Journal of Economic Methodology, 2017
Vol. 24, No. 2, 150–165, https://doi.org/10.1080/1350178X.2017.1309753

The empirical adequacy of cumulative prospect theory and its implications for normative assessment

Glenn W. Harrison[a,b]* and Don Ross[b,c,d]

[a]*Department of Risk Management and Insurance and Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, Atlanta, USA;* [b]*School of Economics, University of Cape Town, Cape Town, South Africa;* [c]*School of Sociology and Philosophy, University College Cork, Cork, Ireland;* [d]*Center for Economic Analysis of Risk, Robinson College of Business, Georgia State University, Atlanta, USA*

(*Received 18 January 2017; accepted 14 March 2017*)

Much behavioral welfare economics assumes that expected utility theory (EUT) does not accurately describe most human choice under risk. A substantial literature instead evaluates welfare consequences by taking cumulative prospect theory (CPT) as the natural default alternative, at least where description is concerned. We present evidence, based on a review of previous literature and new experimental data, that the most empirically adequate hypothesis about human choice under risk is that it is heterogeneous, and that where EUT does not apply, more choice is characterized by rank-dependent utility models than by CPT. Most of the apparently loss-averse choice behavior results from probability weighting rather than from direct disutility experienced when an outcome is framed as a loss against an idiosyncratic reference point. We then consider implications of this finding for methodological debates about how to model welfare effects of policies, and argue that abandonment of a dogmatic belief in CPT as the correct theory of risk human choice exposes a conceptual error that is widespread in behavioral welfare economics. We provide concluding reflections on second-order, philosophical issues around the grounding of normative commitments in policy-focused economics.

Keywords: behavioral welfare economics; cumulative prospect theory; expected utility theory; rank-dependent utility; well being

1. Introduction

It is lately a cliché in the social and behavioral sciences that effective policy design should be targeted at people as they actually are, rather than as they would be if idealized normative theories were descriptively accurate. In welfare economics this advice is often expressed by suggesting that normative assessment should recognize, in light of results of decades of behavioral experimentation, that people are not expected utility maximizers.[1] Among such critics, Sugden (2004, 2009) is unusual in arguing that welfare assessment should not be carried out on the basis of any model of revealed or constructed preferences at all. Most others take the view, as analyzed directly by Bleichrodt, Pinto, and Wakker (2001), that individual utility functions are most usefully modeled on the assumption that they conform to the cumulative prospect theory (CPT) of Tversky and Kahneman (1992), rather than expected utility theory (EUT).

*Corresponding author. Email: gharrison@gsu.edu

A common strategy in applied behavioral economics has been to gather risky choice data, such as choices between pairs of lotteries, from experimental samples thought to be representative of a potential policy treatment population, and then run 'horse races' between EUT and CPT to determine which theory best estimates the data at a pooled or individual level. Such data are often taken as evidence for whichever of the two models yields the best fitting estimation. Ross (2005, pp. 174–176) argues that this approach is inherently inconclusive because CPT is a complex model with more moving parts than EUT. Specifically, CPT incorporates the hypothesis of idiosyncratic weighting of probabilities and then adds, when an outcome prospect includes what a subject perceives as a loss against a reference point, aversion to such losses: in the notation of Tversky and Kahneman (1992), $\lambda > 1$. Either or both of these features can generate the behavioral patterns that are commonly referred to as 'loss aversion.' In an empirical application of CPT, the location of the reference point must be assumed or, if the experimental design and data allow, estimated.[2] Thus there are more ways in which a body of choice data can conform to CPT than to EUT. Given that all data involve noise and measurement error, EUT's fewer degrees of freedom entail that the horse race approach stacks the odds against it.[3] Still more problematically, the horse race method imports the implicit assumption that all subjects in the sample are best modeled by one theory or the other. However, whenever analysts have employed methods that allow within-sample heterogeneity to be observed, they have found it.

We review these issues in Section 2. For the moment, consider the economist who pronounces CPT the winner of her horse race where descriptive modeling is concerned, and then proceeds to welfare analysis on the basis of this conclusion. She might follow Savage (1954) and regard EUT as the normatively correct model of 'rational' decision, even if her sample, or people in general, do not typically conform to it in their behavior. Then she might identify her subjects' welfare losses by reference to the difference between their expected outcomes when they choose in accordance with the specific parameterization of CPT she has identified and their expected outcomes had they chosen in accordance with EUT.[4] Following Bleichrodt et al. (2001), this allows her to both 'respect' a subject's possible loss aversion and/or probability weighting in the welfare assessment, but also to offer the subject a path to potential welfare improvement through 'de-biasing' of the subject's choice function.

Our objective is to outline a superior methodology for descriptive modeling of risky choice, review the empirical results that have been obtained by use of this methodology, and then offer some philosophical reflections on implications of these results for normative economics.

In Section 2, we review data on incentivized choice under risk with outcomes that subjects plausibly viewed as losses, using analytical methods that avoid the horse race method, and which allow heterogeneity of risk preference structures within samples to be identified. We show that these data do not support the assumption that CPT is the most empirically adequate alternative to EUT. In Section 3, we consider implications of this finding for methodological debates about how to model welfare effects of policies, and argue that abandonment of a dogmatic belief in CPT as the 'correct' theory of risky human choice exposes a conceptual error that pervades self-consciously 'behavioral' welfare economics. In Section 4, we discuss second-order, philosophical, issues around the grounding of normative commitments in policy-focused economics.

2. CPT as a descriptive theory

Assume that utility of income is defined by a utility function U(x), where x is the lottery prize. Under EUT the probabilities for each outcome x_j, $p(x_j)$, are those that are induced by the experimenter, so expected utility is simply the probability weighted utility of each outcome in each lottery. Once the utility function is estimated, it is a simple matter to evaluate the implications for risk aversion. The concept of risk aversion traditionally refers to 'diminishing marginal utility,' which is driven by the curvature of the utility function, which is in turn characterized by the second derivative of the utility function, reflecting aversion to variability of outcomes.

CPT is not the only alternative to EUT. The rank-dependent utility (RDU) model of Quiggin (1982) extends the EUT model by allowing for decision weights on the utility of lottery outcomes. These decision weights reflect probability weights on objective probabilities. The decision weights are defined after ranking prizes from largest to smallest. The largest prize receives a decision weight equal to the weighted probability for that prize: the decision weight reflects the probability weight of getting at least that prize. The decision weight on the second largest prize is the probability weight of getting at least that second largest prize, minus the decision weight of getting at least the highest prize. Similarly for other prizes.

RDU in fact denotes a family of specifications that all include probability weighting of outcomes and allow for the integration of laboratory earnings or losses with background wealth. EUT is nested within RDU, as the special case in which there is no probability weighting.

The key innovation of CPT, in comparison to EUT and RDU, is to allow sign-dependent preferences, where risk attitudes depend on whether the agent is evaluating a gain or a loss as perceived by the agent. Kahneman and Tversky (1979) introduced the notion of sign-dependent preferences, stressing the role of the reference point when evaluating lotteries. They defined loss aversion as the notion that the disutility of losses weighs more heavily than the utility of comparable gains.

Tversky and Kahneman (1992, p. 309) popularized the functional forms we often see for loss aversion:

$$\begin{aligned} U(m) &= m^{1-\alpha}/(1-\alpha) && \text{when } m \geq 0 \\ U(m) &= -\lambda[(-m)^{1-\beta}/(1-\beta)] && \text{when } m < 0, \end{aligned}$$

where λ is what we will call the *utility* loss aversion parameter. Here we have the introduction of the assumption that the degree of utility loss aversion for small unit changes is the same as the degree of utility loss aversion for large unit changes: the same λ applies locally to gains and losses of the same monetary magnitude around 0 as it does globally to any size gain or loss of the same magnitude. This is not a criticism, but highlights a restrictive parametric specification.

Abdellaoui, Bleichrodt, and Paraschiv (2007, p. 1662) provide a clear statement of the 'exchange rate assumptions' used to define the utility loss aversion parameter λ in the literature. Tversky and Kahneman (1992) used $-U(-1)/U(1)$, and others have used $U'(-x)/U'(x)$, $-U(-x)/U(x)$, $U(x)-U(y) \leq U(-y)-U(-x)\ \forall\ x > y \geq 0$. One can make the exchange rate assumptions formally *de minimus* by defining an index of loss aversion solely in terms of the directional derivatives at the reference point, $U' \rightarrow (0)/U' \leftarrow (0)$, as proposed by Köbberling and Wakker (2005). But this has the very unfortunate effect, emphasized by Wakker (2010, p. 247), that global properties of loss aversion are being driven by extremely local properties of estimated utility functionals, which puts great strain on empirics and functional form assumptions.

One immediate implication of this last point for normative economics: to assign a specific CPT utility function to an actual person is to make a very strong empirical claim about utility loss aversion, for which production of appropriate evidence will be correspondingly demanding. The specific utility function might be a parametric function or not; in either case it needs to be carefully and reliably elicited, conditional on assumptions about 'the' reference point for this individual in this setting.[5] This implication necessarily increases the risk involved in offering policy advice based on such an assumption about the estimation of utility loss aversion.

What if the decision weights for the gain domain differ from the decision weights for the loss domain? There is nothing *a priori* in CPT to rule this out. Even if the basic utility functions for gains and losses are linear, and conventional utility loss aversion is absent ($\lambda = 1$), this could induce the same behavior as if there were utility loss aversion. This is called *probabilistic* loss aversion by Schmidt and Zank (2008, p. 213). Imagine that there is no probability weighting on the gain domain, so the decision weights are the objective probabilities, but that there is some probability weighting on the loss domain. Then one could easily have losses weighted more than gains, from the implied decision weights.

Anyone using the expression 'loss aversion' must allow for there to be two psychological pathways to account for the differential risk premium generated by either utility loss aversion and/or probabilistic loss aversion. This matters for normative economics because one might take a different stance on the welfare significance of one form of loss aversion than on the other form of loss aversion. Again, there is nothing here that is radical from the perspective of CPT *per se*, although it does raise questions for advocates of CPT who are also advocates of specific empirical regularities.[6] We insist on remaining agnostic about those regularities. At the very least, claims about the validity of CPT in general should be kept separate from claims about the validity of particular empirical strains of CPT.

What does the literature say on the empirical evidence for CPT? The record is remarkable, and deserves scrutiny by methodologists. Virtually no studies have estimated a structural model of CPT in which all tasks were for real payoffs, quite apart from whether or not an incentive-compatible elicitation procedure is used. And those very few that have met these criteria find little evidence for CPT. This is a history of thought that is easy to read, which makes it even more puzzling that it is clearly not read. Harrison and Swarthout (2016) provide an extensive, detailed literature review, which we summarize here.

2.1. Base camp for CPT: Tversky and Kahneman (1992)

Tversky and Kahneman (1992) gave their 25 subjects a total of 64 choices. Their subjects received $25 to participate in the experiment, but rewards were not salient, so their choices had no monetary consequences. They had 28 choices in the gain frame, and 28 in the loss frame. The terms 'gain frame' and 'loss frame' refer here to lotteries in which all prizes are (weakly) gains or losses. A further eight tasks involved mixed-frame gambles, where the term 'mixed frame' refers to lotteries in which some prizes are (strictly) gains and some are (strictly) losses.

Tversky and Kahneman (1992) estimate a structural model of CPT using nonlinear least squares, and at the level of the individual. In addition to the functional forms for utility and loss aversion, they propose the generic probability weighting function $\omega(p) = p^{\gamma}/(p^{\gamma} + (1 - p)^{\gamma})^{1/\gamma}$, and allow one parameter $\gamma +$ for gains and another

parameter $\gamma-$ for losses. Remarkably, they then report the *median* point estimate, for each structural parameter, over the 25 estimated values. So over all 25 subjects, and using our notation, the median value for α was 0.88, the median value for β was also 0.88, the median value of λ was 2.22, the median value of $\gamma+$ was 0.61, and the median value of $\gamma-$ was 0.69. These parameter estimates are remarkable in three respects, given the prominence they have received in the literature.

- First, whenever one sees point estimates estimated for individuals, one can be certain that there are many 'wild' estimates from an *a priori* perspective, so reporting the median value alone might be quite unrepresentative of the average value, and provides no information whatsoever on the variability across subjects.
- Second, there is no mention at all of standard errors, so we have no way of knowing, for example, if the oft-repeated value of λ is statistically significantly different from 1.
- Third, the median value of any given parameter is not linked in any manner to the median value of any other parameter: these are *not the values of some representative, median subject*, which is often how they are implicitly portrayed. The subject that actually generated the median value of λ, for instance, might have had any value for α, β, $\gamma+$, and $\gamma-$.

These shortcomings of the Tversky and Kahneman (1992) study have not led anyone to replicate their experiments with salient rewards and report complete sets of parameter estimates with standard errors. The fault is not that of Tversky and Kahneman (1992), who otherwise employed quite modern methods, but the subsequent CPT literature. Anybody casually using these estimates as statistically representative must not care about rigor in empirical work (e.g. Barberis & Huang, 2008, p. 2071ff).

2.2. *Hypothetical choices*

Many studies that claim to have structural estimates of CPT use choices made entirely over hypothetical outcomes. This is not the place to revisit the tired debate over the unreliability of hypothetical choices in experiments. If anyone claims that they reliably provide the same results, they simply have not read the literature, as reviewed by Harrison (2006, 2014). Providing a fixed cash payment for participation is not the same thing as providing salient rewards that vary with the choices made.

Many other studies that claim to have structural estimates of CPT use choices over gain frames with real payoffs, but then use hypothetical payoffs for choices over loss frames or mixed frames. For example, Abdellaoui et al. (2008) asked real questions in the gain frame, but only hypothetical survey questions in the loss and mixed frames. The line of argument used to justify this approach is worth stating carefully. The latest version is in Abdellaoui and Kemel (2014, p. 1856), who note that:

> The implementation of real incentives for monetary consequences in individual choice under risk is known to be somewhat problematic when it comes to loss/mixed prospects (…) First, an implementation of this type imposes the playing out of loss/mixed questions for real, which is ethically questionable (…) Second, using an initial endowment with real losses could be costly given that one has to elicit utility on a sufficiently wide interval of monetary losses to observe its curvature. Furthermore, subjects benefitting from a prior endowment may integrate the payoffs and then not perceive any loss.

Take the three points of this argument seriously, sentence by sentence:

- The ethical issue arises if one recruits subjects to a lab and requires them, by some means, to cover any losses from their private wealth. And that would also imply, as they separately note, a clear potential for sample selection bias, since subjects would have to be told about this possibility before participating. But no ethical issue arises if subjects face losses out of a house endowment or an earned endowment, as long as they do not incur net losses over the session.[7]
- How is it that the interval of monetary losses has to be significantly greater than the interval of monetary gains? Nothing in CPT requires this, and indeed most of the popular 'exchange rate assumptions' defining the utility loss aversion parameter λ require that one study choices that entail the same small gains or losses of *exactly* the same magnitude around the reference point.
- The fact that subjects 'may integrate the payoffs and not observe any loss' is exactly the thing being tested when one tests CPT against EUT or RDU! This is like saying that implementing real losses is 'problematic' because it will show that CPT is not empirically supported, hardly a strong position to defend.

Is CPT so empirically fragile that we have to resort to arguments like these to defend it?

2.3. *No mixed frame choices*

Estimation of utility loss aversion is logically impossible without mixed frame choices. The λ parameter scales up the valuation of lotteries in the loss frame equally, so can have no effect on such choices under CPT, and of course the λ parameter has no effect at all on the valuation of lotteries in the gain frame. This is theory, not some empirical issue.

Bruhin, Fehr-Duda, and Epper (2010) estimated parametric models of what they referred to as CPT that assumed that the utility loss aversion parameter λ was 1, noting wryly that 'our specification of the value function seems to lack a prominent feature of prospect theory, loss aversion …' (p. 1382). They did this because their design only included lotteries in the gain frame and the loss frame, and none in the mixed frame. They did provide real incentives for decisions, and employed a house endowment to cover losses. But they cannot estimate utility loss aversion, argued by many to be a core feature of CPT. The same problem occurs in other designs.

2.4. *Some new evidence*

Harrison and Swarthout (2016) report experiments designed to test CPT against EUT and RDU in a controlled laboratory setting. They designed a battery of tests that allows identification of all of the parameters of the EUT, RDU, and CPT models, and that allows estimation of a wide range of risk preferences. The battery of 100 binary choices had gain-framed lotteries, loss-framed lotteries, and mixed-framed lotteries, and all losses were framed as coming out of a house endowment.

The sample consisted of 177 undergraduate students and 94 MBA students from the Georgia State University population. The domain of net prizes for the undergraduates spanned \$0 to \$70, and spanned \$0 to \$750 for the MBA students. Separate models of EUT, RDU, and CPT risk preferences were estimated for each subject. Nested and non-nested hypothesis tests were then used to compare the models for each subject.

There are two major findings of relevance here. First, the evidence is that a clear majority of individuals in the sample do *locally asset integrate*. That is, they see the loss frame for what it is, a frame, and behave as if they evaluate the net payment rather than the gross loss when one is presented to them. This finding is fatal to the direct application of CPT to these data. It also sets a serious behavioral bar for moving beyond the simplest framing of losses. In effect, CPT fails to be a descriptively accurate model for these subjects because they asset integrate, at least locally over the gross and net prizes presented to them. By any standard statistical metric, CPT is a *descriptively* inferior model of behavior.

The second major finding is that RDU emerges as the most important non-EUT model of risk preferences from a descriptive perspective, not CPT. This reminds us also that many champions of CPT are actually championing evidence for probability weighting over gains, and just calling that CPT (e.g. Abdellaoui, l'Haridon, & Paraschiv, 2013).

When a separate sample of 58 undergraduate subjects covered losses out of an *earned* endowment, from a general knowledge quiz, the support for CPT increased slightly, but it was still not close to being the modal specification.

2.5. *Econometrics*

There is a methodologically unfortunate divide between those doing theory, those designing experiments, and those doing econometrics. Each is complementary, and the evaluation of descriptive models of risk preferences is one place where that complementarity matters.

One expression of this methodological divide that matters is variance in the manner in which behavioral errors are handled. Does one view these as arising *in the model* when the individual forms a preference for an outcome, forms the scalar evaluation of a lottery, compares the evaluation of one lottery with another, or when the final choice is being operationalized? All are found in the literature, and it is extremely difficult to identify more than one of them at a time, to try to determine which behavioral error story is best for an individual. And, of course, the behavioral error story that best characterizes one individual need not be the story that best characterizes other individuals. Finally, the evaluation of the behavioral error story is normally conditional on a specific model of risk preferences, which might itself be a poor fit for that individual.

There are also trade-offs between the econometric models used to accommodate individual heterogeneity. Most modern researchers agree that the 'representative agent' is a fiction that must be discarded, but how? One can pool data across individuals and allow deep, structural parameters to be (linear) functions of observable task or decision-maker characteristics (e.g. Harrison & Rutström, 2008). These models accommodate some minimal heterogeneity, and tend to be relatively easy to estimate reliably and consistently. The next step is to model each individual, which of course allows for unobserved individual heterogeneity (e.g. Harrison & Ng, 2016; Hey & Orme, 1994). The problem here is that one invariably comes up with some 'wild' estimates for certain individuals and models, where 'wild' is in relation to our priors. Or one can model pooled behavior with random coefficients (e.g. Andersen, Harrison, Hole, Lau, & Rutström, 2012), or in Bayesian jargon as hierarchical Bayesian estimation (e.g. Nilsson et al., 2011). These methods retain the benefits of pooled estimation, but allow for unobserved individual heterogeneity.

Another step toward modeling heterogeneity is to allow individuals to be making choices as if there is some probability that one model of risk preferences is generating behavior, and a residual probability that some other model of risk preferences is generating behavior. These 'mixture models' can be implemented at the level of the choice observation (e.g. Harrison & Rutström, 2009) or the level of the individual (e.g. Conte, Hey, & Moffatt, 2011). The latter entails an assumption that *every* choice of an individual is either characterized by one or other model: for instance, that the individual is solely an EUT decision-maker or solely a CPT decision-maker, for each and every choice that the individual makes. Or one can even allow mixtures to occur at the level of the model itself, as in certain 'dual criteria' theories from psychology (e.g. Andersen, Harrison, Lau, & Rutström, 2014). Mixture models can be usefully viewed as allowing for process-heterogeneity, where the word 'process' derives from these mixtures being weighted averages of the likelihoods of distinct data generating processes.

Bayesian approaches allow one to combine the strengths of each of these approaches. Pooled models, conditioned on observable characteristics, can be used as a prior for an individual: knowing the characteristics of the individual we can infer individual-specific point estimates and covariances from the pooled model. Data from the individual can be then combined with these priors. For those individuals with clear behavior patterns, the data will dominate the prior; and for those individuals with 'wild' behavior, the prior will have to play a greater role to generate a well-behaved posterior inference. Hierarchical Bayesian methods then also allow all of the process-heterogeneity that mixture models allow.

3. Welfare assessment and grades of paternalism

The widespread view that welfare should be assessed on the basis of behaviorally derived utility functions rather than EUT, at least when the latter is interpreted following Savage (1954) and Binmore (2009) as an explicitly normative model, is primarily based on concerns about paternalism (Sugden, 2009). This motivation has considerable surface plausibility if CPT is assumed as the default descriptive model. Recall that CPT incorporates two paths to loss aversion: 'probability loss aversion' and 'utility loss aversion.' In the standard technical sense of cognitive science, probability weighting is a kind of *representation*, which entails that it can, though it need not, involve cognitive or perceptual *error* (Dretske, 1991). By contrast, λ is a response operator, most naturally interpreted as reflecting a sentimental influence on behavior and cognition.[8] To the extent that a person experiences direct sentimental disutility from losses *per se*, whenever she interprets an outcome as a loss, then it seems straightforwardly presumptuous to maintain that a policy-maker should override this aspect of her psychology.[9]

However, in Section 2 we argued that there is little or no empirical evidence for the default interpretation that most observed loss aversion is utility loss aversion. Evidence to date is more consistent with the hypothesis that this behavior most often reflects subjective weighting of probabilities that diverges from objective probability distributions. There are plausible circumstances under which this can be rationalized and so implies no error. In general, agents may have reasons to use representational heuristics to supplement 'rationality' as modeled for application in 'small worlds' when their real decision contexts are 'large worlds' (Chew & Sagi, 2008).[10] Imagine a world of uncertainty or ambiguity where the agent has special aversion to an extreme downside event, such as bankruptcy or death. Then she might place greater weight on these worst outcomes, as a heuristic. Or imagine a world, familiar to actuaries, in which agents have

to make decisions over ε probabilities with massive consequences, but know how poorly they infer tails of distributions where data are sparse, and therefore rely more and more heavily on parametric assumptions. Or yet further, imagine a world in which an agent does not trust the process generating the probabilities, and suspects some strategic concerns; as discussed by Schneeweiss (1973) and Kadane (1992), this can rationalize what looks like ambiguity aversion. However, in the absence of an identified reason as to why an agent might have a subjective belief that renders probability weighting a reasonable way to mitigate uncertain or ambiguous risks, the default assumption is that probability weighting is a representational error.

The finding that most loss aversion appears to reflect RDU rather than CPT undermines the case for abandoning EUT as the normative standard for welfare assessment, at least as far as some literature on paternalism is concerned. Le Grand and New (2015) argue that 'government' paternalism, ideally exercised by a disinterested agent seeking to bring a society closer to a Pareto frontier, is *prima facie* acceptable when its object is correction of objectively verifiable error. Sugden (2009) opposes the 'soft' paternalism of nudge advocates (Sunstein & Thaler, 2003a, 2003b) who side with target agents' deliberative preferences against their (putatively) more impulsive preferences because, he argues, this violates legitimate and characteristically human indulgence in inconsistency.[11] However, RDU does not imply any sort of inconsistency; and Sugden (2009) does not suggest that public policy is obliged to respect cognitive *ignorance*. Were someone to defend the view that policy should respect ignorance or perceptual error, this would represent a radical challenge to the broadly liberal status quo in political philosophy. Most liberals, for example, think that a person whose sentiments incline her to smoke is best off if allowed to do so, at least out of range of unwilling second-hand victims and if she is willing to pay the smoking-attributable cost of health care. But these same liberals also think that public resources should be spent on ensuring that the smoker is not ignorant of the likely effects of her behavior on her physical health. Consistency with such opinions suggests that public policy should also be designed to correct more general misperceptions about objective probabilities, and by implication objective risks, such as those represented by rank-dependent preferences in small-world contexts.

The argument just given implies maintaining EUT as a normative standard for policy assessment. However, in the context of challenges from behavioral dissenters against standard economics, it would beg questions to leave the standard economic concept of welfare,[12] which features in the case for privileging EUT, unexamined. How do behavioral economists challenge the idea that optimization of social welfare[13] is an appropriate basis for policy assessment? Are there alternative standards with which the welfare standard competes? The evidence presented in Section 2 undermines the assumption that most people's risky choices are best described by CPT. Do arguments against the welfare standard, or in favor of alternatives to it, rest to any extent on this undermined assumption?

In the realms of policy assessment frequented by both economists and philosophers there has been considerable inconsistency in interpreting the relationship between welfare, as studied by economists, and more diffuse conceptions of well-being that philosophers have examined under the broad influence of Aristotle. Many economists have avoided engagement with the philosophers' concepts on suspicion that these concepts reflect paternalism, since it seems that promoting well-being as distinct from welfare must necessarily involve privileging some sources of utility, those the philosopher regards as most robust under ideal deliberation, from others. The philosophers Tiberius and Plakias (2010), in a finely balanced review of concepts of well-being, clearly

acknowledge this concern and distance themselves from orthodox Aristotelians on the basis of it. On the other hand, Sen (1999) sharply criticizes 'welfarism' as a pinched foreshortening of well-being that dogmatically refuses to consider information that is not represented in utility functions. In their rhetoric, behavioral economists have frequently sided with Sen (1999) on this question (see Davis, 2003, 2010; Ross, 2005, Chapter 4), on the (imprecise) grounds that by incorporating the psychological wellsprings of choice into formal models, behavioral work appropriately widens the informational basis of normative economic theory and 'humanizes' the economic agent. A common populist slogan is that behavioral economics unifies 'Homo economicus' with *Homo sapiens*. Thus, whereas mainstream economists have tended to keep their distance from the philosopher's idea of well-being, behavioral economists have often written as if their refreshed concept of welfare should at least be a proxy for well-being, if not in fact equivalent to it.[14]

This has been the conceptual backdrop for most welfare analyses that have abjured EUT as the relevant normative standard. At first glance, and with CPT taken as the default alternative to EUT, this can seem to generate a nice consensus. The philosopher who thinks that overriding a CPT chooser's sentimental aversion to loss would disregard her autonomy, and therefore harm her well-being, will make the same policy judgment as an economist who observes that a paternalistically overruled CPT chooser would be forced into accepting diminution of her subjective utility, at least in any choice frame where there is significant risk of what she perceives as a loss.

However, attending to the distinction between CPT and RDU, and to the basis for empirically identifying choices that, respectively, satisfy them, shows that something is wrong with this reasoning. Once we distinguish subjective sentimental disutility that the policy-maker should not override from losses in expected value that result from perceptual or cognitive error, we see that the economist who takes for granted that the overruled CPT chooser must suffer a welfare loss has smuggled in the unjustified assumption that the agent's entire utility decrement under paternalism is attributed to utility loss aversion. But since CPT incorporates probability weighting also, the hypothetical behavioral economist we are considering cannot know whether the expected value gain from paternalism might not more than compensate the agent for any sentimental utility loss she suffers. By contrast, when RDU is the standard then, in the absence of evidence-based rationalization of the agent's probability pessimism, the economist should expect that correction by reference to EUT would be welfare improving, *ceteris paribus*.[15] The philosopher might or might not object to correction, depending on the extent of normative libertarianism to which she subscribes. But, as noted earlier, only a relatively extreme and highly revisionary libertarianism will support resistance to informing people about objective probability distributions.

This argument in turn suggests that the behavioral economist's 'consensus' assimilation of welfare and well-being represents conceptual *muddling* rather than grand reconciliation. This conclusion could be derived at a purely abstract, hypothetical level even if evidence suggested that most human choices under risk were best characterized by CPT. But if in fact CPT's history of horse race victories over EUT is explained by the fact that many people have rank-dependent preferences, then the wedge between welfare and philosophers' conceptions of well-being is directly relevant to real, pressing, policy assessment methodologies.

More can be said about the implications of the empirical results, we now argue, if we attend in more detail to what philosophers intend when they distinguish well-being from welfare.

4. Second-order normativity

Suppose it is granted, at least for purposes of argument, that non-paternalistic policy can legitimately seek to correct non-rationalizable probability weighting distortions but not sentimental loss aversion. In that case, we might then propose an explanation of this judgment in terms of a second-order normative principle that has been widely shared among philosophers, going back at least to Hume. This principle is that people's sentimental (or, more narrowly, emotional) responses should be respected by non-paternalistic policy-makers, but that no such 'consumer sovereignty' restriction applies to cognitive errors. The classic principle is perhaps best illustrated by reference to the most historically important effort to complicate it, that of John Stuart Mill (1863). He, famously, refused to accept the view shared by his own father and by Bentham that a taste for pushpin should be regarded as equally worthy of promotion by policy as a taste for poetry. To reconcile this opinion with his anti-paternalism, Mill argued that the difference between the pushpin fan and the poetry lover is not merely sentimental after all, but is cognitively grounded: the poetry lover has acquired *knowledge* about the available range of experiential quality that the untutored pushpin fan lacks. Thus, for Mill, teaching the philistine about poetry resembles, in our terms, correcting probability weighting more than it resembles overruling sentimental loss aversion.

The kind of distinction Mill aimed to draw is at least broadly similar to what we find in the most sophisticated contemporary accounts of well-being by philosophers who, using Sen's (1999) language mentioned earlier, resist 'welfarism.' Tiberius and Plakias (2010) articulate and defend what they call a 'values-based life satisfaction (VBLS) account' of well-being. According to this account '… it is satisfaction with how one's life is going overall *with respect to one's values* that counts as well-being. In other words, life satisfaction constitutes well-being when it is a response to how life is going according to certain standards, and these standards are provided by a person's values' (p. 421). The key concept to be theoretically refined here is that of a *value*. Tiberius and Plakias (2010, pp. 422–423) identify

> … three features that … are important if values are to play a role in an account of well-being:
>
> First, values must be normative from the point of view of the person who has them: that is, a person takes her values to provide good reasons for doing things. This must be the case if values are to answer the problem of normative arbitrariness. Second, values include an affective component; part of what it is to care about something in the way distinctive of valuing is to have some positive emotional response toward it. This must be the case if values are to provide the ground for the positive attitude of life satisfaction. Third, values are relatively stable, as they must be on our view since well-being itself is relatively stable.
>
> … if values are to solve the problem of normative arbitrariness, they must be subject to standards of correctness or appropriateness. We think that our rough characterization of values … suggests two such standards, which we might call the standard of affective appropriateness and the standard of information … (V)alues that are sustained by false beliefs are unlikely to be stable because new information will put pressure on them to change. Moreover, values based on false beliefs about what one's emotional needs are, or what one will find satisfying, are unlikely to produce a positive emotional response over the long term.
>
> … According to VBLS, life satisfaction is a positive cognitive/affective attitude toward one's life as whole, and life satisfaction constitutes well-being when it is not defeated by considerations that undermine its normative authority.

The key feature of this account for our purposes is that it is normative at second-order. That is, it can provide *personal* normative standards for assessing first-order *institutional* policy preferences.

A philosophical utilitarian would object to the VBLS on grounds that life satisfaction is simply a source of utility that, in principle, competes with other typical sources of utility such as wealth and hedonic pleasure. We have nothing to add to the argument of Tiberius and Pakias's against philosophical utilitarianism except to note that, contrary to widespread populist views about mainstream economics, the leading methodologists of the neoclassical synthesis, particularly Pareto, Hicks, and Samuelson, rejected philosophical utilitarianism, urging instead that philosophy be left to philosophers while economists got on with economics (see Mandler (1999) for careful history). The relationship between well-being, as a *second-order* norm for personal policy, and welfare, as a *first-order* norm for public policy, could then have the following structure. A policy-maker could defend her personal commitment to promoting welfare-optimizing policies, and set an example for similar commitment by other policy-makers and by economists, on grounds that preference for policies that move society closer to, or prevent it falling further away from, a Pareto frontier, is a value that satisfies the criteria of Tiberius and Plakias (2010). And the policy-maker could defend drawing people's attention to their cognitive and perceptual errors as being generally compatible with citizens' VBLS. Furthermore, it is a value the policy-maker or economist is particularly likely to find informationally and affectively stable because it allows them to make full use of their professional knowledge and expertise.

This appeal to the authority of moral philosophers[16] is intended to serve two purposes. First, it shows how, in general, one can resist the muddling of welfare and well-being that casual reliance on CPT as a dogmatic generalization about human risky choice has, according to us, encouraged. Second, the VBLS grounds, in a second-order normative structure, a basis for thinking that correcting common distortions in probability weighting is appropriate.

We conclude by briefly explaining what we have in mind by the second point. Welfare measures as economists have developed them must serve at least two essential criteria. First, they must encapsulate the normative concern that, we contend, really *is* central to the normative case for bringing economics to bear on policy design and policy choice. This is the view that resources that serve human wealth should not be wasted. Welfare criteria are *efficiency* measures. A community with RDU risk preferences in small worlds *must* waste some utility *because* their choices incorporate confused perceptions of probabilities. Correcting RDU preferences by reference to EUT is a strategy for reducing this inefficiency. Second, welfare criteria must allow for comparison of relationships between choices and outcomes in ways that are objectively defensible. Two agents' probability weighting metrics are measurably comparable. Thus RDU, like EUT, is a *practical* model of utility for empirical welfare economics. CPT, because its application is highly sensitive to identification of reference points that are expected to vary idiosyncratically, is a relatively impractical model for empirical welfare economics. We conjecture that this largely explains why CPT, despite the enormous attention it has received from theorists, has been as sparsely tested as our review in Section 2 indicates: it is *difficult* to directly empirically test, because it requires the experimenter to induce a real loss frame. The awkwardness of CPT as a welfare measure further explains why theorists attracted to it have been tempted to confound welfare with other normative concepts.

None of this would be much to the pragmatic point if most human risky choice actually seemed to be best described by CPT; we would just have to live with the practical challenges. But, fortunately for theorists and policy architects alike, the facts that we presently have do not point in that direction.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes

1. See Bernheim (2009), Manzini and Mariotti (2014), Rubinstein and Salant (2012) and Sugden (2004, 2009).
2. The definition of 'the' reference point is subtle and difficult, and was deliberately left 'outside the model' by Kahneman and Tversky (1979). It is also related to debates over the modeling of asset integration assumptions: see Cox and Sadiraj (2006). Kőszegi and Rabin (2007) and Schmidt, Starmer, and Sugden (2008) consider the theoretical implications of loss aversion relative to a *stochastic reference point*, defined in terms of *subjective beliefs* about outcomes of the lottery. Kőszegi and Rabin (2007, p. 1051) recognize that '… relatively little evidence on the determinants of reference points currently exists.' The only operational theory of endogenous reference points, albeit outside of CPT, comes from the disappointment aversion model of Gul (1991), who proposed the certainty equivalent as the reference point for a lottery.
3. Our concern here should not be confused with the misuse of Ockham's Razor to motivate a principle according to which a model with fewer parameters is, all else being equal, superior to a model with more parameters. Whether or not one agrees with that principle, and we do not in general, these methods apply mechanical corrections to the standard log-likelihood test of two models to penalize them for less parsimonious specifications. The trade-off between precision of model fit and parsimony of model specification, if one is needed at all, is far more subtle than these corrections can possibly encompass. For instance, one always desires less restrictive parametric functional forms for utility functions and probability weighting functions, but these typically require more parameters (particularly if one insists on non-parametric functions). How is this trade-off, among many others, captured through parameter counting and mechanical penalty factors?
4. When we say 'in accordance with EUT,' we refer to the EUT specification estimated for this subject given her observed choices.
5. Thaler and Johnson (1990) illustrate many different ways to construct reference points for the same individual.
6. One such regularity is 'the fourfold pattern of risk attitudes,' although there are several such patterns to consider (Scholten and Read (2014)). Another is 'inverse-S' probability weighting over gains, which is far from the norm.
7. Allowing subjects to incur a loss from their own money runs into legal issues, since the right to extract money from gamblers normally vests with governments.
8. Psychologists Charpentier, De Neve, Xinyi, Roiser, and Sharot (2016) interpret probability loss aversion as reflecting varying perceptual salience of sentiments, and utility loss aversion as reflecting direct sentiments, where they refer to sentiments as 'feelings.' In a salient experiment in which subjects are asked to report feelings, in addition to making lottery choices, their results echo those of Harrison and Swarthout (2016): they find probability loss aversion but no utility loss aversion. Charpentier et al. (2016) correctly included a loss frame, a gain frame, and a mixed frame in their incentivized choice battery, and used a house endowment to cover losses.
9. It is always possible that a person might change her choice behavior after it is brought to her attention that her loss aversion reduces her expected value from a prospect. In the absence of specific evidence for such a preference, however, the charge of paternalism would stick if the person were advised as though she were an EUT chooser.

10. A referee suggested that perhaps the most difficult and significant policy situations in which paternalistic interventions are considered are such large world choice contexts. There is no evident metric for assessing this reasonable possibility; but it can hardly be disputed that applications of EUT and CPT, which presuppose small worlds, have featured massively in the literature on nudging.
11. A referee points out that the preference inconsistency mainly addressed in this debate is alleged to result from experienced regret, which motivates, and arguably might justify, an intervention in support of the later, wiser version of the agent.
12. By this we intend reference to a comprehensive and canonical representation of this concept, such as the classic Graaff (1957).
13. Again, we refer here to social welfare in the broad sense of standard welfare economics, not merely to outputs of social welfare functions in the sense of Sen (1970).
14. We will pass over the large literature on another candidate proxy for well-being, hedonic (episodic or lifetime) satisfaction. We endorse the basis for the brisk rejection of this given by Tiberius and Plakias (2010, pp. 405–407).
15. We can say how *ceteris paribus* would generally be fleshed out here: by reference to any utility loss that might result from whatever mechanism is used to implement the correction.
16. We call this an appeal to authority because we have not reviewed their arguments here. Of course we would not highlight the view if we were not impressed by the arguments.

References

Abdellaoui, M., Bleichrodt, H., & Paraschiv, C. (2007). Measuring loss aversion under prospect theory: A parameter-free approach. *Management Science, 53*, 1659–1674.

Abdellaoui, M., Bleichrodt, H., & L'Haridon, O. (2008). A tractable method to measure utility and loss aversion under prospect theory. *Journal of Risk and Uncertainty, 36*, 245–266.

Abdellaoui, M., & Kemel, E. (2014). Eliciting prospect theory when consequences are measured in time units: 'Time is not money'. *Management Science, 60*, 1844–1859.

Abdellaoui, M., l'Haridon, O., & Paraschiv, C. (2013). Individual vs. couple behavior: An experimental investigation of risk preferences. *Theory and Decision, 75*, 175–191.

Andersen, S., Harrison, G. W., Hole, A. R., Lau, M., & Rutström, E. E. (2012). Non-linear mixed logit. *Theory and Decision, 73*, 77–96.

Andersen, S., Harrison, G. W., Lau, M., & Rutström, E. E. (2014). Dual criteria decisions. *Journal of Economic Psychology, 41*, 101–113.

Barberis, N., & Huang, M. (2008). Stocks as lotteries: The implications of probability weighting for security prices. *American Economic Review, 98*, 2066–2100.

Bernheim, B. D. (2009). Behavioral welfare economics. *Journal of the European Economic Association, 7*, 267–319.

Binmore, K. (2009). *Rational decisions*. Princeton: Princeton University Press.

Bleichrodt, H., Pinto, J., & Wakker, P. (2001). Using descriptive findings of prospect theory to improve the prescriptive use of expected utility. *Management Science, 47*, 1498–1514.

Bruhin, A., Fehr-Duda, H., & Epper, T. (2010). Risk and rationality: Uncovering heterogeneity in probability distortion. *Econometrica, 78*, 1375–1412.

Charpentier, C., De Neve, J.-E., Xinyi, L., Roiser, J., & Sharot, T. (2016). Models of affective decision making: How do feelings predict choice? *Psychological Science, 27*, 763–775.

Chew, S. H., & Sagi, J. (2008). Small worlds: Modeling attitudes toward sources of uncertainty. *Journal of Economic Theory, 139*, 1–24.

Conte, A., Hey, J., & Moffatt, P. (2011). Mixture models of choice under risk. *Journal of Econometrics, 162*, 79–88.

Cox, J., & Sadiraj, V. (2006). Small- and large-stakes risk aversion: Implications of concavity calibration for decision theory. *Games and Economic Behavior, 56*, 45–60.

Davis, J. (2003). *The theory of the individual in economics*. London: Routledge.

Davis, J. (2010). *Individuals and identity in economics*. Cambridge: Cambridge University Press.

Dretske, F. (1991). *Explaining behavior*. Cambridge, MA: MIT Press.

Graaff, J. d. V. (1957). *Theoretical welfare economics*. Cambridge: Cambridge University Press.

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

Harrison, G. W. (2006). Hypothetical bias over uncertain outcomes. In J. A. List (Ed.), *Using experimental methods in environmental and resource economics* (pp. 41–69). Northampton, MA: Edward Elgar.
Harrison, G. W. (2014). Real choices and hypothetical choices. In S. Hess & A. Daly (Eds.), *Handbook of choice modeling* (pp. 236–254). Northampton, MA: Edward Elgar.
Harrison, G. W., & Ng, J. M. (2016). Evaluating the expected welfare gain from insurance. *Journal of Risk and Insurance, 83*, 91–120.
Harrison, G., & Rutström, E. E. (2009). Expected utility *and* prospect theory: One wedding and a decent funeral. *Experimental Economics, 12*, 133–158.
Harrison, G., & Swarthout, J. T. (2016). *Cumulative prospect theory in the laboratory: A reconsideration* (CEAR Working Paper No. 2016-05). Atlanta: Center for Economic Analysis of Risk, Robinson College of Business, Georgia State University.
Hey, J., & Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica, 62*, 1291–1326.
Kadane, J. (1992). Healthy scepticism as an expected-utility explanation of the phenomena of Allais and Ellsberg. *Theory and Decision, 32*, 57–64.
Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*, 263–292.
Köbberling, V., & Wakker, P. (2005). An index of loss aversion. *Journal of Economic Theory, 122*, 119–131.
Kőszegi, B., & Rabin, M. (2007). Reference-dependent risk attitudes. *American Economic Review, 97*, 1047–1073.
Le Grand, J., & New, B. (2015). *Government paternalism: Nanny state or helpful friend?*. Princeton: Princeton University Press.
Mandler, M. (1999). *Dilemmas in economic theory*. Oxford: Oxford University Press.
Manzini, P., & Mariotti, M. (2014). Welfare economics and bounded rationality: The case for model-based approaches. *Journal of Economic Methodology, 21*, 343–360.
Mill, J. S. (1863). *Utilitarianism*. London: Parker, Son and Bourne.
Nilsson, H., Rieshamp, J., & Wagenmakers, E.-J. (2011). Hierarchical bayesian parameter estimation for cumulative prospect theory. *Journal of Mathematicla Psychology, 55*, 84–93.
Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior and Organization, 3*, 323–343.
Ross, D. (2005). *Economic theory and cognitive science: Microexplanation*. Cambridge, MA: MIT Press.
Rubinstein, A., & Salant, Y. (2012). Eliciting welfare preferences from behavioral datasets. *The Review of Economic Studies, 79*, 375–387.
Savage, L. (1954). *The foundations of statistics*. New York, NY: Wiley.
Schmidt, U., Starmer, C., & Sugden, R. (2008). Third-generation prospect theory. *Journal of Risk & Uncertainty, 36*, 203–223.
Schmidt, U., & Zank, H. (2008). Risk aversion in cumulative prospect theory. *Management Science, 54*, 208–216.
Schneeweiss, H. (1973). The Ellsberg paradox from the point of view of game theory. *Selecta Statistica Canadiana, 1*, 65–78.
Scholten, M., & Read, D. (2014). Prospect theory and the 'forgotten' fourfold pattern of risk preferences. *Journal of Risk & Uncertainty, 48*, 67–83.
Sen, A. (1970). *Collective choice and social welfare*. San Francisco: Holden Day.
Sen, A. (1999). *Development as freedom*. New York, NY: Random House.
Sugden, R. (2004). The opportunity criterion: Consumer sovereignty without the assumption of coherent preferences. *American Economic Review, 94*, 1014–1033.
Sugden, R. (2009). Market simulation and the provision of public goods: A non-paternalistic response to anomalies in environmental evaluation. *Journal of Environmental Economics and Management, 57*, 87–103.
Sunstein, C., & Thaler, R. (2003a). Libertarian paternalism. *American Economic Review, Papers and Proceedings, 93*, 175–179.
Sunstein, C., & Thaler, R. (2003b). Libertarian paternalism is not an oxymoron. *The University of Chicago Law Review, 70*, 1159–1202.

Thaler, R., & Johnson, E. (1990). Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice. *Management Science, 36*, 643–660.

Tiberius, V., & Plakias, A. (2010). Well-being. In J. Doris & the Moral Psychology Research Group (Eds.), *The moral psychology handbook* (pp. 402–432). Oxford: Oxford University Press.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representations of uncertainty. *Journal of Risk and Uncertainty, 5*, 297–323.

Wakker, P. (2010). *Prospect theory for risk and ambiguity*. New York, NY: Cambridge University Press.