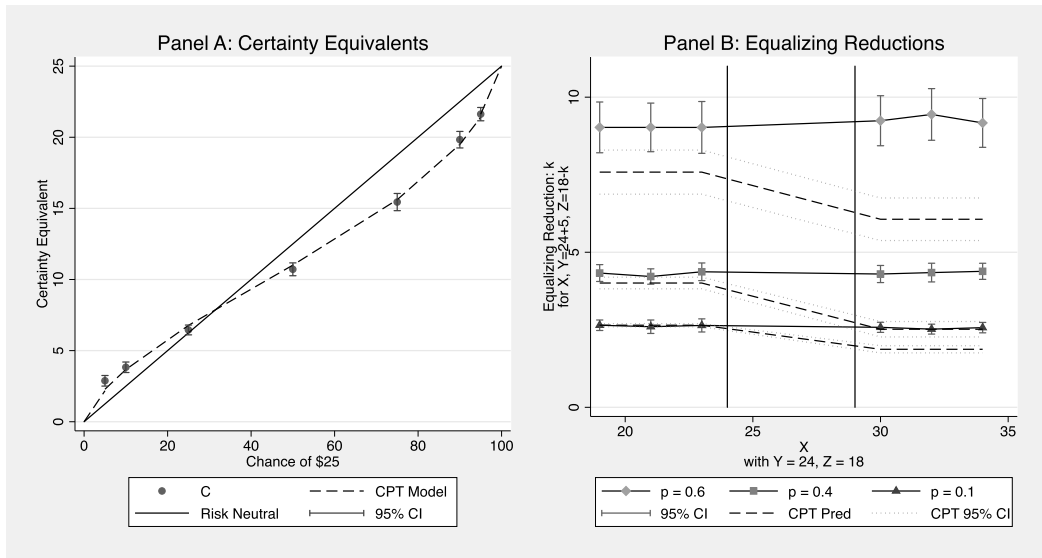
FIGURE 3.—Certainty equivalents and equalizing reductions. *Notes*: Panel A: Mean behavior for C estimated from interval regression (Stewart (1983)) of experimental response on indicators for probability vectors. Standard errors clustered at individual level to provide 95% confidence interval. Appendix Table AII, column (1) of the Supplemental Material (Bernheim and Sprenger (2020)) provides corresponding estimates. Dashed line corresponds to predicted CPT behavior with $\alpha = 0.941$ (s.e. $= 0.019$) and $\gamma = 0.715$ (0.015); standard errors clustered at individual level. Panel B: Mean behavior for k estimated from interval regression of experimental response on indicators for probability vectors interacted with indicators for value of X. Standard errors clustered at individual level to provide 95% confidence interval. Appendix Table AIV of the Supplemental Material (Bernheim and Sprenger (2020)) provides corresponding estimates. Dashed line corresponds to predicted values of equation (3) for CPT decisionmaker with risk preference parameters $\alpha = 0.941$ (s.e. $= 0.019$) and $\gamma = 0.715$ (0.015). Standard errors clustered at individual level and calculated using the delta method to provide 95% confidence interval.

($F_{1,142} = 341.2$, $p < 0.01$). The value of the probability weighting parameter is reasonably close to the estimate of Tversky and Kahneman (1992) ($\gamma = 0.61$), and coincides with the findings of Wu and Gonzalez (1996), who estimated $\gamma = 0.71$. The dashed line in Figure 3, Panel A, shows the in-sample model fit, which closely matches actual behavior.[32]

The main take-away from this analysis is that our subjects exhibit typical risk preferences, and do so within the context of the experiment's overall structure. Accordingly, there is no reason to suspect that we have an unrepresentative subject pool, or that our subjects fail to treat each task as an isolated decision problem.

4.2. *Average Equalizing Reductions*

Panel B of Figure 3 presents equalizing reductions, $\overline{k}$ and $\underline{k}$, for each value of $\overline{X}$ and $\underline{X}$. We exhibit separate results for each of the three probability vectors. To determine average behavior, we performed an interval regression describing the equalizing reduction, $\underline{k}$ or

---

[32]The correlation coefficient for predicted and actual certainty equivalents is 0.93, and a regression of the true certainty equivalent on the model's prediction yields a slope coefficient of 0.998 (clustered s.e. $= 0.020$), a constant of 0.102 (0.214), and an R-squared value of 0.86. The null hypothesis that the constant is 0 and the predicted value's true coefficient is 1 is not rejected ($F_{2,142} = 0.17$, $p = 0.84$).

$\overline{k}$, as a function of indicators for the probability vectors interacted with indicators for the value of $\overline{X}$ or $\underline{X}$.[33] The figure exhibits the estimated mean equalizing reductions along with 95% confidence intervals.[34] Vertical lines at $Y = \$24$ and $Y + m = \$29$ partition the figure into three regions, one showing $\underline{k}$, another showing $\overline{k}$, and a transitional region. Corresponding statistics appear in Table III, Panel A.

As explained in Section 2, our analysis involves two steps. In the first step, we use the equalizing reductions to measure changes in the relative decision weights, $\frac{\omega_Y}{\omega_Z}$, associated with reversals in the ranks of X and Y. Our main finding is apparent in Figure 3, Panel B, and Table III, Panel A. For each of our three probability vectors, our estimates of $\underline{k}$ and $\overline{k}$ are virtually identical, indicating either that probability weighting is not rank-dependent, or that $\pi(\cdot)$ is nearly linear. Upon reaching this conclusion, we turn to the second step, in which we draw inferences about the shape of the probability weighting function from responses *in the same tasks* to variations in probabilities, holding ranks fixed. Once again, our main finding is apparent in Table III, Panel A: both ϕ and $\overline{\phi}$ vary substantially across the three probability vectors, which indicates that $\pi(\cdot)$ is highly nonlinear. Thus, to rationalize the pattern of equalizing reductions, one must posit rank-independent probability weighting.

4.2.1. *The Effect of Payoff Rank on Decision Weight (Step 1)*

The equalizing reductions, $\underline{k}$ and $\overline{k}$, are strikingly similar for each fixed probability vector. For $\{p, q, 1 - p - q\} = \{0.6, 0.3, 0.1\}$, the mean value of $\underline{k}$ is 9.02 (clustered s.e. = 0.39), while the mean value of $\overline{k}$ is 9.28 (0.38). The difference, $\overline{k} - \underline{k} = 0.26$ (0.17), is not statistically different from zero $\chi^2(1) = 2.31$, $p = 0.13$. In the last column of Panel A, we compute the change $\Delta \log(k)$ resulting from the change in ranks, and present it as an estimate of the change in the log of the relative probability weights, $\Delta \widehat{\log(\frac{w_Y}{w_Z})}$. For the first probability vector, the point estimate is $+0.03$ (0.02). Consequently, our nonparametric estimates of the equalizing reduction rule out a decline in w_Y/w_Z greater than 1%. We obtain similar results for the other probability vectors.[35]

In the context of the CPT model, the log changes in equalizing reductions reported in Table III, Panel A should closely approximate the change in probability weight assigned to outcome Y as ranks change. Viewed in this light, the equalizing reductions reported in Table III reflect a striking absence of rank dependence. Based on our nonparametric

---

[33]Virtually identical results are obtained when using OLS and the midpoint of the interval.

[34]Standard errors are estimated clustered at the individual level. Appendix Table AIV of the Supplemental Material (Bernheim and Sprenger (2020)) provides corresponding estimates.

[35]For $\{p, q, 1 - p - q\} = \{0.4, 0.3, 0.3\}$, the mean value of $\underline{k}$ is 4.31 (0.12), while the mean value of $\overline{k}$ is 4.34 (0.12). The difference, $\overline{k} - \underline{k} = 0.04$ (0.09), is not statistically different from zero $\chi^2(1) = 0.18$, $p = 0.67$. The 95% confidence interval for $\Delta \widehat{\log(\frac{w_Y}{w_Z})}$ is $[-0.03, 0.05]$, indicating that we can reject a decline in the ratio w_Y/w_Z more extreme than 3%. For $\{p, q, 1 - p - q\} = \{0.1, 0.3, 0.6\}$, the mean value of $\underline{k}$ is 2.63 (0.08), while the mean value of $\overline{k}$ is 2.56 (0.07). The difference, $\overline{k} - \underline{k} = -0.07$ (0.06), is not statistically different from zero $\chi^2(1) = 1.70$, $p = 0.19$. The 95% confidence interval for $\Delta \widehat{\log(\frac{w_Y}{w_Z})}$ is $[-0.07, 0.01]$, indicating that we can reject a decline in the ratio w_Y/w_Z more extreme than 7%. While these statistics do not account for multiple hypotheses, a joint test also fails to reject the null hypothesis that equalizing reductions are constant across ranks, $\chi^2(3) = 4.50$, $p = 0.21$. In Appendix Tables AIII and AV of the Supplemental Material (Bernheim and Sprenger (2020)) , we reproduce these analyses with individual fixed effects and robust standard errors, and reach identical conclusions.

TABLE III

EQUALIZING REDUCTIONS[a]

| | Panel A: Mean Behavior and Estimated Rank Dependence | | | Panel B: CPT Estimates and Predicted Rank Dependence | | |
| | | | | Equalizing Reductions $Y > \underline{X} > Z$ $\gamma = 0.784\,(0.020)$ | Equalizing Reductions $\overline{X} > Y > Z$ $\gamma = 0.830\,(0.021)$ | Certainty Equivalents $\gamma = 0.715\,(0.015)$ |
$\{p, q, 1-p-q\}$	$\underline{k}$ $\underline{\phi} \equiv \frac{m}{\underline{k}}\frac{q}{1-p-q}$	$\overline{k}$ $\overline{\phi} \equiv \frac{m}{\overline{k}}\frac{q}{1-p-q}$	$\Delta\widehat{\log(\frac{w_Y}{w_Z})}$ [95% Conf.]	$\Delta\widehat{\log(\frac{w_Y}{w_Z})}$ [95% Conf.]	$\Delta\widehat{\log(\frac{w_Y}{w_Z})}$ [95% Conf.]	$\Delta\widehat{\log(\frac{w_Y}{w_Z})}$ [95% Conf.]
$\{0.6, 0.3, 0.1\}$	9.02 (0.39) 1.66 (0.07)	9.28 (0.38) 1.62 (0.07)	0.03 (0.02) [−0.01, 0.07]	−0.18 (0.01) [−0.21, −0.15]	−0.14 (0.02) [−0.18, −0.11]	−0.23 (0.01) [−0.25, −0.21]
$\{0.4, 0.3, 0.3\}$	4.31 (0.12) 1.16 (0.03)	4.34 (0.12) 1.15 (0.03)	0.01 (0.02) [−0.03, 0.05]	−0.35 (0.03) [−0.41, −0.28]	−0.27 (0.04) [−0.34, −0.20]	−0.47 (0.03) [−0.52, −0.41]
$\{0.1, 0.3, 0.6\}$	2.63 (0.08) 0.95 (0.03)	2.56 (0.07) 0.98 (0.03)	−0.03 (0.02) [−0.07, 0.01]	−0.22 (0.02) [−0.30, −0.19]	−0.18 (0.03) [−0.24, −0.13]	−0.35 (0.03) [−0.40, −0.30]

[a]Panel A: Mean behavior for $\underline{k}$ and $\overline{k}$ estimated from interval regression (Stewart (1983)) of experimental response on indicators for probability vector interacted with indicator for whether $X > Y$. Calculated values of $\underline{\phi} \equiv \frac{m}{\underline{k}}\frac{q}{1-p-q}$ and $\overline{\phi} \equiv \frac{m}{\overline{k}}\frac{q}{1-p-q}$. Estimated change in relative decision weights, $\Delta\log(w_Y/w_Z)$, calculated as $\Delta\log(k)$. Standard errors clustered at individual level and calculated using the delta method, in parentheses. See Appendix Table AIII, column (1) and Appendix Table AIV of the Supplemental Material (Bernheim and Sprenger (2020)) for detail. Panel B: Predicted change in probability weight for CPT decisionmaker with probability weighting estimated from equalizing reductions with $Y > \underline{X} > Z$, from equalizing reductions with $\overline{X} > Y > Z$, or from certainty equivalents. Estimated probability weighting parameter noted for each prediction. Change in relative decision weights, $\Delta\log(w_Y/w_Z)$, calculated as $\log(\pi(p+q) - \pi(p)) - \log(\pi(q))$ for estimated weighting function. Standard errors clustered at individual level and calculated using the delta method, in parentheses.

estimates, we can rule out any change in probability weight for outcome Y in excess of 1% for the first probability vector, 3% for the second, and 7% for the third.

Another way to gauge the magnitudes of the observed changes in the equalizing reductions is to parameterize the probability weighting function and determine the amount of curvature needed to rationalize these changes under the maintained hypothesis of CPT. We adopt Tversky and Kahneman's (1992) specification and focus on the curvature parameter (γ). The estimates of $\Delta \log(\frac{w_Y}{w_Z})$ in the third column of Table III, Panel A, imply the following: for $\{p, q, 1 - p - q\} = \{0.6, 0.3, 0.1\}$, $\gamma = 1.040$ (clustered s.e. $= 0.027$); for $\{p, q, 1 - p - q\} = \{0.4, 0.3, 0.3\}$, $\gamma = 1.006$ (0.014); for $\{p, q, 1 - p - q\} = \{0.1, 0.3, 0.6\}$, $\gamma = 0.971$ (0.022).[36] With so little curvature in the probability weighting function, decision weights cannot be meaningfully rank-dependent.

Because we have conducted our analysis for a small number of probability vectors, we cannot claim to have ruled out nonlinearities in probability weighting comprehensively (under the maintained hypothesis of rank dependence). Technically, recalling equation (4), what we have shown is that $\pi(\cdot)$ has the same average slope on the intervals $[0, 0.3]$, $[0.1, 0.4]$, $[0.4, 0.7]$, and $[0.6, 0.9]$. This finding does not rule out fortuitous "wiggles" on $[0, 0.9]$, nor systematic nonlinearities within the interval $[0.9, 1]$, such as a discontinuity at unity. Addressing the latter possibility is straightforward: one simply elicits equalizing reductions for values of X below Z. We implement that strategy as part of an extension in Section 5.3.

Before moving on to the next step in our analysis, we highlight one additional implication of our findings: even under PT, the flatness of the $X - k$ schedule within the regimes $X > Y$ and $X < Y$ implies either that the reference point is invariant with respect to X (contrary to the theories of Bell (1985), Loomes and Sugden (1986)), or that the function $u(\cdot)$ is approximately linear. As we show below, other features of our data are inconsistent with the second possibility.

4.2.2. *The Effect of Probability on Decision Weight (Step 2)*

We now turn our attention to Step 2, in which we draw inferences about the shape of the probability weighting function from responses (in the same tasks) to variations in probabilities holding ranks fixed. If the probability weighting function is indeed linear (as the first step of our analysis implies under the maintained hypothesis of rank dependence), then we should find that $\underline{\phi}$ and $\overline{\phi}$ are insensitive to these variations. And yet, as shown in Table III, Panel A, the mean value of $\underline{\phi}$ ranges from 1.66 (0.07) when

---

[36]Under the maintained hypothesis of CPT, for the case of $\overline{X} > Y > Z$,

$$\overline{k} = Z - u^{-1}\left(u(Z) - \left[\frac{\pi(p+q) - \pi(p)}{1 - \pi(p+q)}\right][u(Y+m) - u(Y)]\right),$$

while for $Y > \underline{X} > Z$,

$$\underline{k} = Z - u^{-1}\left(u(Z) - \left[\frac{\pi(q)}{1 - \pi(p+q)}\right][u(Y+m) - u(Y)]\right).$$

We assume these equalities are satisfied up to an additive error term for the midpoint of k defined by experimental choice, and use the functional forms of Tversky and Kahneman (1992): $u(x) = x^\alpha$, $\pi(p) = p^\gamma/(p^\gamma + (1-p)^\gamma)^{1/\gamma}$. We then estimate the parameters γ and α by minimizing the sum of squared residuals for the corresponding nonlinear combined regression equation for $\underline{k}$ or $\overline{k}$ for each probability vector separately (combining values of $\overline{X} > Y > Z$ and $Y > \underline{X} > Z$). The estimates for the parameter α are: for $\{p, q, 1 - p - q\} = \{0.6, 0.3, 0.1\}$, $\alpha = 0.211$ (clustered s.e. $= 0.100$); for $\{p, q, 1 - p - q\} = \{0.4, 0.3, 0.3\}$, $\alpha = 0.711$ (0.055); for $\{p, q, 1 - p - q\} = \{0.1, 0.3, 0.6\}$, $\alpha = 1.063$ (0.055).

$p = 0.6$, to 1.16 (0.03) when $p = 0.4$, to 0.95 (0.03) when $p = 0.1$. Similarly, the mean value of $\overline{\phi}$ ranges from 1.62 (0.07) when $p = 0.6$, to 1.15 (0.03) when $p = 0.4$, to 0.98 (0.03) when $p = 0.1$. In both cases, we reject the hypothesis of equality across values of p ($\chi^2(1) = 110.7$, p-value < 0.01) for $\underline{\phi}$, and ($\chi^2(2) = 103.6$, p-value < 0.01) for $\overline{\phi}$.

How should we interpret the magnitudes of these differences in $\underline{\phi}$ and $\overline{\phi}$ across distinct values of p? Focusing on the regime in which $Y > \underline{X}$, we have

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

In other words, the percentage change in $\underline{\phi}$ across values of p roughly equals the percentage change in the average slope of $\pi(\cdot)$ on $[p + q, 1]$. When we decrease $p + q$ from 0.9 to 0.7, we have $\Delta \log(\underline{\phi}) = -0.36$ (0.03). When we decrease $p + q$ from 0.7 to 0.4, we have $\Delta \log(\underline{\phi}) = -0.20$ (0.03). The values of $\Delta \log(\underline{\phi})$ indicate the average slope of $\pi(\cdot)$ declines by approximately 30% (i.e., $\exp(-0.36) - 1$) between the regions $[0.9, 1]$ and $[0.7, 1]$, and 18% (i.e., $\exp(-0.20) - 1$) between the regions $[0.7, 1]$ and $[0.4, 1]$.

Using the estimates of $\underline{\phi}$ in Table III, we can determine how the average slope of $\pi(\cdot)$ varies across the adjacent subintervals $[0.4, 0.7]$, $[0.7, 0.9]$, and $[0.9, 1]$, thereby narrowing down the location of the nonlinearities.[37] Specifically, we obtain $\log(\frac{\pi(0.9)-\pi(0.7)}{0.2}) - \log(\frac{\pi(1)-\pi(0.9)}{0.1}) = -0.60$ (0.06) and $\log(\frac{\pi(0.7)-\pi(0.4)}{0.3}) - \log(\frac{\pi(0.9)-\pi(0.7)}{0.2}) = -0.21$ (0.06). These calculations imply a reduction in the average slope of $\pi(\cdot)$ of approximately 45% (i.e., $\exp(-0.60) - 1$) between the regions $[0.9, 1]$ and $[0.7, 0.9]$, and a further reduction of approximately 19% (i.e., $\exp(-0.21) - 1$) between the regions $[0.7, 0.9]$ and $[0.4, 0.7]$. In both cases, we reject the null hypothesis of constant average slope over the relevant region: for $[0.9, 1]$ to $[0.7, 0.9]$, $\chi^2(1) = 106.4$ (p-value < 0.01); and for $[0.7, 0.9]$ to $[0.4, 0.7]$, $\chi^2(1) = 10.1$ (p-value < 0.01). Thus, the responses to variations in probabilities (with fixed ranks) imply that $\pi(\cdot)$ is highly nonlinear on the interval $[0.4, 1]$. Notably, this nonlinearity is not confined to a small neighborhood around 1. On the contrary, the implied slope of $\pi(\cdot)$ varies considerably within the interval $[0.4, 0.9]$.

Another way to gauge the magnitudes of the observed changes in equalizing reductions across values of p is to parameterize the probability weighting function and determine the amount of curvature needed to rationalize these changes. Once again we adopt Tversky and Kahneman's (1992) specification and focus on the curvature parameter (γ). For choices with $Y > \underline{X} > Z$, we estimate $\gamma = 0.830$ (0.022), while for choices with $\overline{X} > Y > Z$, we estimate $\gamma = 0.784$ (0.020).[38] Significantly, these magnitudes align

---

[37]Note that

$$\log\left(\frac{\frac{\pi(0.9) - \pi(0.7)}{0.2}}{\frac{\pi(1) - \pi(0.9)}{0.1}} \right) = \log\left(\frac{3}{2}\left(\frac{\frac{\pi(1) - \pi(0.7)}{0.3}}{\frac{\pi(1) - \pi(0.9)}{0.1}} \right) - \frac{1}{2} \right),$$

and

$$\log\left(\frac{\frac{\pi(0.7) - \pi(0.4)}{0.3}}{\frac{\pi(0.9) - \pi(0.7)}{0.2}} \right) = \log\left(2\left(\frac{\frac{\pi(1) - \pi(0.4)}{0.6}}{\frac{\pi(1) - \pi(0.9)}{0.1}} \frac{\frac{\pi(1) - \pi(0.9)}{0.1}}{\frac{\pi(0.9) - \pi(0.7)}{0.2}} \right) - \frac{1}{3}\left(\frac{\frac{\pi(1) - \pi(0.9)}{0.1}}{\frac{\pi(0.9) - \pi(0.7)}{0.2}} \right) - \frac{2}{3} \right).$$

[38]Following the procedure in footnote 36, we again assume that the formulas for $\overline{k}$ and $\underline{k}$ are satisfied up to an additive error term for the midpoint of k defined by experimental choice, and use the functional forms of

much more closely with our estimate from the certainty equivalent tasks ($\gamma = 0.715$) than with those based on cross-regime changes in k ($\gamma \approx 1$).

Moreover, the estimates of γ reported in the preceding paragraph imply that, under the maintained hypothesis of rank dependence, we ought to have observed substantial differences between $\underline{k}$ and $\overline{k}$ for each fixed probability vector in the first step of our analysis. These predictions appear in the first and second columns of Table III, Panel B. As is clear from the table, with either $\gamma = 0.830$ or $\gamma = 0.784$ (along with the associated estimates of α), the implied percentage reductions in the relative decision weight on outcome Y resulting from a reversal of X and Y far exceed the observed percentage declines in the measured equalizing reductions. For example, when $\gamma = 0.830$ and $\{p, q, 1 - p - q\} = \{0.6, 0.3, 0.1\}$, the implied value of $\Delta \log(\frac{\omega_Y}{\omega_Z})$ is -0.14 (0.02), which translates into a 13% decline in $\frac{\omega_Y}{\omega_Z}$ as ranks change. Predictions for the other probability vectors, and for $\gamma = 0.784$, are even more extreme. Regardless of whether we use $\gamma = 0.830$ or $\gamma = 0.784$, and for every probability vector we consider, the 95% confidence interval for the predicted value of $\Delta \log(\frac{\omega_Y}{\omega_Z})$ (under the maintained hypothesis of rank dependence) does not overlap with the corresponding interval for the observed difference, which is always concentrated near zero.[39]

We also performed analogous calculations using the estimated value of the curvature parameter obtained from the certainty equivalent tasks, $\gamma = 0.715$, along with the associated estimate of α. These calculations are of interest because economists routinely adopt these types of CPT calibrations for the purpose of studying applied problems; prominent examples using the probability weighting parameters derived from the binary lotteries in Tversky and Kahneman (1992) include Benartzi and Thaler (1995), Barberis and Huang (2008), and Barberis, Mukherjee, and Wang (2016). Such work proceeds from the assumption that valuations of binary lotteries reveal the values of "deep" CPT preference parameters that are stable across a wide range of contexts. We use the certainty equivalent calibration to predict the effect of rank reversals on equalizing reductions; results appear in the third column of Table III, Panel B. Once again, the predicted changes are an order of magnitude greater than the observed changes. Figure 3, Panel B superim-

---

Tversky and Kahneman (1992): $u(x) = x^\alpha$, $\pi(p) = p^\gamma/(p^\gamma + (1 - p)^\gamma)^{1/\gamma}$. We then estimate the parameters γ and α by minimizing the sum of squared residuals for the corresponding nonlinear regression equations for $\underline{k}$ and $\overline{k}$ using data from the regimes $\overline{X} > Y > Z$ and $Y > \underline{X} > Z$ separately (combining probability vectors). For $\overline{X} > Y > Z$, we estimate $\alpha = 1.024$ (clustered s.e. $= 0.082$) and $\gamma = 0.830$ (0.022). For $Y > \underline{X} > Z$, we estimate $\alpha = 0.911$ (0.063), and $\gamma = 0.784$ (0.020).

[39]With mean estimates and standard errors for predicted $\Delta \log(w_Y/w_Z)$ and actual $\Delta \log(k)$ and assumptions of normality for both, hypothesis tests for equality between predicted and actual rank dependence are easily implemented via calculation of the following test statistic:

$$z = \frac{\widehat{\Delta \log(w_Y/w_Z)} - \Delta \log(k)}{\sqrt{\text{s.e.}\big(\widehat{\Delta \log(w_Y/w_Z)}\big)^2 + \text{s.e.}\big(\Delta \log(k)\big)^2}}.$$

Under the null hypothesis of equality, the distribution of z is standard normal. For calibrations based on our certainty equivalent tasks, we find $z = 12.0$ $(p < 0.01)$, $z = 13.9$ $(p < 0.01)$, and $z = 9.7$ $(p < 0.01)$ for $\{p, q, 1 - p - q\} = \{0.6, 0.3, 0.1\}$, $\{0.4, 0.3, 0.3\}$, and $\{0.1, 0.3, 0.6\}$, respectively. For calibrations based on our equalizing reduction tasks with $\overline{X} > Y$, these statistics are $z = 6.9$ $(p < 0.01)$, $z = 6.8$ $(p < 0.01)$, and $z = 4.5$ $(p < 0.01)$; and for calibrations based on our equalizing reduction tasks with $\underline{X} < Y$, they are $z = 8.8$ $(p < 0.01)$, $z = 9.0$ $(p < 0.01)$, and $z = 6.2$ $(p < 0.01)$. Applying a Bonferroni correction for multiple hypotheses for these $n = 9$ tests requires a p-value below $0.01/9 = 0.001$ or a z-score above 3.26 to retain the interpretation of 1% statistical signifcance. As all test statistics lie above this value, they are clearly robust to a multiple hypotheses correction.

poses the predicted values of $\underline{k}$ and $\overline{k}$ over the measured values. It shows the predicted values along with 95% confidence intervals, which appear as dashed lines.[40] Substantial discontinuities are readily apparent.[41] Importantly, the figure shows that—according to the calibrated CPT model—a rank-altering change in X should have roughly the same effect on the equalizing reduction as moving 30% of the probability mass from the high to low outcome under the fixed ranking $Y > \underline{X} > Z$. Subjects respond strongly (and as predicted) to the latter change, but not at all to the former.

One can, of course, object to comparisons between preference parameters estimated from certainty equivalent tasks, and from equalizing reduction tasks, on the grounds that subjects might use different probability weighting functions for different classes of tasks. However, this hypothesized failure of "procedural invariance" (Starmer (2000)) does not apply to our comparisons involving responses of equalizing reductions to, respectively, changes in probabilities and changes in ranks within the same tasks.

In summary, Step 1 uses variation in outcome ranks (with fixed probabilities) to establish the absence of meaningful nonlinearities in the probability weighting function on the subinterval $[0, 0.9]$ under the maintained hypothesis of rank dependence, while Step 2 uses variation in probabilities (with fixed ranks) to establish the presence of substantial nonlinearities on the subinterval $[0.4, 1]$, and, critically, within the overlapping portion $[0.4, 0.9]$. The most natural explanation for the direct conflict between our findings in Step 1 and Step 2 is that probability weighting is indeed nonlinear but rank-independent, as envisioned in PT.

Technically, one could potentially reconcile our main findings with CPT by positing a probability weighting function with fortuitously located "wiggles." For example, if the slope of $\pi(\cdot)$ increased by 23% (i.e., $\exp(0.21) - 1$) between $[0.1, 0.4]$ and $[0.4, 0.6]$, decreased by 57% (i.e., $\exp(-0.84) - 1$) between $[0.4, 0.6]$ and $[0.6, 0.7]$, and increased by 132% (i.e., $\exp(0.84) - 1$) between $[0.6, 0.7]$ and $[0.7, 0.9]$, then the average slopes on $[0.1, 0.4]$, $[0.4, 0.7]$, and $[0.6, 0.9]$ would be identical (as we found in Step 1 under the maintained hypothesis of rank dependence), and the average slope on $[0.4, 0.7]$ would be 19% lower than the average slope on $[0.7, 0.9]$ (as we found in Step 2).[42] As far as we

---

[40]We obtain closed-form solutions for $\underline{k}$ and $\overline{k}$ based on Eq. (3), and derive standard errors using the delta method.

[41]Appendix Figure A.4 of the Supplemental Material (Bernheim and Sprenger (2020)) exhibits similar predictions using the values of γ and α implied by the responses of equalizing reductions to variations in probabilities holding ranks fixed.

[42]Note that if $\frac{\pi(0.4)-\pi(0.1)}{0.3} = \frac{\pi(0.7)-\pi(0.4)}{0.3} = \frac{\pi(0.9)-\pi(0.6)}{0.3}$,

$$\log\left(\frac{\frac{\pi(0.7)-\pi(0.6)}{0.1}}{\frac{\pi(0.9)-\pi(0.7)}{0.2}}\right) = \log\left(3\left(\frac{\frac{\pi(0.7)-\pi(0.4)}{0.3}}{\frac{\pi(0.9)-\pi(0.7)}{0.2}}\right) - \frac{2}{3}\right) = \log\left(3(\exp(-0.21)) - \frac{2}{3}\right) = -0.84,$$

$$\log\left(\frac{\frac{\pi(0.9)-\pi(0.7)}{0.2}}{\frac{\pi(0.6)-\pi(0.4)}{0.2}}\right) = 0,$$

and

$$\log\left(\frac{\frac{\pi(0.4)-\pi(0.1)}{0.3}}{\frac{\pi(0.6)-\pi(0.4)}{0.2}}\right) = \log\left(\frac{1}{3}\left(\frac{\frac{\pi(0.7)-\pi(0.6)}{0.1}}{\frac{\pi(0.6)-\pi(0.4)}{0.2}}\right) + \frac{2}{3}\right) = \log\left(\frac{1}{3}(\exp(-0.84)) + \frac{2}{3}\right) = -0.21.$$

know, nothing in the literature suggests that the probability weighting function exhibits this distinctively peculiar wiggle. In any case, we speculate that one could rule it out using our methods by eliciting equalizing reductions for additional probability vectors.

Our main results also leave open the possibility that a rank-dependent probability weighting function with the requisite wiggle could exhibit nonlinearities on $[0.9, 1]$, such as a discontinuity at unity. We examine and reject that possibility in Section 5.3 by extending our analysis to lotteries with payoff ranks $Y > Z > X$. This rejection should come as no surprise in light of several considerations. First, with these limited nonlinearities, CPT loses its power to explain some of the puzzles that account for its popularity, such as the common ratio effect. Second, as is apparent in Figure 3, Panel A, choices in the certainty equivalent tasks imply a completely different pattern of nonlinearities. While it is certainly conceivable that the *degree* of probability weighting might vary with the nature of the elicitation task due to a failure of procedural invariance, it is less plausible that the *qualitative nature* of probability weighting would fundamentally change from task to task—and if it does, then there is certainly no justification for using conventionally parameterized CPT models to describe real-world phenomena, such as portfolio choices (as in Barberis (2018)). Moreover, with the sole exception of implying rank dependence where none exists, the CPT model calibrated with certainty equivalent tasks predicts choices in the equalizing reduction tasks with reasonable accuracy. For example, as long as payoffs respect the ranking $Y > \underline{X} > Z$, the predictions track behavior reasonably well, matching differences in levels of equalizing reductions across probability vectors.[43]

4.2.3. *Which Model Best Explains the Data on Equalizing Reductions?*

To underscore the fact that the CPT model cannot simultaneously account for the responses to probabilities and the absence of responses to ranks in the equalizing reduction tasks, we use the data for the equalizing reduction tasks (and only those data) to estimate a model of the form

$$k = \delta k_C(\alpha, \gamma) + (1 - \delta)k_P(\alpha, \gamma) + \epsilon, \tag{6}$$

where k is the midpoint of the equalizing reduction interval defined by experimental choice, $k_C(\alpha, \gamma)$ is the equalizing reduction implied by a CPT model with parameters (α, γ), k_P is the value implied by an otherwise identical PT model (without rank dependence), and ϵ is a disturbance term.[44] We interpret this model as depicting a population consisting of CPT and PT decisionmakers in the proportions δ and $1 - \delta$, respectively. If k_C and k_P represent the average equalizing reductions for the CPT and PT decisionmakers, respectively, then the expected equalizing reduction for a randomly selected member of the population is $\delta k_C + (1 - \delta)k_P$, precisely as the model claims.

Estimating this equation with nonlinear least squares (and clustering errors at the subject level), we obtain: $\alpha = 0.684$ (0.059), $\gamma = 0.794$ (0.019), and $\delta = -0.105$ (0.094). Notice that the probability weighting function exhibits substantial curvature. As a result, δ must be close to zero to account for the absence of noticeable differences between $\overline{k}$ and $\underline{k}$.

---

[43]Given that the calibrated CPT model is nearly linear in payoffs, the difference between its predictions and those of a risk-neutral objective function are primarily attributable to probability weighting.

[44]The formulation of $\overline{k}_C$ and $\underline{k}_C$ is noted in footnote 36, and

$$\overline{k}_P = \underline{k}_P = Z - u^{-1}\left(u(Z) - \left[\frac{\pi(q)}{\pi(1 - p - q)}\right]\left[u(Y + m) - u(Y)\right]\right).$$

Note that this formulation features rank-independent treatment of the probabilities of receiving Y and Z.

Indeed, the point estimate of the weight on the CPT model is negative, and we can reject the hypothesis that it is greater than 8 percent (0.08) with 95 percent confidence. Reestimating the model with the constraint that $\delta \in [0, 1]$, we find $\delta = 0$. Thus, PT accounts for subjects' choices in the equalizing reduction tasks better than any other mixture of PT and CPT.

The preceding finding suggests that the PT formulation of probability weighting does a better job fitting the data than either the CPT or EU formulations. Each of these models is nested in Eq. (6), with $\delta = 0$, $\delta = 1$, and $\gamma = 1$ corresponding to the PT, CPT, and EU formulations (respectively). Assuming the error, ϵ, is distributed $N(0, \sigma^2)$, we estimate these restricted models via maximum likelihood to provide standard fit comparisons.[45] The PT formulation of $\delta = 0$ provides the highest likelihood and its improved fit over both CPT and the more parsimonious EU formulation is validated by both the Akaike and Bayes Information Criteria. Thus, we find that the PT formulation of probability weighting fits the data substantially better than the other two models.

5. ROBUSTNESS CHECKS

In the following subsections, we examine various potential explanations for our results other than rank independence, including the possibilities that our findings may reflect unrepresentative subjects (Section 5.1), order effects (Section 5.2), "cancellation" of common outcomes (Section 5.3), and inattention to the parameters of the decision tasks (Section 5.4). We present additional evidence that rules out each of these possibilities. In Appendix F of the Supplemental Material (Bernheim and Sprenger (2020)), we provide supplemental analyses that examine alternative formulations for CPT using the functional forms of Prelec (1998), explore behavior in tasks where $X = \$25$ (which implies that adding $m = \$5$ to $Y = \$24$ changes the ranking), and include potentially confused subjects who switch more than once in a given task. These exercises all yield the same conclusion: decision weights are rank-independent (or nearly so), and CPT is soundly rejected.

5.1. *Subject-Level Analysis*

First we address the possibility that our findings might reflect the behavior of unrepresentative subjects or other forms of heterogeneity. Each subject in our experiment provides us with data on equalizing reductions and certainty equivalents. Accordingly, we can replicate our analysis at the subject level. For each subject and each probability vector, we calculate the average $\underline{k}$ and $\overline{k}$ for values of $\underline{X} < Y$ and $\overline{X} > Y + m$, respectively, and compute the actual change, $\Delta \log(k)$.[46] Additionally, we use each subject's certainty equivalent data to estimate their CPT risk preference parameters based on equation (5), and then use the curvature parameter, γ, to predict the change in the log of relative decision weights—$\Delta \log(w_Y/w_Z) = \log(\pi(p + q) - \pi(p)) - \log(\pi(q))$—and, hence, in equalizing reductions, for each probability vector.

Panel A of Figure 4 presents the distributions of the predicted values of $\Delta \log(w_Y/w_Z)$ and the actual values of $\Delta \log(k)$ for each of the three probability vectors, along with their

---

[45]The maximum likelihood estimates are as follows: for $\delta = 0$ (PT), $\alpha = 0.724$ (clustered s.e. $= 0.049$), $\gamma = 0.789$ (0.019), LL $= -6520.2$, AIC $= 13{,}046.5$, BIC $= 13{,}064.0$; for $\delta = 1$ (CPT), $\alpha = 0.593$ (0.051), $\gamma = 0.915$ (0.015), LL $= -6557.4$, AIC $= 13{,}120.7$, BIC $= 13{,}138.3$; for $\gamma = 1$ (EU), $\alpha = 0.366$ (0.065), LL $= -6567.3$, AIC $= 13{,}138.6$, BIC $= 13{,}150.3$.

[46]This calculation is based on the midpoints of the intervals for $\underline{k}$ or $\overline{k}$ implied by each subject's switch point.
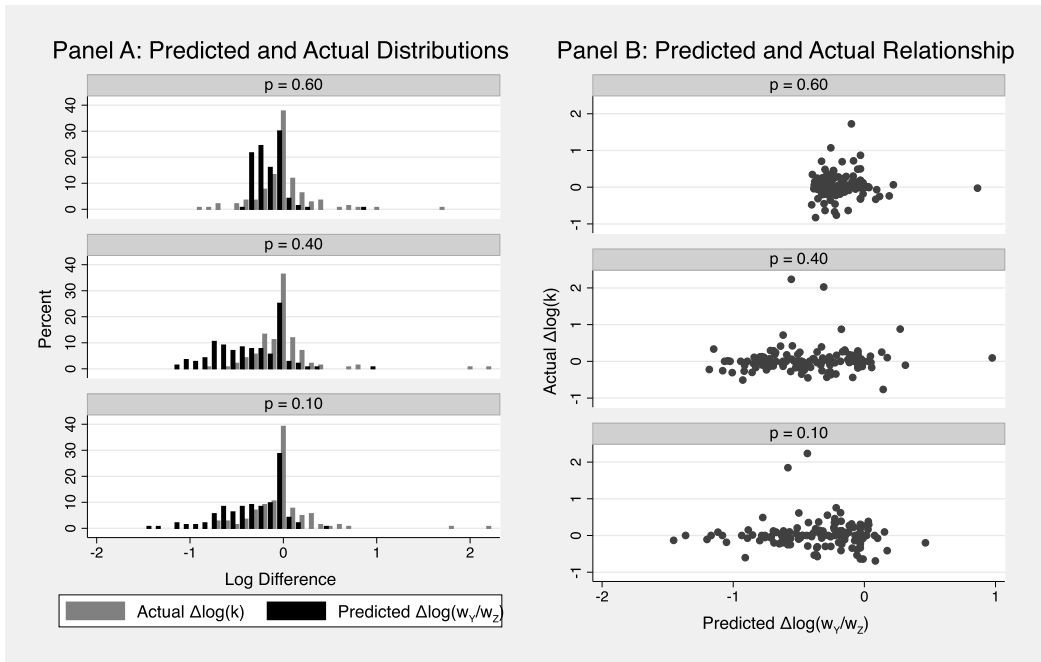
FIGURE 4.—Individual results. *Notes*: Mean actual value of $\Delta \log(k)$ calculated for each individual in each probability set. Predicted value of $\Delta \log(w_Y/w_Z)$ calculated for each individual based on equation (3) for CPT decisionmaker with parameters estimated from each individual's certainty equivalent responses.

relationship for our 143 subjects. For $\{p, q, 1 - p - q\} = \{0.6, 0.3, 0.1\}$, the median actual value for $\Delta \log(k)$ is 0 and the interquartile range is $[-0.05, 0.12]$. Results for the other probability vectors are similar.[47] Thus, the changes in equalizing reductions for our subjects, and hence the changes in their relative decision weights, are concentrated within a small band around zero. Though the equalizing reductions exhibit almost no rank dependence, substantial rank dependence is predicted according to the estimated subject-level CPT models; see also Figure 4, Panel A.[48] Panel B shows the subject-level relationship between predicted and actual rank dependence. For every probability vector, the correlation between the predicted value, $\Delta \log(w_Y/w_Z)$, and the actual value, $\Delta \log(k)$, is indistinguishable from zero.[49] We obtain similar results when we use subject-level estimates of the curvature parameter for the probability weighting function based on responses to variations in probabilities for the equalizing reduction tasks rather than for the certainty

---

[47]For $\{p, q, 1 - p - q\} = \{0.4, 0.3, 0.3\}$, the median actual value for $\Delta \log(k)$ is 0 and the interquartile range is $[-0.10, 0.10]$, while for $\{p, q, 1 - p - q\} = \{0.1, 0.3, 0.6\}$, the median actual value for $\Delta \log(k)$ is 0 and the interquartile range is $[-0.11, 0.10]$.

[48]For $\{p, q, 1 - p - q\} = \{0.6, 0.3, 0.1\}$, the median predicted value for $\Delta \log(w_Y/w_Z)$ is -0.17, for $\{p, q, 1 - p - q\} = \{0.4, 0.3, 0.3\}$, it is -0.33, and for $\{p, q, 1 - p - q\} = \{0.1, 0.3, 0.6\}$, it is -0.23. Wilcoxon signed rank tests for equivalent distributions across predicted and actual rank dependence yield the following test statistics: $z = 7.00$ ($p < 0.01$), $z = 8.77$ ($p < 0.01$), and $z = 7.54$ ($p < 0.01$) for $\{p, q, 1 - p - q\} = \{0.6, 0.3, 0.1\}$, $\{0.4, 0.3, 0.3\}$, and $\{0.1, 0.3, 0.6\}$, respectively.

[49]Correlations between predicted and actual values are $\rho = 0.01$, ($p = 0.86$), $\rho = 0.06$, ($p = 0.46$), and $\rho = -0.03$ ($p = 0.69$) for $\{p, q, 1 - p - q\} = \{0.6, 0.3, 0.1\}$, $\{0.4, 0.3, 0.3\}$, and $\{0.1, 0.3, 0.6\}$, respectively.

equivalent tasks.[50] Thus, our conclusions do not rely on an assumption of procedural invariance.

Our subject-level results provide striking evidence of rank-independent probability weighting. The majority of subjects exhibit no evidence of quantitatively consequential rank dependence as ranks change from $\overline{X} > Y$ to $\underline{X} < Y$. Further, the individualized CPT models predict sizable percentage changes in relative weights for many subjects, and those predictions bear no relation to the actual magnitudes. Plainly, our main findings are not driven by unrepresentative subjects or other forms of heterogeneity.

5.2. *Order Effects*

Our main findings exploit within-subject variation in payoff rank. If a subject's early responses in the tasks used to elicit equalizing reductions somehow anchor their later responses, that approach could obscure rank dependence. We note, however, that responses often change considerably at the individual level from one block of tasks to the next. For example, between the first and second block of tasks, 59% of individual responses differ and 37% of responses differ by more than 25 percent.[51] Order effects certainly do not appear to dampen within-subject responses to changes in the probability vector, hence it is hard to see why they should do so for rank-order effects.

To address any residual concerns about anchoring, we replicate our analysis using only the first task block for each subject. Recall that payoff rank is fixed within each block. It follows that, with this alternative approach, we identify possible rank dependence entirely from between-subject variation using responses that are untainted by anchoring. For such between-subjects analysis, we rely only on the variation in the ordering of tasks across sessions. To account for selection on observable characteristics, we additionally include measures of gender, age, and cognitive ability from a post-study questionnaire and each

---

[50]Unlike our certainty equivalent estimates, where predictions for rank dependence can be made for every subject, estimates from equalizing reductions do not reliably converge within 200 iterations for every subject. For $\underline{X} < Y$, predictions for rank dependence can be made for 133 of 143 subjects, while for $\overline{X} > Y$, predictions can be made for 137 subjects. Predictions of $\Delta \log(w_Y/w_Z)$ can be made for all three methodologies for 130 subjects. For one of these subjects, we predict $\Delta \log(w_Y/w_Z) > 10$. For the remaining 129 subjects, the cross-subject correlations between the measured $\Delta \log(k)$ and predicted $\Delta \log(w_Y/w_Z)$ are 0.02 when inferring γ from certainty equivalent tasks, 0.13 when inferring from equalizing reduction tasks with $\underline{X} < Y$, and 0.04 when inferring from equalizing reduction tasks with $\overline{X} > Y$. The cross-subject correlations between predicted values of $\Delta \log(w_Y/w_Z)$ based on certainty equivalent tasks and based on equalizing reduction tasks with $\underline{X} < Y$ is 0.46; for equalizing reduction tasks with $\overline{X} > Y$, the correlation is 0.42. The cross-subject correlation between predicted values of $\Delta \log(w_Y/w_Z)$ based on equalizing reduction tasks with $\underline{X} < Y$ and those with $\overline{X} > Y$ is 0.49.

[51]The 143 subjects make three decisions in each task block, yielding 429 potential differences across the first and second task blocks. Of these, 175 responses (41%) exhibit no change, 100 responses (23%) increase, and 154 (36%) decrease. The order of the first two task blocks has no measurable relationship with changes. Forty-four subjects began with $\underline{X} < Y$ first and then proceeded to $\overline{X} > Y$, giving 132 potential differences. Of these, 57 responses (43%) exhibit no change, 34 responses (26%) increase, and 41 (31%) decrease. Forty-one subjects began with $\overline{X} > Y$ first and then proceeded to $\underline{X} < Y$, giving 123 potential differences. Of these, 50 responses (41%) exhibit no change, 33 responses (27%) increase, and 40 (33%) decrease. Under the estimated CPT model, one would expect more frequent decreases for subjects with $\underline{X} < Y$ first and more frequent increases for subjects with $\overline{X} > Y$ first. The differences in response across the first two blocks does not seem localized to a limited number of subjects, with only 28 of 143 subjects (20%) exhibiting no change across any of the three decisions, and 45 (31%) exhibiting a change in all three.

subject's average certainty equivalent in their binary lottery tasks as a measure of risk aversion.[52]

Table IV presents between-subjects results based on the first task blocks with and without the controls noted above.[53] We see a hint of rank dependence, particularly for $\{p, q, 1 - p - q\} = \{0.6, 0.3, 0.1\}$, without controls in Panel A. With controls in Panel B, we find essentially no differences between $\underline{k}$ and $\overline{k}$. In all cases, the degree of observed rank dependence falls far short of the various CPT benchmark predictions from Table III, Panel B.

Can one construe the small differences between $\underline{k}$ and $\overline{k}$ without controls as limited evidence of rank-dependent probability weighting and CPT? In our view, any such inference would be unwarranted. With 143 subjects in total, the subsamples that first faced $\underline{X} < Y$ and $\overline{X} > Y$ are of modest size, and consequently not perfectly matched. Indeed, we find some hints of selection across these subsamples in observable characteristics such as gender, cognitive ability, and average risk aversion, all of which correlate highly with equalizing reductions (see Appendix Table AIII, Columns (5) and (6) of the Supplemental Material (Bernheim and Sprenger (2020)) for detail).[54] Absent controls for this potential selection, the differences in characteristics inflate $\overline{k}$ relative to $\underline{k}$, spuriously producing the appearance of modest rank dependence.[55]

5.3. *Rank Dependence and Cancellation*

Our experiment could also obscure rank dependence if decisionmakers adopt the heuristic practice of "canceling" common elements across available lotteries before evaluating them (see Wu (1994), Weber and Kirsner (1997)). Specifically, subjects may cancel the probability p chance of winning X common to both L and L_e when determining their equalizing reduction. In that case, they would ignore the changes in payoff ranks that occur when the value of X passes through either Y or Z.

To determine whether our results are attributable to cancellation, we designed a modified equalizing reduction task. In this task, we add m to X while reducing Y and Z by k. That is, the modified equalizing reduction identifies the value of k that makes a subject indifferent between the lottery $L = (\{p, q, 1 - p - q\}, \{X, Y, Z\})$ and the lottery

---

[52]Cognitive ability is measured with the three-question Cognitive Reflection Test introduced and validated in Frederick (2005).

[53]See Appendix Table AIII, columns (3) and (5) of the Supplemental Material (Bernheim and Sprenger (2020)) for further detail on these regressions. Of the 143 subjects in the primary sample, 21 had the $\underline{X} = 19$ block first, 23 had the $\underline{X} = 21$ block first, 21 had the $\underline{X} = 23$ block first, 22 had the $\overline{X} = 30$ block first, 19 had the $\overline{X} = 32$ block first, and 37 had the $\overline{X} = 34$ block first.

[54]Subjects who first faced $\underline{X} < Y$ ($\overline{X} > Y$) are 51% (45%) male, with Cognitive Reflection Test scores of 2.28 (2.02), and average certainty equivalents of 11.96 (11.35). Of these comparisons, the difference in risk aversion has a two-sided t-test p-value of 0.05 and the difference in cognitive ability has a two-sided p-value of 0.15. An omnibus test of selection from the regression of assignment to $\underline{X} < Y$ first on the controls of Table IV, Panel B yields $F(5, 134) = 1.60$ ($p = 0.17$), suggestive of the potential for selection on observables. These differences are of no consequence for the main portion of our analysis, which relies on within-subject variation, but could be influencing the results of Table IV, Panel A.

[55]In Appendix Table AVI of the Supplemental Material (Bernheim and Sprenger (2020)), we also present a specification that includes controls for gender, age, and cognitive ability, but omits our risk aversion proxy. The latter specification addresses any concerns that the average certainty equivalent in binary lottery tasks might be sensitive to the types of three-outcome lottery tasks ($\underline{X} < Y$ versus $\overline{X} > Y$) the subject encounters first, a possibility we regard as remote. Results for that specification are similar to those reported in Table IV, Panel B.

TABLE IV

EQUALIZING REDUCTIONS BETWEEN SUBJECTS[a]

$\{p, q, 1-p-q\}$	Panel A: First Task Block (without Controls)			Panel B: First Task Block (with Controls)		
	$\underline{k}$	$\overline{k}$	$\Delta \log(\widehat{\frac{w_Y}{w_Z}})$ [95% Conf.]	$\underline{k}$	$\overline{k}$	$\Delta \log(\widehat{\frac{w_Y}{w_Z}})$ [95% Conf.]
$\{0.6, 0.3, 0.1\}$	9.81 (0.65)	8.71 (0.56)	−0.12 (0.09) [−0.30, 0.06]	9.77 (0.60)	9.13 (0.56)	−0.07 (0.09) [−0.24, 0.10]
$\{0.4, 0.3, 0.3\}$	4.78 (0.19)	4.41 (0.19)	−0.08 (0.06) [−0.20, 0.04]	4.65 (0.22)	4.61 (0.21)	−0.01 (0.07) [−0.14, 0.12]
$\{0.1, 0.3, 0.6\}$	3.16 (0.16)	2.88 (0.12)	−0.09 (0.07) [−0.22, 0.04]	3.01 (0.20)	3.07 (0.16)	0.02 (0.09) [−0.15, 0.19]

[a]Mean behavior for $\underline{k}$ and $\overline{k}$ estimated from interval regression (Stewart (1983)) of experimental response on indicators for probability vector interacted with indicator for whether $X > Y$. Estimated change in relative decision weights, $\Delta \log(\widehat{w_Y/w_Z})$, calculated as $\Delta \log(k)$. Standard errors clustered at individual level and calculated using the delta method, in parentheses. See Appendix Tables AIII, columns (3) and (5) and Table AVIII of the Supplemental Material (Bernheim and Sprenger (2020)) for detail. Panel A: No controls; 143 total subjects. Panel B: controls include age, gender, Cognitive Reflection Task score, and mean certainty equivalent from seven certainty equivalents tasks; 135 total subjects.

$L_e = (\{p, q, 1-p-q\}, \{X+m, Y-k, Z-k\})$. Note that L and L_e share no common outcome with identical probabilities and payoffs, so there is nothing for the decisionmaker to cancel.

We can use these modified equalizing reductions as the basis for an alternative test of CPT probability weighting. Suppose in particular that we measure one equalizing reduction, k, for a lottery with $X > Y > Z$, another, k', for a lottery with $Y - k' > X' + m > X' > Z$, and a third, k'', for a lottery with $Z - k'' > X'' + m$. In that case, CPT implies

$$\frac{m}{k} \approx \frac{\pi(p+q) - \pi(p)}{\pi(p)} \frac{du/dY}{du/dX} + \frac{1 - \pi(p+q)}{\pi(p)} \frac{du/dZ}{du/dX}, \tag{7}$$

$$\frac{m}{k'} \approx \frac{\pi(q)}{\pi(p+q) - \pi(q)} \frac{du/dY}{du/dX'} + \frac{1 - \pi(p+q)}{\pi(p+q) - \pi(q)} \frac{du/dZ}{du/dX'}, \tag{8}$$

and

$$\frac{m}{k''} \approx \frac{\pi(q)}{1 - \pi(1-p)} \frac{du/dY}{du/dX''} + \frac{\pi(1-p) - \pi(q)}{1 - \pi(1-p)} \frac{du/dZ}{du/dX''}. \tag{9}$$

The differences between k, k', and k'' will depend on rank-dependent probability weights for X, X', X'', and Y along with marginal utilities for X, X', X'', Y, and Z. To build intuition, we will start with the case where utility is linear (so that marginal utility is fixed), while probability weights potentially vary. For small m, CPT predicts a discontinuous change in the equalizing reduction as the value of X passes through Y:

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

As long as p is not too large (so that $\pi(p) > p$), the concavity of the probability weighting curve for low probabilities implies that $\pi(p) > \pi(p+q) - \pi(q)$. It then follows that $\log(k) - \log(k') > 0$, which means that the equalizing reduction decreases discontinuously

as X passes below Y. Likewise, for small m, CPT also implies a discontinuous change in the equalizing reduction as the value of X passes through Z:

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

As long as p is not too large (so that $\pi(1-p) < 1-p$), the convexity of the probability weighting curve for large probabilities implies $1 - \pi(1-p) > \pi(p+q) - \pi(q)$. It then follows that $\log(k') - \log(k'') < 0$, which means that the equalizing reduction increases discontinuously as X passes below Z.

The key implication of the preceding analysis is that, under CPT, variations in X should produce discontinuities in the modified equalizing reduction of opposite signs depending on whether the value of X passes through Y (the high payoff) or Z (the low payoff). Our strategy is to test this distinctive implication. The following calculations provide an illustration using the parameterized probability weighting function of Tversky and Kahneman (1992). For $p = 0.4$, $q = 0.3$, we have $\pi(0.4) = 0.37$, $\pi(0.7) - \pi(0.3) = 0.22$, and $1 - \pi(1 - 0.4) = 0.53$, which implies opposing discontinuities: $\log(k) - \log(k') = 0.76$, and $\log(k') - \log(k'') = -1.40$. Similarly, for $p = 0.6$, $q = 0.2$, we have $\pi(0.6) = 0.47$, $\pi(0.8) - \pi(0.2) = 0.35$, and $1 - \pi(1 - 0.6) = 0.63$, which likewise implies opposing discontinuities: $\log(k) - \log(k') = 0.53$, and $\log(k') - \log(k'') = -1.17$.

The preceding discussion assumes utility is linear. In the Appendix, we generalize our analysis to the case of concave utility, proving that the $X - k$ schedule continues to exhibit two discontinuities of opposite signs (see Proposition 3). In addition, variations in X that leave ranks unaffected can change the modified equalizing reduction. From Eqs. (7) through (9), it is apparent that k, k', and k'' are proportional to, respectively, du/dX, du/dX', and du/dX''. Assuming concavity, these marginal utilities decline as the payoff rises, which imparts a downward slope to the $X - k$ schedule within each of the three regimes (the "curvature effect"). For instance, with the standard parameterization of utility, $u(X) = X^{\alpha}$, we have $\log(k) = K - (1 - \alpha)\log(X)$ (for some constant K), along with similar expressions for k' and k''. Under EU and PT, we would expect to see a continuous, downward-sloping curve of this form. Notice that, for discrete changes in the value of X, the rank dependence effect reinforces the curvature effect when X crosses Z, but opposes it when X crosses Y. Accordingly, if the rank dependence effect is important, we should observe substantially different changes in the modified equalizing reductions in these two cases.

We implemented our supplemental design with 84 Stanford subjects in the Fall of 2017 and Winter of 2018. Each subject completed 18 modified equalizing reduction tasks. In each task, subjects made a series of decisions between two lotteries. We examined two probability vectors, $\{p, q, 1 - p - q\} = \{0.4, 0.3, 0.3\}$ and $\{p, q, 1 - p - q\} = \{0.6, 0.2, 0.2\}$, corresponding to the examples provided above. In each task, $Y = \$36$, $Z = \$18$, $m = \$4$ and k ranged from \$0 to \$9.75 in \$0.25 increments. We used nine values of $X/X'/X''$ for each probability vector, $\{2, 3, 4, 20, 21, 22, 38, 39, 40\}$; notice that X increased by a factor of 20 across all tasks. We vary X over this wide range to ensure that the differences are salient to the subjects, and we verify this salience by demonstrating below that subjects respond to the variation in X as one would predict in light of the utility function's curvature. Appendix Figure A5 of the Supplemental Material (Bernheim and Sprenger (2020)) exhibits a sample task.[56] As before, subjects also completed the seven Prospect

---

[56]Note that subjects had no opportunity to choose values of k that would have changed payoff ranks within any task.

Theory elicitation tasks of Tversky and Kahneman (1992) which we used to generate one set of benchmark predictions for behavior in the new environment. Seventy-two of 84 subjects (85.7%) completed all tasks without multiple switching.[57]

Figure 5 presents results from this supplemental experiment. As before, the elicited certainty equivalents reproduce the S-shaped pattern commonly associated with probability weighting. Based on these data, we estimate $\alpha = 0.982$ (clustered s.e. $= 0.024$) and $\gamma = 0.766$ (0.021), and reject at all conventional levels the null hypothesis of a linear weighting function, $\gamma = 1$, $F_{1,71} = 121.9$ ($p < 0.01$). Using these parameters, we generate a benchmark by predicting modified equalizing reductions.[58] Panel B shows the predicted pattern for each probability vector. The predicted $X - k$ schedule exhibits striking non-monotonicity.[59] Additionally, the curvature effect discussed above leads to decreasing values of k within the regions $X > Y > Z$, $Y > X' > Z$, and $Y > Z > X''$.

To address potential concerns about failures of procedural invariance, we also generate alternative benchmarks analogous to those presented in Section 4.2.2. Specifically, we exploit variation in probabilities within each of the three regimes to estimate the curvature parameter of the probability weighting function,[60] and then use that estimate to predict changes in the modified equalizing reduction. Appendix Figure A6 of the Supplemental Material (Bernheim and Sprenger (2020)) contrasts the benchmarks generated from these estimates with actual behavior. Our conclusions are robust with respect to these alternative benchmarks.

In contrast to the predictions based on the calibrated CPT models, the average value of the actual modified equalizing reduction shows no hint of non-monotonicity in X. Though subjects respond strongly to differences in probabilities, there is no indication that payoff ranks matter, either as CPT predicts or otherwise. The observed declines in equalizing reductions are smooth and monotonic, and the overall patterns are qualitatively consistent with the curvature effect noted above. At the individual level, only 4 of 72 subjects exhibit the non-monotonic pattern predicted by CPT for both probability vectors.[61]

Analogously to Section 4.2.3, we can ask whether the PT, CPT, or EU (linear) formulation of probability weighting provides the best explanation for our data on modified equalizing reductions. Estimating again Eq. (6) with standard errors clustered at the individual level yields $\alpha = 0.948$ (0.017), $\gamma = 0.809$ (0.025), and $\delta = -0.039$ (0.062). The probability weighting function continues to exhibit substantial curvature, but the rank-dependent CPT version receives approximately zero weight. Interestingly, the estimated probability weighting parameter from this specification is quite similar to that obtained

---

[57]Incorporating multiple switching subjects by removing only their multiple switching observations yields virtually identical results.

[58]For these predictions, we assume a reference point of $r = 0$ and calculate the value k that solves for indifference between $L = (\{p, q, 1 - p - q\}, \{X, Y, Z\})$ and $L_e = (\{p, q, 1 - p - q\}, \{X + m, Y - k, Z - k\})$ at the estimated CPT parameters.

[59]The $X - k$ schedule does not feature sharp discontinuities because of our use of non-infinitesimal values of m and k. There exist two transitional regions where either: (1) $Z > X$ and $Z - k < X + m$, or (2) $Y > X$ and $Y - k < X + m$.

[60]Using the functional forms of Tversky and Kahneman (1992) and an additive disturbance, we estimate Eqs. (7), (8), and (9) using nonlinear least squares with standard errors clustered at the individual level. The estimates for γ are 0.785 (0.032), 0.863 (0.024), and 0.741 (0.042) for $X > Y > Z$, $Y > X' > Z$, and $Y > Z > X''$, respectively. The estimates for α are 0.844 (0.074), 0.437 (0.364), and 1.066 (0.021) for $X > Y > Z$, $Y > X' > Z$, and $Y > Z > X''$, respectively.

[61]For the purpose of this calculation, we treat the subject as conforming to the predicted pattern for a particular probability vector if their average values of k and k'' are at least 25 cents greater than their average value of k'.
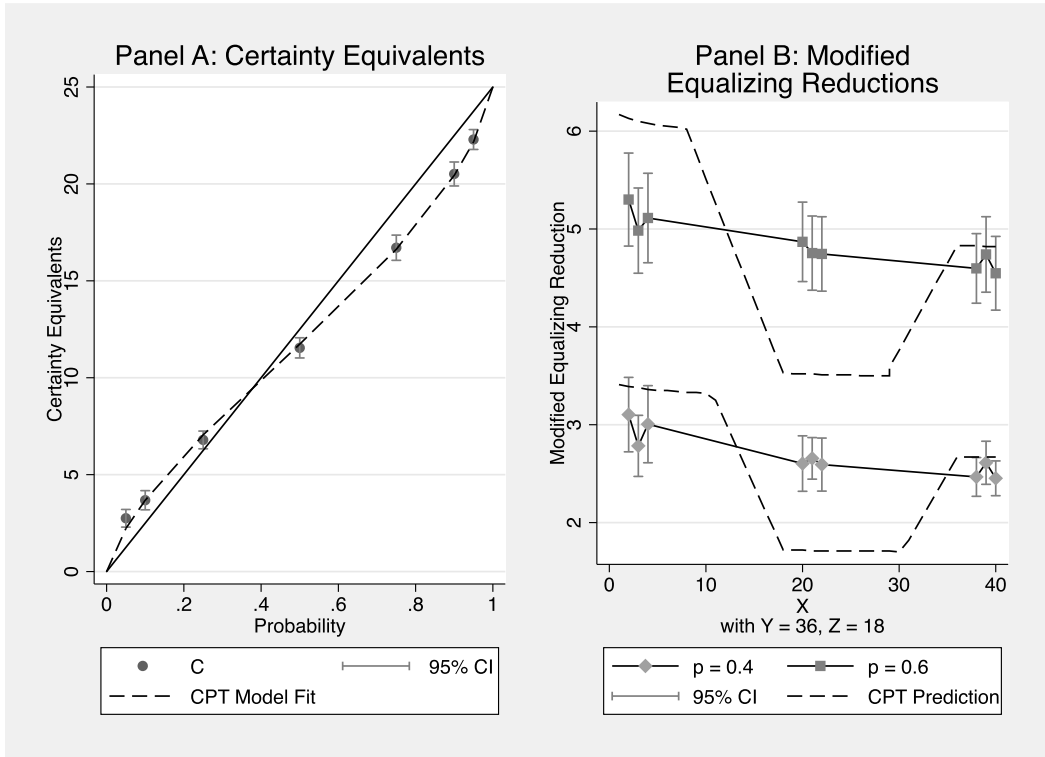
FIGURE 5.—Certainty equivalents and modified equalizing reductions. *Notes*: Panel A: Mean behavior for C estimated from interval regression (Stewart (1983)) of experimental response on indicators for probability vectors. Standard errors clustered at individual level to provide 95% confidence interval. Dashed line corresponds to predicted CPT behavior with $\alpha = 0.982$ (s.e. $= 0.024$) and $\gamma = 0.766$ (0.021); standard errors clustered at individual level. Panel B: Mean behavior for modified equalizing reduction estimated from interval regression of experimental response on indicators for probability vectors interacted with indicators for value of X. Standard errors clustered at individual level to provide 95% confidence interval. Dashed line corresponds to predicted equalizing reductions for CPT decisionmaker with risk preference parameters $\alpha = 0.982$ and $\gamma = 0.766$.

from certainty equivalents for the same subjects, which suggests that there is no meaningful failure of procedural invariance; that is, the modified equalizing reduction design does not, itself, dramatically alter the shape of the probability weighting function.[62]

Figure 5 confirms that the evidence of rank independence in our main experiment is not an artifact of cancellation. Subjects attend to payoff values (including X) and probability vectors when determining their equalizing reductions, but there is no indication that the rank of outcomes per se influences behavior.

5.4. *Random Choice*

Finally, it is important to rule out the possibility that we detect no rank dependence because our subjects ignore the parameters of their decision tasks (either in general, or X in

---

[62]Additionally, the similarity in results for estimated probability weighting parameters between our initial and modified designs suggests that the two samples do not differ meaningfully along this critical dimension.

particular) and make their choices more or less randomly. This hypothesis is inconsistent with the following findings.

First, as we have noted, equalizing reductions are highly responsive to variations in the probability vector. Shifting 20% of the probability mass from the high-payoff outcome to the low-payoff outcome, thereby reducing p from 0.6 to 0.4, cuts the average equalizing reduction in half (see Table III and Figure 3, Panel B). Moving another 30% of the probability mass from the high payoff to the low payoff, thereby reducing p from 0.4 to 0.1, halves the equalizing reduction again. According to our parameterized CPT models, the induced changes in payoff ranks should be just as consequential, in that they ought to induce similar changes in probability weights, yet they have no effects. More generally, the parameterized CPT model generally predicts behavior with reasonable accuracy out of sample with the exception of variations involving rank dependence.

Second, our analysis of subject-level results (Section 5.1) shows that estimates of the percentage change in k induced by a change in payoff ranks cluster tightly around zero. In other words, as a general matter, changing payoff ranks induces very little change in the equalizing reduction, which means we rarely observe subjects who exhibit economically meaningful rank dependence. In contrast, if choices were truly random, we would often observe economically meaningful rank dependence simply by chance: the distributions for the change in equalizing reduction resulting from a change in ranks would still be centered around zero, but they would be far more diffuse (see Appendix D of the Supplemental Material (Bernheim and Sprenger (2020)) for simulations). Thus, the consistent absence of rank dependence, documented in Panel A of Figure 4, rejects the random choice hypothesis.

Third, our analysis of the cancellation hypothesis (Section 5.3) shows clearly that equalizing reductions respond strongly to changes in X precisely in the way theory predicts they should based on utility curvature. (See in particular Panel B of Figure 5, in which the relationships between actual equalizing reductions and X are robustly downward-sloping.) In light of these results, one cannot plausibly claim that our experiments induce subjects to ignore variations in the value of X.

6. EVENT SPLITTING AND VIOLATIONS OF DOMINANCE

Our analysis casts doubt on the empirical validity of the assumption about rank-dependent probability weighting that lies at the core of CPT. What type of model should behavioral economists consider in its place? One possibility is that PT is correct, in which case people should actually exhibit the implied violations of first-order stochastic dominance that motivated the formulation of CPT in the first place. After we obtained our main results, we fielded an additional treatment designed to investigate that hypothesis.

We conducted the follow-up experiment at Stanford University and UC San Diego during the Spring and Fall of 2015. A total of 214 subjects completed the experiment, and 182 exhibited no instances of multiple switching.[63] We elicited certainty equivalents first for binary lotteries, such as a 40% chance of receiving $30 and a 60% chance of receiving $20. In one task, the lower payoff was more likely, and in another, it was less likely. Then, we elicited certainty equivalents for related "split-event" lotteries, which we created by splitting the more likely event in each of the binary lotteries. For the preceding example,

---

[63]We recruited 126 at UC San Diego and 88 at Stanford. Those who exhibited no instances of multiple switching include 99 subjects from UC San Diego and 83 from Stanford. Subjects from the two locations made qualitatively similar choices.

a "split-event" lottery would take the following form: a 40% chance of receiving $30, a 30% chance of receiving $20 + ϵ, and a 30% chance of receiving $20 − ϵ. Across tasks, ϵ took on the following values: $0.50, $1, $2, and $3. Note that this split-event design is also immune to critiques involving the cancellation hypothesis, as subjects choose between lotteries and sure amounts rather than between two lotteries with common elements. Subjects also completed a series of seven Prospect Theory elicitation tasks involving binary lotteries, as before.

Most empirical parameterizations of PT imply that a 60% probability receives substantially less than twice the weight of a 30% probability. For lotteries with outcomes above the reference point, splitting an event occurring with 60% probability into two similar events, each occurring with 30% probability, should therefore increase certainty equivalents dramatically. It follows that PT predicts a sharp downward discontinuity at $\epsilon = 0$. As noted in Section 1, such discontinuities imply violations of dominance. In contrast, CPT predicts that responses vary smoothly with ϵ and thereby avoid dominance violations. Importantly, under CPT the nature of the outcome that is split influences the direction of response. Splitting a good event occurring with 60% probability into two similar events with 30% probability creates a lottery with greater proportionate weight on the better outcome, increasing valuations.[64] Splitting a bad event occurring with 60% probability into two similar events with 30% probability creates a lottery with greater proportionate weight on the worse outcome, decreasing valuations.[65] Hence, CPT can generate an asymmetric prediction in this setting. This prediction differs in turn from that of EU, which predicts weakly decreasing (increasing) valuations for all types of split events provided utility is concave (convex).[66]

Figure 6 presents our findings. As shown in Panel A, the Prospect Theory elicitation tasks exhibit the hallmark pattern of probability weighting. In Panel B, we use the fitted probability weighting to predict the effect of event-splitting on certainty equivalents.[67] The predicted certainty equivalents vary smoothly but asymmetrically with ϵ under CPT, and feature a sharp increase at zero under PT. In contrast, the means of the actual certainty

---

[64]Consider a 60% probability of $30, 40% probability of $20. Split the 60% probability of $30 into two 30% probabilities of $30 ± ϵ. The split lottery will yield higher utility if

$$\pi(0.3)u(30+\epsilon|r)+\big[\pi(0.6)-\pi(0.3)\big]u(30-\epsilon|r)+\big[1-\pi(0.6)\big]u(20|r) > \pi(0.6)u(30|r)+\big[1-\pi(0.6)\big]u(20|r).$$

Under the approximation $u(30+\epsilon|r)-u(30|r)=u'(30|r)\epsilon$ and $u(30|r)-u(30-\epsilon|r)=u'(30|r)\epsilon$, this expression reduces to

$$2\pi(0.3) > \pi(0.6),$$

a property generally satisfied by S-shaped weighting functions used in the literature.

[65]Consider a 60% probability of $20, 40% probability of $30. Split the 60% probability of $20 into two 30% probabilities of $20 ± ϵ. The split lottery will yield lower utility if

$$\pi(0.4)u(30|r)+\big[\pi(0.7)-\pi(0.4)\big]u(20+\epsilon|r)+\big[1-\pi(0.7)\big]u(20-\epsilon|r) < \pi(0.4)u(30|r)+\big[1-\pi(0.4)\big]u(20|r).$$

Under the approximation $u(20+\epsilon|r)-u(20|r)=u'(20|r)\epsilon$ and $u(20|r)-u(20-\epsilon|r)=u'(20|r)\epsilon$, this expression reduces to

$$2\pi(0.7) < 1 + \pi(0.4),$$

a property generally satisfied by S-shaped weighting functions used in the literature.

[66]Another valuable aspect of this design is that certainty equivalents are used for elicitation of both the CPT parameters and split-event valuations. Hence, any elicitation issues related to certainty effects plausibly affect both elements of the design.

[67]Using the same estimation strategy as before, we obtain the following parameter values: $\alpha = 0.975$ (clustered s.e. $= 0.019$) and $\gamma = 0.671$ (0.013).
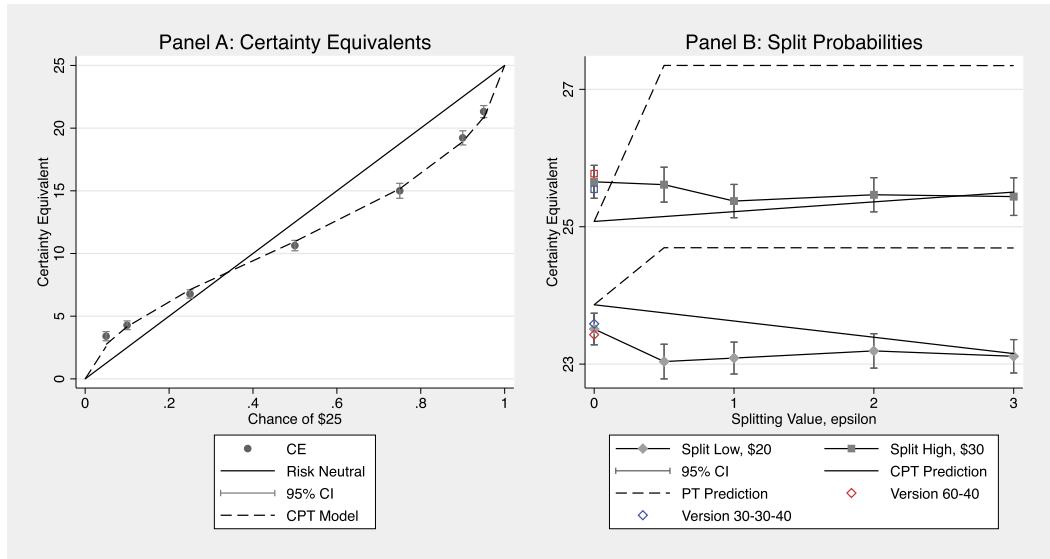
FIGURE 6.—Certainty equivalents and split probabilities. *Notes*: Panel A: Average certainty equivalent, C, estimated from an interval regression (Stewart (1983)) of elicited certainty equivalents on the probability of winning $25. Confidence intervals based on standard errors clustered at the subject level. Dashed line corresponds to CPT predictions with $\alpha = 0.975$ (s.e. $= 0.019$) and $\gamma = 0.671$ (0.013); standard errors clustered at individual level. Panel B: Average certainty equivalent, C, estimated from interval regressions of elicited certainty equivalents on the value of ϵ. Confidence intervals based on standard errors clustered at subject level. For $\epsilon = 0$, separate averages reported based on presentation style (either 60%–40% or 30%–30%–40% with identical payoffs for the first two events). Dashed line corresponds to prediction assuming PT at aggregate parameter values. Solid line corresponds to prediction assuming CPT at aggregate parameter values.

equivalents both *decrease* sharply when we split the low-outcome event, and then level off. Specifically, moving from $\epsilon = 0$ to $\epsilon = 0.5$ reduces the average certainty equivalent by $0.47 (clusted s.e. $= 0.11$), ($z = 4.16$, $p < 0.01$).[68] We observe a qualitatively similar though somewhat muted pattern when we split the high-outcome event.[69]

The findings in Panel B of Figure 6 are inconsistent with PT, CPT, and EU, and therefore call for an alternative explanation. A unified theory must account simultaneously for all three patterns discussed in this study: (1) robust evidence of probability weighting based on behavioral responses to variations in probabilities, (2) rank independence in equalizing reductions, and (3) the sharp drop in certainty equivalents that results from splitting an event. EU is inconsistent with (1) and (3). In light of (1), CPT is inconsis-

---

[68]Test statistics are derived from interval regressions (Stewart (1983)) of certainty equivalents on indicators for ϵ. Standard errors are clustered at the subject level.

[69]The apparent distaste for splitting an event, which turns a binary lottery into a trinary one, is not an artifact of differences in presentation. As shown in Figure 6, certainty equivalents are essentially the same regardless of whether we present a binary lottery as a 60%–40% gamble, or as a 30%–30%–40% gamble with identical payoffs for the first two events. Recall that we employed a single presentation of binary lotteries for each subject, so this finding reflects between-subject comparisons. One notable phenomenon discussed by Birnbaum (2008) is a sensitivity of lottery valuations to description of events. Describing two lotteries as ({0.85, 0.10, 0.05}; {100, 50, 50}) and ({0.85, 0.10, 0.05}; {100, 100, 7}) leads to different hypothetical binary choice patterns than ({0.85, 0.15}; {100, 50}) and ({0.95, 0.05}; {100, 7}). This failure of 'coalescing' is one of a number of violations reviewed by Birnbaum (2008) and is clearly at odds with CPT. Interestingly, in incentivized tasks, we do not see the failure of coalescing noted by Birnbaum (2008) for hypothetical choice.

tent with (2) and (3), and PT is inconsistent with (3). Alternatives to CPT that likewise incorporate rank-dependent probability weighting are also rejected.

One possibility is to reformulate PT probability weighting in terms of normalized weights—that is, $\frac{\pi(p_k)}{\pi(p_1)+\cdots+\pi(p_K)}$ rather than simply $\pi(p_k)$. That model accounts for the downward discontinuity observed when splitting the low-payoff event, but implies an upward discontinuity when splitting the high-payoff event, which we do not observe. It also precludes the theory from accounting for the well-known *certainty effect*, which Allais famously described as a "preference for security in the neighborhood of certainty" (Allais (2008)), for which there is considerable evidence (see, e.g., Camerer (1992), Harless and Camerer (1994)).

A second and more promising possibility is that the observed behavior reflects a combination of standard PT and a form of complexity aversion: people may prefer lotteries with fewer outcomes because they are easier to understand.[70] One can think of the certainty effect as a special case of this more general phenomenon. Under this hypothesis, sufficiently small values of ϵ lead subjects to see a lottery as binary rather than trinary, which discretely increases their certainty equivalents. Because the PT probability-weighting effect works in the opposite direction, the composite effect of reducing ϵ to zero can be positive or negative.

The aforementioned theory can, in principle, account for the somewhat different patterns observed when we split the high-payoff and low-payoff events. Higher stakes may reduce the magnitude of the complexity effect by making subjects more willing to ponder their prospects. Additionally, if subjects think of discrepancies between payoffs in relative terms, a higher payoff will tend to enlarge the "neighborhood" within which they implicitly "merge" events yielding similar outcomes. Both implications are consistent with the pattern observed in Panel B of Figure 6, but we acknowledge that this is an ex post rationalization for a somewhat limited collection of results rather than a bona fide and systematic test of the theory.

It is important to acknowledge that complexity effects, like PT, potentially give rise to violations of dominance. However, the implied violations are explicable, because they involve the selection of options that are easier to understand. To illustrate, in our experiment, splitting a 60% chance of receiving $20 into a 30% chance of receiving $20.50 and a 30% chance of receiving $19.50 decreases the certainty equivalent by $0.47. Assuming the addition of $0.50 to the $19.50 payoff increases the certainty equivalent by less than $0.47, we plainly have a violation of dominance: the individual attaches more value to the original lottery than to the revised split-event lottery even though the latter dominates the former.[71]

7. CONCLUSION

The main lessons of this study concern the empirical validity of rank-dependent probability weighting, an assumption that lies at the core of Cumulative Prospect Theory (CPT).

---

[70]We are certainly not the first to discuss complexity as a driver of choice. A body of relevant work has highlighted the effects of complexity and the number of decision options on choice (see, e.g., Iyengar and Lepper (2000), Iyengar, Jiang, and Huberman (2003), Iyengar and Kamenica (2010), Sonsino and Mandelbaum (2001)). Additionally, Stodder (1997) drew links between complexity and Allais-style behavior in a theory involving costly calculation.

[71]Viewing complexity as a rationale for indirect violations of dominance between binary and trinary lotteries may also yield insights concerning other indirect dominance violations between binary and degenerate lotteries, such as those documented by Gneezy, List, and Wu (2006) and Andreoni and Sprenger (2011).

We conduct an experimental investigation involving the elicitation of 'equalizing reductions.' A meaningful degree of rank dependence implies that these equalizing reductions should change discontinuously when a shift in the payoff associated with an unrelated realization alters the ranking of payoffs, but should otherwise remain constant. Indeed, we have shown that the percentage change in the equalizing reduction precisely (and nonparametrically) measures the percentage change in the relative decision weights applied to the outcomes. Based on standard parameterizations of CPT as well as our own estimates for our subject pool, these discontinuities should be substantial. And yet we find no evidence of the predicted pattern at either the aggregate or individual level, based on both within- and between-subjects analysis. Thus, we conclude that an empirical foundation for rank-dependent probability weighting is absent. We demonstrate that models with reference distributions (notably Koszegi and Rabin (2006, 2007)) have similar implications, and hence we falsify them as well. Other features of our results are contrary to theories of payoff-sensitive reference points (as in Bell (1985), Loomes and Sugden (1986)).

Our findings pose serious challenges for future research on choice under uncertainty. If CPT has taken the PT agenda in the wrong direction by promoting the assumption of rank-dependent probability weighting (which several other recent theories embrace), how can we reconcile PT with the presumed absence of implied dominance violations? As we demonstrate in a supplemental experiment involving the effects of "event splitting," those violations indeed do not arise. Instead, we observe a different type of violation that is at odds not only with PT, but also with CPT and EU. An important direction for future research is to explore tractable explanations for this finding that do not involve rank dependence. One potential explanation involves a general form of aversion to complexity. That notion rationalizes the behavioral patterns discussed in this study, but whether it survives rigorous and systematic testing remains to be seen.

A potential defense of rank dependence is that it is an assumption of convenience: it renders PT more tractable for applications, but is rarely required to account for observed behavior, including anomalies. This argument strikes us as odd. First, rank dependence undeniably plays a critical role in a number of applications, and is responsible for generating particular results. We noted several examples in footnote 9. If there is no empirical support for rank dependence, then plainly those applications require reexamination. Second, in other applications, either (1) it is known that rank dependence is inessential for generating the results of interest, or (2) it is not known. Case (1) cannot arise unless the researcher has conducted the analysis without assuming rank dependence and found it tractable, in which case there is no reason to employ rank dependence as an assumption of convenience. In case (2), the possibility remains that the result of interest may be an artifact of an assumption that lacks empirical validity.

Additional research is plainly required to test the robustness of our findings. In twenty-five years since the publication of Tversky and Kahneman (1992), there have been numerous studies of probability weighting for binary lotteries, but decidedly few systematic studies of rank dependence. Further work is needed to resolve the relevance of rank dependence in a variety of experimental and naturally occurring contexts. If the findings of the current study replicate, the profession could either discard CPT and other models that employ or (as in the case of Koszegi and Rabin's theory of reference distributions) imply rank dependence, or augment the CPT model with additional degrees of freedom to accommodate the absence of rank dependence in the data. If the former path is followed, valuable research could focus on the importance of complexity in decisionmaking under uncertainty. If the latter path is followed, the model will need to be augmented with novel

degrees of freedom, as preexisting hypotheses concerning heuristic cancellation and the nature of reference dependence cannot rationalize the data.

APPENDIX: PROOFS AND PROPOSITION 3

Proofs of Propositions 1 and 2

PROOF OF PROPOSITION 1: Noting that $\overline{k}$ and $\underline{k}$ converge to zero as $m \to 0$, we apply l'Hospital's rule (along with continuity of the log function) to obtain

$$\lim_{m \to 0} \Delta \log(k) = \Delta \log\left(\lim_{m \to 0} \frac{dk}{dm}\right).$$

Fixing any value of $X \neq Y, Z$ and choosing m sufficiently small so that the orderings of X, Y, Z and $X, Y + m, Z - k$ (and hence the decision weights w_X, w_Y, and w_Z) are the same, we see from Eq. (2) that

$$\frac{dk}{dm} = \frac{w_Y}{w_Z}\left(\frac{u'(Y + m)}{u'(Z - k)}\right).$$

Because u is continuously differentiable at Y and Z, we have

$$\lim_{m \to 0} \frac{dk}{dm} = \frac{w_Y}{w_Z}\left(\frac{u'(Y)}{u'(Z)}\right).$$

But then

$$\Delta \log\left(\lim_{m \to 0} \frac{dk}{dm}\right) = \log\left(\frac{\overline{w}_Y}{\overline{w}_Z}\left(\frac{u'(Y)}{u'(Z)}\right)\right) - \log\left(\frac{\underline{w}_Y}{\underline{w}_Z}\left(\frac{u'(Y)}{u'(Z)}\right)\right)$$

$$= \Delta \log\left(\frac{w_Y}{w_Z}\right). \qquad \qquad Q.E.D.$$

PROOF OF PROPOSITION 2: As in the proof of Proposition 1,

$$\lim_{m \to 0} \Delta \log(k) = \Delta \log\left(\lim_{m \to 0} \frac{dk}{dm}\right).$$

Choosing m sufficiently small so that the orderings of the payoffs (and hence the decision weights w_X, w_Y, and w_Z) are the same, we implicitly differentiate the indifference condition, Eq. (2), using u_1 and u_2 to stand for the partial derivates of the utility function with respect to the payoff and the reference point, respectively, to obtain

$$\frac{dk}{dm} = \left(w_Y\left[u_1(Y + m, r) + \left(u_2(Y + m, R) - u_2(Y, R)\right)\frac{dR}{dm}\right]\right.$$
$$\left. + w_Z\left(u_2(Z - k, R) - u_2(Z, R)\right)\frac{dR}{dm}\right)$$
$$\left/\left(w_Z\left[u_1(Z - k, r) + \left(u_2(Z, R) - u_2(Z - k, R)\right)\frac{dR}{dk}\right]\right.\right.$$
$$\left. + w_Y\left(u_2(Y, R) - u_2(Y + m, R)\right)\frac{dR}{dk}\right).$$

Given the (local) continuous differentiability of u and R, it follows that

$$\lim_{m \to 0} \frac{dk}{dm} = \frac{w_Y}{w_Z}\left(\frac{u_1(Y, R)}{u_1(Z, R)}\right).$$

But then, for fixed n,

$$\Delta \log\left(\lim_{m \to 0} \frac{dk}{dm}\right) = \log\left(\frac{\overline{w}_Y}{\overline{w}_Z}\left(\frac{u'(Y, \overline{R}^n)}{u'(Z, \overline{R}^n)}\right)\right) - \log\left(\frac{\underline{w}_Y}{\underline{w}_Z}\left(\frac{u'(Y, \underline{R}^n)}{u'(Z, \underline{R}^n)}\right)\right)$$

$$= \Delta \log\left(\frac{w_Y}{w_Z}\right) + \log\left(\left(\frac{u'(Y, \overline{R}^n)}{u'(Y, \underline{R}^n)}\right)\left(\frac{u'(Z, \underline{R}^n)}{u'(Z, \overline{R}^n)}\right)\right).$$

Using the local continuity of R and the local continuous differentiability of u, we have

$$\lim_{n \to \infty} \log\left(\left(\frac{u'(Y, \overline{R}^n)}{u'(Y, \underline{R}^n)}\right)\left(\frac{u'(Z, \underline{R}^n)}{u'(Z, \overline{R}^n)}\right)\right) = 0.$$

The desired conclusion follows immediately.   *Q.E.D.*

Proposition 3: Opposing Discontinuities for Modified Equalizing Reductions

Define the (marginal) modified equalizing reduction as $e = \frac{dk}{dm}|_{m=0}$. In this appendix section, we establish that e exhibits two discontinuities in X, of opposite signs, at $X = Y$ and $X = Z$.

PROPOSITION 3: *Suppose u is continuously differentiable and (weakly) concave, $\pi(p) > \pi(p+q) - \pi(q)$, and $1 - \pi(1-p) > \pi(p+q) - \pi(q)$.*
  (i) *Consider two infinite sequences $X_n^1 \downarrow Y$ and $X_n^2 \uparrow Y$. Then $\lim_{n \to \infty} \frac{e_n^1}{e_n^2} > 1$.*
  (ii) *Consider two infinite sequences $X_n^3 \downarrow Z$ and $X_n^4 \uparrow Z$. Then $\lim_{n \to \infty} \frac{e_n^3}{e_n^4} < 1$.*

PROOF: We begin with part (i). Define

$$J(\theta) = \left(\frac{\pi(q)}{\pi(p+q) - \pi(q)}\right)\frac{du}{dY} + \left(\frac{1 - \pi(p+q)}{\pi(p+q) - \pi(q)}\right)\theta,$$

$$K(\theta) = \left(\frac{\pi(p+q) - \pi(p)}{\pi(p)}\right)\frac{du}{dY} + \left(\frac{1 - \pi(p+q)}{\pi(p)}\right)\theta,$$

and

$$D_H(\theta) \equiv \frac{J(\theta)}{K(\theta)}.$$

Because u is continuously differentiable (so that, in the limit, we can cancel terms involving $\frac{du}{dX_n^1}$ and $\frac{du}{dX_n^2}$), we have $D_H(\frac{du}{dZ}) \equiv \lim_{n \to \infty} \frac{e_n^1}{e_n^2}$. Notice that $D_H(\frac{du}{dY}) > 0$, and that $\frac{du}{dY} \leq \frac{du}{dZ}$ (due to the concavity of u). If $\frac{du}{dY} = \frac{du}{dZ}$, then the conclusion is immediate, so we will assume in what follows that $\frac{du}{dY} < \frac{du}{dZ}$.

Suppose contrary to the claim that $D_H(\frac{du}{dZ}) \leqq 1$. Then, because D_H is continuous in θ on the pertinent domain, there exists some value $\theta^H \in (\frac{du}{dY}, \frac{du}{dZ}]$ for which $D_H(\theta^H) = 1$ and $D_H(\theta) > 1$ for all $\theta \in [\frac{du}{dY}, \theta^H)$. Moreover, because D_H is continuously differentiable in θ on the pertinent domain, we must also have $D'_H(\theta^H) \leq 0$. Observe that

$$D'_H(\theta) = \frac{1 - \pi(p+q)}{(K(\theta))^2} \left[K(\theta) \left(\frac{1}{\pi(p+q) - \pi(q)} \right) - J(\theta) \left(\frac{1}{\pi(p)} \right) \right].$$

Noting that $J(\theta^H) = K(\theta^H)$, we have

$$D'_H(\theta^H) = \frac{1 - \pi(p+q)}{K(\theta^H)} \left[\left(\frac{1}{\pi(p+q) - \pi(q)} \right) - \left(\frac{1}{\pi(p)} \right) \right] > 0,$$

which is a contradiction.

Now turn to part (ii). Define

$$M(\theta) = \left(\frac{\pi(q)}{1 - \pi(1-p)} \right) \theta + \left(\frac{\pi(1-p) - \pi(q)}{1 - \pi(1-p)} \right) \frac{du}{dZ},$$

$$N(\theta) = \left(\frac{\pi(q)}{\pi(p+q) - \pi(q)} \right) \theta + \left(\frac{1 - \pi(p+q)}{\pi(p+q) - \pi(q)} \right) \frac{du}{dZ},$$

and

$$D_L(\theta) \equiv \frac{M(\theta)}{N(\theta)}.$$

Because u is continuously differentiable (so that, in the limit, we can cancel terms involving $\frac{du}{dX_n^3}$ and $\frac{du}{dX_n^4}$), we have $D_L(\frac{du}{dY}) \equiv \lim_{n \to \infty} \frac{e_n^3}{e_n^4}$. Notice that $D_L(\frac{du}{dZ}) < 0$. Once again, concavity of u ensures that $\frac{du}{dY} \leq \frac{du}{dZ}$, and if $\frac{du}{dY} = \frac{du}{dZ}$, the conclusion is immediate, so we will assume in what follows that $\frac{du}{dY} < \frac{du}{dZ}$.

Suppose contrary to the claim that $D_L(\frac{du}{dY}) \geq 1$. Then, because D_L is continuous in θ on the pertinent domain, there exists some value $\theta^L \in [\frac{du}{dY}, \frac{du}{dZ})$ for which $D_L(\theta^L) = 1$ and $D_L(\theta) < 1$ for all $\theta \in (\theta^L, \frac{du}{dZ}]$. Moreover, because D_L is continuously differentiable in θ on the pertinent domain, we must also have $D'_L(\theta^L) \leq 0$. Observe that

$$D'_L(\theta) = \frac{\pi(q)}{(N(\theta))^2} \left[N(\theta) \left(\frac{1}{\pi(p+q) - \pi(q)} \right) - M(\theta) \left(\frac{1}{1 - \pi(1-p)} \right) \right].$$

Noting that $M(\theta^L) = N(\theta^L)$, we have

$$D'_H(\theta^L) = \frac{\pi(q)}{N(\theta^L)} \left[\left(\frac{1}{\pi(p+q) - \pi(q)} \right) - \left(\frac{1}{1 - \pi(1-p)} \right) \right] > 0,$$

which is a contradiction. $\hspace{4cm}$ *Q.E.D.*

REFERENCES

ABDELLAOUI, M. (2000): "Parameter-Free Elicitation of Utility and Probability Weighting Functions," *Management Science*, 46 (11), 1497–1512. [1368]

ALLAIS, M. (1953): "Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Americaine," *Econometrica*, 21 (4), 503–546. [1363]

——— (2008): "Allais Paradox," in *The New Palgrave Dictionary of Economics* (Second Ed.), ed. by S. N. Durlauf and L. E. Blume. Palgrave Macmillan. [1402]

ANDREONI, J., AND C. SPRENGER (2011): "Uncertainty Equivalents: Testing the Limits of the Independence Axiom," Working Paper. [1402]

——— (2012): "Risk Preferences Are not Time Preferences," *American Economic Review*, 102 (7), 3357–3376. [1368]

BARBERIS, N. (2018): "Psychology-Based Models of Asset Prices and Trading Volume," in *Handbook of Behavioral Economics: Applications and Foundations 1*, Vol. 1, ed. by B. D. Bernheim, S. DellaVigna, and D. Laibson. Elsevier, 79–175. [1390]

BARBERIS, N., AND M. HUANG (2008): "Stocks As Lotteries: The Implications of Probability Weighting for Security Prices," *American Economic Review*, 98 (5), 2066–2100. [1388]

BARBERIS, N., A. MUKHERJEE, AND B. WANG (2016): "Prospect Theory and Stock Returns: An Empirical Test," *Review of Financial Studies*, 29 (11), 3068–3107. [1368,1388]

BARSEGHYAN, L., F. MOLINARI, T. O'DONOGHUE, AND J. C. TEITELBAUM (2015): "Bracketing and Risk Preferences: Identification and Estimation in Field Data," Working Paper. [1368,1375]

BELL, D. E. (1985): "Disappointment in Decision Making Under Uncertainty," *Operations Research*, 33 (1), 1–27. [1366,1374,1386,1403]

BENARTZI, S., AND R. H. THALER (1995): "Myopic Loss Aversion and the Equity Premium Puzzle," *Quarterly Journal of Economics*, 110, 79–98. [1388]

BERNHEIM, B. D., AND C. SPRENGER (2020): "Supplement to 'On the Empirical Validity of Cumulative Prospect Theory: Experimental Evidence of Rank-Independent Probability Weighting'," *Econometrica Supplemental Material*, 88, https://doi.org/10.3982/ECTA16646. [1367,1368,1374-1377,1379,1381-1385,1389, 1391,1394-1397,1399]

BIRNBAUM, M. H. (2008): "New Paradoxes of Risky Decision Making," *Psychological Review*, 115 (2), 463–501. [1367,1368,1376,1401]

BLEICHRODT, H., AND J. L. PINTO (2000): "A Parameter-Free Elicitation of the Probability Weighting Function in Medical Decision Analysis," *Management Science*, 46 (11), 1485–1496. [1368]

BOOIJ, A. S., AND G. VAN DE KUILEN (2009): "A Parameter-Free Analysis of the Utility of Money for the General Population Under Prospect Theory," *Journal of Economic Psychology*, 30, 651–666. [1368]

BOOIJ, A. S., B. M. S. VAN PRAAG, AND G. VAN DE KUILEN (2010): "A Parametric Analysis of Prospect Theory's Functionals for the General Population," *Theory and Decision*, 68, 115–148. [1368]

CAMERER, C. F. (1992): "Recent Tests of Generalizations of Expected Utility Theory," in *Utility: Theories, Measurement, and Applications*, ed. by W. Edwards. Norwell, MA: Kluwer, 207–251. [1402]

CAMERER, C. F., AND T.-H. HO (1994): "Violations of the Betweenness Axiom and Nonlinearity in Probability," *Journal of Risk and Uncertainty*, 8 (2), 167–196. [1364,1369]

CHATEAUNEUF, A., J. EICHBERGER, AND S. GRANT (2007): "Choice Under Uncertainty With the Best and Worst in Mind: Neo-Additive Capacities," *Journal of Economic Theory*, 137, 538–567. [1370]

CUBITT, R. P., C. STARMER, AND R. SUGDEN (1998): "On the Validity of the Random Lottery Incentive System," *Experimental Economics*, 1, 115–131. [1380]

DIECIDUE, E., AND P. P. WAKKER (2001): "On the Intuition of Rank-Dependent Utility," *The Journal of Risk and Uncertainty*, 23 (3), 281–298. [1368]

EPPER, T., AND H. FEHR-DUDA (2015): "Balancing on a Budget Line: Comment on Andreoni and Sprenger (2012)'s 'Risk Preferences Are not Time Preferences'," *American Economic Review*, 105 (7), 2261–2271. [1368]

FENNEMA, H., AND P. WAKKER (1996): "A Test of Rank-Dependent Utility in the Context of Ambiguity," *Journal of Risk and Uncertainty*, 13, 19–35. [1367,1376]

——— (1997): "Original and Cumulative Prospect Theory: A Discussion of Empirical Differences," *Journal of Behavioral Decision Making*, 10, 53–64. [1364,1369]

FREDERICK, S. (2005): "Cognitive Reflection and Decision Making," *Journal of Economic Perspectives*, 19 (4), 25–42. [1394]

FRIEDMAN, M., AND L. J. SAVAGE (1948): "The Utility Analysis of Choices Involving Risk," *Journal of Political Economy*, 56 (4), 279–304. [1363]

GNEEZY, U., J. A. LIST, AND G. WU (2006): "The Uncertainty Effect: When a Risky Prospect Is Valued Less Than Its Worst Possible Outcome," *The Quarterly Journal of Economics*, 121 (4), 1283–1309. [1402]

GONZALEZ, R., AND G. WU (1999): "On the Shape of the Probability Weighting Function," *Cognitive Psychology*, 38, 129–166. [1368,1372]

HARLESS, D. W., AND C. F. CAMERER (1994): "The Predictive Utility of Generalized Expected Utility Theories," *Econometrica*, 62 (6), 1251–1289. [1402]

HARRISON, G. W., M. I. LAU, E. E. RUTSTROM, AND M. B. WILLIAMS (2005): "Eliciting Risk and Time Preferences Using Field Experiments: Some Methodological Issues," in *Field Experiments in Economics*, Vol. 10 (Research in Experimental Economics), ed. by J. Carpenter, G. W. Harrison, and J. A. List. Greenwich and London: JAI Press. [1381]

HOLT, C. A. (1986): "Preference Reversals and the Independence Axiom," *The American Economic Review*, 76 (1), 508–515. [1379]

HOLT, C. A., AND S. K. LAURY (2002): "Risk Aversion and Incentive Effects," *The American Economic Review*, 92 (5), 1644–1655. [1381]

IYENGAR, S. S., AND E. KAMENICA (2010): "Choice Proliferation, Simplicity Seeking, and Asset Allocation," *Journal of Public Economics*, 94 (7–8), 530–539. [1402]

IYENGAR, S. S., AND M. R. LEPPER (2000): "When Choice Is Demotivating: Can One Desire Too Much of a Good Thing?" *Journal of Personality and Social Psychology*, 79, 995–1006. [1402]

IYENGAR, S. S., W. JIANG, AND G. HUBERMAN (2003): "How Much Choice Is Too Much: Contributions to 401(k) Retirement Plans," Pension Research Council Working Paper 2003-10. [1402]

JACOBSON, S., AND R. PETRIE (2009): "Learning From Mistakes: What Do Inconsistent Choices Over Risk Tell Us?" *Journal of Risk and Uncertainty*, 38 (2). [1381]

KAHNEMAN, D., AND A. TVERSKY (1979): "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, 47 (2), 263–291. [1363,1364]

KARNI, E., AND Z. SAFRA (1987): ""Preference Reversal" and the Observability of Preferences by Experimental Methods," *Econometrica*, 55 (3), 675–685. [1379]

KOSZEGI, B., AND M. RABIN (2006): "A Model of Reference-Dependent Preferences," *Quarterly Journal of Economics*, 121 (4), 1133–1165. [1366,1374,1375,1403]

——— (2007): "Reference-Dependent Risk Attitudes," *The American Economic Review*, 97 (4), 1047–1073. [1366,1374,1375,1403]

LOOMES, G., AND R. SUGDEN (1986): "Disappointment and Dynamic Consistency in Choice Under Uncertainty," *Review of Economic Studies*, 53 (2), 271–282. [1366,1374,1386,1403]

LOPES, L. L. (1984): "Risk and Distributional Inequality," *Journal of Experimental Psychology: Human Perception and Performance*, 10 (4), 465–485. [1368]

LOPES, L. L., AND G. 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. [1368]

MASATLIOGLU, Y., AND C. RAYMOND (2016): "A Behavioral Analysis of Stochastic Reference Dependence," *American Economic Review*, 106 (9), 2760–2782. [1375]

MEIER, S., AND C. SPRENGER (2010): "Present-Biased Preferences and Credit Card Borrowing," *American Economic Journal—Applied Economics*, 2 (1), 193–210. [1381]

PRELEC, D. (1998): "The Probability Weighting Function," *Econometrica*, 66 (3), 497–527. [1372,1391]

QUIGGIN, J. (1982): "A Theory of Anticipated Utility," *Journal of Economic Behavior and Organization*, 3, 323–343. [1364,1365]

SAMUELSON, P. A. (1952): "Probability, Utility, and the Independence Axiom," *Econometrica*, 20 (4), 670–678. [1363]

SAVAGE, L. J. (1954): *The Foundations of Statistics*. New York: J. Wiley. [1363]

SCHMEIDLER, D. (1989): "Subjective Probability and Expected Utility Without Additivity," *Econometrica*, 57 (3), 571–587. [1376]

SONSINO, D., AND M. MANDELBAUM (2001): "On Preference for Flexibility and Complexity Aversion: Experimental Evidence," *Theory and Decision*, 51 (2–4), 197–216. [1402]

STARMER, C. (2000): "Development in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice Under Risk," *Journal of Economics Literature*, 38 (2). [1389]

STARMER, C., AND R. SUGDEN (1991): "Does the Random-Lottery Incentive System Elicit True Preferences? An Experimental Investigation," *The American Economic Review*, 81 (4), 971–978. [1380]

STEWART, M. B. (1983): "On Least Squares Estimation When the Dependent Variable Is Grouped," *The Review of Economic Studies*, 50 (4), 737–753. [1382,1383,1385,1395,1398,1401]

STODDER, J. (1997): "Complexity Aversion: Simplification in the Herrnstein and Allais Behaviors," *Eastern Economic Journal*, 23 (1), 1–15. [1402]

TANAKA, T., C. CAMERER, AND Q. NGUYEN (2010): "Risk and Time Preferences: Experimental and Household Data From Vietnam," *American Economic Review*, 100 (1), 557–571. [1368]

TVERSKY, A., AND C. R. FOX (1995): "Weighing Risk and Uncertainty," *Psychological Review*, 102 (2), 269–283. [1363,1368,1372,1382]

TVERSKY, A., AND D. KAHNEMAN (1992): "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty*, 5 (4), 297–323. [1363-1365,1368,1372,1374,1379,1382,1383, 1386-1388,1396,1397,1403]

VON NEUMANN, J., AND O. MORGENSTERN (1944): *Theory of Games and Economic Behavior*. Princeton University Press. [1363]

WAKKER, P., I. EREV, AND E. U. WEBER (1994): "Comonotonic Independence: The Critical Test Between Classical and Rank Dependent Utility Theories," *Journal of Risk and Uncertainty*, 9, 195–230. [1367,1376]

WEBER, E. U. (1994): "From Subjective Probabilities to Decision Weights: The Effect of Asymmetric Loss Functions on the Evaluation of Uncertain Outcomes and Events," *Psychological Bulletin*, 115 (2), 228–242. [1368]

WEBER, E. U., AND B. KIRSNER (1997): "Reasons for Rank-Dependent Utility Evaluation," *Journal of Risk and Uncertainty*, 14 (1), 41–61. [1366,1367,1376,1394]

WU, G. (1994): "An Empirical Test of Ordinal Independence," *Journal of Risk and Uncertainty*, 9, 39–60. [1366, 1367,1375,1394]

WU, G., AND R. GONZALEZ (1996): "Curvature of the Probability Weighting Function," *Management Science*, 42 (12), 1676–1690. [1368,1383]

Co-editor Itzhak Gilboa handled this manuscript.