

# Varieties of Paternalism and the Heterogeneity of Utility Structures

Glenn W. Harrison<sup>1</sup>

Don Ross<sup>2</sup>

1 – Department of Risk Management and Insurance and Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, USA, Robinson College of Business, Georgia State University. Harrison is also affiliated with the School of Economics, University of Cape Town, South Africa. E-mail: [gharrison@gsu.edu](mailto:gharrison@gsu.edu)

2 - School of Economics, University of Cape Town, South Africa, and Center for Economic Analysis of Risk, Robinson College of Business, Georgia State University, USA. Email: [don.ross@uct.ac.za](mailto:don.ross@uct.ac.za)

## Abstract

A principal source of interest in behavioral economics has been its advertised contributions to policies aimed at ‘nudging’ people away from allegedly natural but self-defeating behavior toward patterns of response thought more likely to improve their welfare. This has occasioned controversies among economists and philosophers around the normative limits of paternalism, especially by technical policy advisors. One recent suggestion has been that ‘boosting,’ in which interventions aim to enhance people’s general cognitive skills and representational repertoires instead of manipulating their choice environments behind their backs, avoids the main normative challenges. A limitation in most of this literature is that it has focused on relatively sweeping policy recommendations and consequently on strong polar alternatives of general paternalism and strict *laissez faire*. We review a real instance, drawn from a consulting project we conducted for an investment bank, of a proposed intervention that is more typical of the kind that economists are more often actually called upon to offer. In this example, the sophistication of current tools for preference attribution, combined with philosophical externalism about the semantics of preferences that makes it less plausible to attribute their literal self-conscious representation to people as propositional attitude content becomes more tightly refined, blocks applicability of the distinction between nudging and boosting. This seems to call for irreducible, context-specific ethical judgment in assessing the appropriateness of the forms of paternalism that economists must actually wrestle with in going about their everyday business.

Keywords: nudging, paternalism, applied economics, risk preferences, investment choices

JEL codes: A11, A13, B40, C44, C54, C93, D01, D14, D63, D81, D91

## *1. Introduction*

A principal source of interest in behavioral economics has been its advertised contributions to policies aimed at ‘nudging’ people away from allegedly natural but self-defeating behavior toward patterns of response thought more likely to improve their welfare. Leading early promotions of this kind of application of behavioral studies are Camerer et al (2003) and Sunstein & Thaler (2003a, 2003b). Recently Grüne-Yanoff & Hertwig (2016) have distinguished nudging, which is based on the heuristics-and-biases (H&B) branch of behavioral economics research (Kahneman & Tversky 1982;

Kahneman 2011), from policies aimed at ‘boosting’, which apply the ‘simple heuristics’ (SH) research program due to Gigerenzer and his colleagues (Gigerenzer et al 1999; Todd et al 2012; Hertwig et al 2013). Nudging and boosting are contrasted as follows. Nudges aim to change a decision-maker’s (DM) ecological context and external cognitive affordances in such a way that the DM will be more likely to choose a welfare-improving option without having to think any differently than before. Nudging is thus open to the charge that it is manipulative (Ashcroft 2011; Conly 2013, p. 8). Its defenders point out that if people are naturally prone to systematic error, then any scaffolding built by any institution unavoidably involves manipulation, so the manipulation in question might as well be benevolent. Boosting, by contrast, involves endowing DMs with enhanced cognitive capacities by teaching them more effective decision principles<sup>1</sup>, which they can choose to apply or not once they have been enlightened. Thus boosting, according to Grüne-Yanoff & Hertwig (2016), avoids manipulating the agents to whom the policies in question are applied, and is to that extent less paternalistic.<sup>2</sup>

An additional contrast relevant to normative assessment is that a nudge would normally be expected to have effects only on the specific behavior to which it is applied, and only in the setting that the nudge adjusts. A boost, on the other hand, to the extent that it alters standing cognitive capacities, and associated behavioral propensities across ranges of structurally similar choice problems, might be hoped to generate what have been called ‘rationality spillovers’ (Cherry, Crocker, & Shogren 2003). Furthermore, boosting might plausibly capacitate people with defenses against non-benevolent nudging by narrowly self-interested parties such as marketers and demagogues.

The classic example of nudging is changing default options. If the policy maker thinks that workers ought to invest in retirement savings plans, then she can make participation in such a plan the outcome if the DM is passive, needing to take action only if she wants to act on a preference not to participate. Grüne-Yanoff & Hertwig’s (2016) leading example of a boost is teaching people to represent the alternatives in risky decisions as natural frequencies, even when they are presented as probabilities. This is thought to improve the quality of choices because some evidence suggests that people are more likely to use ‘accuracy-promoting’ heuristics when reasoning about the former than when reasoning about the latter.

Almost all examples in the literature on both nudges and boosts resemble these in taking the policy maker or the educator as the target community for whose

---

<sup>1</sup> Grüne-Yanoff & Hertwig assume that the principles in question should be effective heuristics in the sense of Gigerenzer *et al* (1999). This reflects the arguable assumption that any general reasoning principle that most people can adopt reliably across a range of decision contexts is by definition an heuristic.

<sup>2</sup> The motivation for boosting as characterized here is similar to that provided by John *et al* (2009) and John *et al* (2011) for what they call a ‘think’ strategy for correcting people’s reasoning errors. These authors make no reference to heuristics. Otherwise, think strategies are a special case of boosting that work through engaging the intended beneficiaries in collective deliberation. The form of boosting we will consider does not involve such deliberation. We share concerns raised by Le Grand & New (2015, p. 142) concerning the practicality and likely effectiveness of think strategies except under rarified circumstances.

consideration the policies are proposed. Though there is typically a general presumption that members of these communities should prefer to avoid *gratuitous* paternalism, it is often assumed that their primary aim is to maximize the probability that DMs influenced by their policy choices or educational interventions will maximize their welfare. Examples are typically constructed in such a way that what is taken to be the welfare-maximizing behavior is transparent.

This frame will strike many economists as problematic. Economists are typically more reluctant than policy makers or pedagogues to help themselves to opinions about what constitutes an agent's welfare. There is a strong tradition in economics of treating preferences as summaries of, or statistical patterns in, actual choices, rather than as independent standards against which to try to regulate decisions. Clearly this is partly because mainstream economics descends historically and intellectually from utilitarian and classical liberal political and moral philosophies that view paternalism as more or less anathema. But economists' suspicion of welfare judgments that aren't derived directly from the observed behavior of the people whose welfare is being judged also has other, more deliberative, sources. First, economists are typically highly sensitive to prospects for unintended consequences of policies. They see these as mainly arising from the interactions of people with heterogeneous preferences, or differing resources, or both, and so are less sanguine than many policy makers about letting normative considerations that are not fully decentralized drive policy choices. A myriad of micro-scale decisions, economists often suppose, will tend toward equilibria in which each participant is making the best choice for herself that she can given the choices of everyone else. Thus economists are often more comfortable making welfare assessments *ex post* rather than *ex ante*. But both nudging and boosting depend on *ex ante* evaluations. Second, economists distinguish between welfare, a technical concept of their own construction that is by definition subjective, but for which they have a well-stocked and venerable analytical tool-kit, from well-being, a broader but vaguer idea on which philosophers have long tolerated and indeed fostered disagreement.

Economists who emphasize the 'positive' nature of their enterprise (Friedman 1953) might simply assert that the merits or downsides of nudging and boosting are none of their concern *ex ante*, just as with all other normative questions. However, over the past couple of decades this has clearly become a minority stance within the discipline. Most economists think that theirs is policy-driven inquiry, in the strong sense that the hierarchy of interesting problems largely derives from the practical requirements of the businesses, governments, and households that seek their advice (Leamer 2012). The majority of economic inquiry is not basic research but is commissioned by clients seeking assistance in policy selection and design.

A more common view is that intervention to modify a target person's behavior can be acceptable paternalism when it corrects (and merely corrects) for failures of the target's rationality,<sup>3</sup> while any proposal for intervention that imposes normative

---

<sup>3</sup> Le Grand & New (2015), in philosophically analyzing government (as opposed to private) paternalism, refer more broadly to corrections of "judgment" rather than corrections of 'rationality.' We endorse their semantic preference. However, in the context where we are characterizing views common among, specifically, economists, 'rationality' is the more accurate term. Le Grand & New defend the normative thesis that justification of paternalism requires identification of a correctible judgment. We

judgments about the best way to live that the target might not share faces a *prima facie* obligation to morally justify the specific usurpation of the target's autonomy. This is the approach of so-called behavioral welfare theorists who argue for appeal to psychological facts about targets to ensure that when the economist's advice implies over-ruling a target's immediate preference, there is good reason to believe that the target's *ex post* preference will accord with the judgment implied by the advice (Bernheim & Rangel 2008). For example, if a person's behavior exhibits conflict between wanting to smoke and wanting to break the addiction, policy should side with the latter preference because, as a matter of psychological fact, few if any ex-smokers regret having quit, while most continuing smokers regret their recurrent lapses of willpower. These kinds of situations involving intrapersonal conflict and ambivalence are sometimes thought to mark the generic enabling conditions for acceptable nudging. Where they do not apply, the view would elaborate, we should try to change people only by teaching (or transparently incentivizing) them, not by manipulating them – that is, we should boost (or hire), not nudge.

We are here concerned with the distinction between nudging and boosting as it applies to what we will argue is a highly representative context of commissioned economic research. What we show is that the economist's need to operate with a technically precise model of the information built into the utility functions assigned to agents exposes problematic simplifications in the way in which the nudging versus boosting distinction is normatively interpreted. In particular, the behavioral welfare theorist's suggested meta-policy as characterized above fails to give the economist helpful advice in the most common sorts of policy situations with which she is called upon to engage.

The claimed normative advantage of boosting over nudging relies, as we explained, on the distinction between altering an agent's inner and outer environments. This might seem relatively straightforward if we assume, as many behavioral economists do, that the utility functions on which welfare analysis is based are grounded in latent cognitive processes on the 'inboard' side of the agent/environment boundary. However, economists model utility in a way that is better captured by externalist/ascriptionist accounts of minds such as Dennett's intentional stance (Ross 2014). This further complicates attempts to apply the nudging/boosting distinction to practical economic welfare assessments.

Economists typically infer agents' subjective assessments of value from their actual choices. This need not be based on an analytic identification of preferences with choices, as in Samuelson's (1937, 1938) original version of revealed preference theory. Ross (2014) argues that is more defensibly based on the philosophical thesis of externalism about the contents of intentional attitude ascriptions (see Appendix A, available at [http://cear.gsu.edu/wp-2016\\_06-varieties-paternalism-heterogeneity-utility-structures/](http://cear.gsu.edu/wp-2016_06-varieties-paternalism-heterogeneity-utility-structures/) for details). According to that thesis, such attitudes, which include

---

conjecture that most economists could be persuaded without much strain to agree that substituting the broader concept of judgment for a narrower concept of rationality would respect their normative concerns. However, incorporating that adjustment here would both require a distracting foray into wider issues in the philosophy of economics, and gratuitously complicate our focus on the interrelationship between economists' normative assumptions and the technical resources they use in welfare analyses.

beliefs as well as preferences, are ascribed by people to others and to themselves in such a way as to rationalize patterns of observed behavior (including utterances). Thus we do not take preferences to be internal psychological states. Intentional attitude ascription is holistic, taking account of all such behavior as is evident. We thus have no quarrel with Hausman's (2011) insistence that preference ascriptions implicate assumptions about beliefs, but we add to this the claim that belief ascriptions likewise implicate assumptions about preferences. The co-dependence of belief ascription and preference ascription is not viciously circular. Intentional attitude ascription is recursive and always open to revision as more evidence arrives. With Binmore (2009) we regard it as misleading to say that a person's preference for some *X* over some *Y* is a *cause* of their choosing *X* over *Y*; on the other hand, behavior that is rationalized by ascribing a preference for *X*'s over *Y*'s can be part of the causal background for predicting or explaining a specific *new* instance of choice of *X* over *Y*. Furthermore, past behavior rationalized by this preference ascription can also be part of the explanatory background for a choice among other contingencies related to *X* and *Y*, and this can be crucial in motivating welfare judgments.

Consider an example due to Harrison & Ng (2016). Suppose we think that a person has chosen an insurance policy that will reduce their utility relative to the state in which they did not choose the policy. If we were forced to say that the choice of the policy necessarily revealed a preference for having the policy over not having the policy, because we derive welfare assessments from preferences, then it would be impossible for any such choice to *ever* be deemed welfare reducing. This would show that the concept had been drained of the content that makes it useful. If we can't even say that a person reduces their welfare when they buy an actuarially unsound insurance policy (which people do), then we'll never be able to say anything about welfare in an applied context. But it would be consistent with taking behavior as the informational basis for preference ascription to hold that the choice was a mistake based on its inconsistency with ascription of a risk preference structure attributed on the basis of a run of the person's *other* behavior. In what follows we use elicited choices of lotteries, made under controlled experimental conditions, as such a basis. Arguably, this is a more direct and less noisy probe of risk preference structure than the choices of investment funds, also made in the lab, with which to make comparisons. Of course attribution of risk preferences derived from the lottery choices to the subjects choosing funds depends on the assumption that to some specified extent subjects' risk preferences are stable across choice contexts. We will illustrate a standard applied economist's application of this form of indirect preference revelation in working through the details of our main case study. We stress that some such assumption is necessary for normative evaluation if one is to allow for behavioral errors.

Our claim is not the implausibly strong one that nudging and boosting are indistinguishable. It is clear enough that changing people's behavior by altering its context and changing their behavior by teaching them new cognitive skills are not in general the same kind of thing, and that this difference is significant where concerns about paternalism arise. Our point, instead, will be to illuminate complexities that arise for this philosophically clear-enough distinction when it is exported from its home territory in purely normative policy and meta-policy debates, into a domain where normative and technical considerations are tightly entangled. In particular, we will argue, a meta-policy according to which boosting is morally unproblematic, while nudging proposals must always be accompanied by responses to concerns about

paternalism, is poorly adapted to the front line of applied economics. The difficulties we will illustrate pose new challenges for understanding the relationship between welfare and well-being, which have been revealed not by philosophical reflection (moral or conceptual), or by psychological discoveries, but by progress in economists' practices for measuring the valuations implicit in choices.

We conduct this exercise by describing a recent consulting project we carried out for a large South African retailer of investment products, and asking whether what we were doing for our client was helping them nudge their customers or helping them boost those customers. We also ask where any potential moral issues of interest arise, and for which parties.

In Section 2 we describe advances in the economic representation of utility and preference structures that give rise to the complications for the economist's normative stance that are our main subject. These advances motivate the design and interpretation of the commissioned experimental research that we describe in Section 3. Section 4 pulls the preceding strands together and gives the main argument of the paper. Section 5 briefly concludes, looking ahead both to philosophical and experimental research avenues indicated by our argument.

## *2. Modeling preference under risk and uncertainty: two alternatives to expected utility theory*

Define the risk premium as the difference between the actuarial expected value of a risky prospect and the certain amount of money an individual would accept in exchange for giving it up. Assume there is no bargaining process causing the individual to strategically mis-state this certainty equivalent if asked for it directly or indirectly.

We consider three core models of decision-making under objective risk. One is Expected Utility Theory (EUT), and posits that the risk premium is explained solely by an aversion to variability of earnings from a prospect. The second model is Rank-Dependent Utility (RDU), and further posits that decision-makers may be pessimistic or optimistic with respect to the probabilities of outcomes. RDU does not rule out aversion to variability of earnings, but augments it with an additional psychological process. The process may be 'latent' or 'virtual' in the sense associated with Dennett's (1987) intentional stance; that is, it might refer not to a specific physical computation 'in a person's head,' but to an equivalence class of relationships between decision contexts and observed choices. Both EUT and RDU assume that individuals asset integrate, in the sense that they net out framed losses from some endowment. The third model is Cumulative Prospect Theory (CPT), which adds an aversion to losses as a possible virtual psychological pathway to the risk premium, and also adds the assumption that gross gains and losses matter because individuals do not locally asset integrate and evaluate net gains or losses. We spend more time on the CPT model, since it is one that many recent writers on nudging and related behavioral topics believe to be descriptively superior, but is in fact not (see Harrison & Swarthout 2016).

### A. Expected Utility 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. Of course, 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 given by the second derivative of the utility function. Although somewhat loose, this can be viewed as characterizing individuals that are averse to mean-preserving increases in the variance of returns.

## B. Rank-Dependent Utility

The RDU model of Quiggin (1982) extends the EUT model by allowing for decision weights on lottery outcomes. These decision weights reflect probability weights on objective probabilities. The decision weights are defined after ranking the 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 the highest prize. Similarly for other prizes.

The Dual Theory (DT) specification of Yaari (1987) is the special case of the RDU model in which the utility function is assumed to be linear. Hence diminishing marginal utility can have no influence on the risk premium, and the only thing that can explain the risk premium is 'probability pessimism.'

## C. Cumulative Prospect Theory

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. The concept of loss aversion, based on sign-dependent preferences, is one that has been formalized in different ways in the literature. Given the popularity of the CPT model, it is important to review the different formalizations, and their varying implications for experimental design and estimation.

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. Here is the key paragraph (p. 279) introducing the concept:

A salient characteristic of attitudes to changes in welfare is that losses loom larger than gains. The aggravation that one experiences in losing a sum of money appears to be greater than the pleasure associated with gaining the same amount of money [...]. Indeed, most people find symmetric bets of the form  $(x, .50; -x, .50)$  distinctly unattractive. Moreover, the aversiveness of symmetric fair bets generally increases with the size of the stake. That is, if  $x > y \geq 0$ , then  $(y, .50; -y, .50)$  is preferred to  $(x, .50; -x, .50)$ . According to [their] equation (1), therefore,  $v(y) + v(-y) > v(x) + v(-x)$  and  $v(-y) - v(-x) > v(x) - v(y)$ . Setting  $y=0$  yields  $v(x) < -v(-x)$ , and letting  $y$  approach  $x$  yields  $v'(x) < v'(-x)$ , provided  $v'$ , the derivative of  $v$ , exists. Thus, the valuation function for losses is steeper than the value function for gains.

Note that there is no presumption here that the difference between  $v(x)$  and  $-v(-x)$  must be a constant,  $\lambda$ . That assumption is never made in Kahneman and Tversky (1979), and appears later in the literature.

When we say that the utility decrement of a unit loss, where the absolute value of  $(x-y)$  defines the unit here, is bigger than the utility increment of a unit gain, we need to be able to compare utility changes in the gain domain and the loss domain. This means that we cannot just have a utility scale that allows any order-preserving transformation: otherwise one could choose utility numbers that violated the hypothesis. In turn, this means that we have to be more restrictive than allowing positive affine transformations, and limit ourselves to defining utility on a ratio scale rather than an interval scale.

Note also the final discussion in the quote from Kahneman and Tversky [1979] about defining loss aversion in terms of the derivatives of the utility function around a zero reference point, which is  $y=0$  in the quote. This suggestion anticipates later proposals for defining loss aversion from Köbberling and Wakker (2005) and others.

Tversky and Kahneman [1992; p. 309] popularized the functional forms we often see for loss aversion, using a constant relative risk aversion (CRRA) specification of utility:

$$U(m) = m^{1-\alpha} / (1-\alpha) \text{ when } m \geq 0$$

$$U(m) = -\lambda[(-m)^{1-\beta} / (1-\beta)] \text{ when } m < 0,$$

and where  $\lambda$  is the loss aversion parameter. Here we have the introduction of the assumption that the degree of loss aversion for small unit changes is the same as the degree of loss aversion for large unit changes: the same  $\lambda$  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, just signposting of a restrictive parametric turn in the specification compared to Kahneman and Tversky (1979).

Abdellaoui, Bleichrodt and Paraschiv (2007; p.1662) provide a clear statement of the ‘exchange rate assumptions’ used to define loss aversion in the literature. For instance, Fishburn and Kochenberger (1979) and Pennings and Smidts (2003) defined loss aversion as  $U'(-x)/U'(x)$ , Tversky and Kahneman (1992) as  $-U(-1)/U(1)$ , Bleichrodt, Pinto and Wakker (2001) as  $-U(-x)/U(x)$ , and Schmidt and Traub (2002; p.235) as  $U(x)-U(y) \leq U(-y)-U(-x) \forall x>y \geq 0$ . 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,  $U'_-(0)/U'_+(0)$ , as proposed by Köbberling and Wakker (2005) and Booij and van de Kuilen (2009). But this has the very unfortunate effect, as emphasized by Wakker (2010; p. 247), that global properties of loss aversion are being driven by very, very local properties of estimated utility functionals, and that puts a great strain on empirics and functional form assumptions.

Let us immediately note an 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, for which production of appropriate evidence will be correspondingly demanding. This can only increase the risk involved in offering policy advice to a would-be nudger based on such an assumption.



What if the decision weights for the gain domain differ from the probability weighting functions 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.

#### D. Evidence from the Laboratory

Harrison and Swarthout (2016) report an extensive series of 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, that provides some 'stress tests' of the axioms underlying the EUT model, and that allows estimation of a wide range of risk preferences. The first criterion means that the battery must have gain framed lotteries, loss framed lotteries, and mixed frame lotteries. The terms 'gain' and 'loss' refer here to lotteries in which all prizes are (weakly) gains or losses, and the term 'mixed' refers here to lotteries in which some prizes are (strictly) gains and some are (strictly) losses. The second criterion means that the battery must include some sets of choices that generate sharp predictions under EUT, such as the classic Allais Paradox set of two choices, and the classic Common Ratio set of two choices. The third criterion means that the battery must recognize that certain risk preferences could make individuals indifferent between the two lotteries in any given choice, and hence generate low power tests of EUT or RDU. And it also means that one should try to generate stakes that are as large as possible, within obvious feasibility constraints of budgets.

A battery of 100 binary choices was developed to meet these criteria. There were 177 undergraduate students and 94 MBA students sampled from the Georgia State University population. The only differences between the two samples were the stakes, with the domain of net prizes for the undergraduates spanning \$0 up to \$70, and spanning \$0 up to \$750 for the MBA students. The battery deliberately included a number of prize contexts for the MBA students that were identical to the domain of prizes that the undergraduates faced, so the pure effect of stake size could be ascertained. 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 is one major finding of relevance here: 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 devastating 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, non-nested hypothesis tests or mixture models (Harrison and Rutström 2009), CPT is a *descriptively* inferior model of behavior. We therefore focus solely on EUT and RDU in our experiments.

### *3. Helping investment product retailers give better customer advice*

In 2014 we accepted a commission for research from a major South African retailer<sup>4</sup> of household investment products (primarily mutual funds, in American terminology). The company's motivation in commissioning the research began from its observation, which is nearly universal in the industry, of many clients buying products that were sensible investments, given the clients' stated savings and earnings goals, only assuming tolerance for pre-specifiable ranges and average durations of decline in net product value, and then selling back the product, or compounding losses by churning its portfolio elements, upon encountering the predicted episodes of decline. The company hoped to reduce the extent of this behavior. In general, a company can seldom expect to maximize its sales volumes, customer base, or brand reputation if many of its customers systematically fail to derive full value from its products due to misuse. Investment portfolios can be unusual where this relationship is concerned, however, because volumes of commissions to providers and their agents are typically driven up, rather than down, when clients over-churn. This incentive to encourage, or not fully discourage, client over-activity is countered by losses of business when disappointed clients withdraw their funds altogether. Over-churning by large proportions of clients can in extreme cases disrupt the performance metrics on a company's funds. We had no access to our client's accounts, so we cannot comment on the mixture of self-interest and social responsibility in its motivations for wishing to see more of its customers behave in a way that optimized their expected returns. But given the prominence of our client's brand, we would be surprised if social responsibility were not a relevant factor.

The company hypothesized that its customers might show greater resilience during periods of portfolio value decline if, when they chose their portfolios, they were presented with richer information about the histories of net value movements in the set of alternative products, formatted in a way thought to correspond to widespread patterns of cognitive adaptedness. The need for us to guard our client's intellectual property precludes our describing this informational and representational intervention in any detail, which would in any event be superfluous to our immediate concerns. It suffices to say that our research consisted mainly in designing, administering, and analyzing a controlled trial of a prototype of the intervention.

The specific research design involved a sample of 193 subjects, who for reasons of convenience related to budget constraints were employees of the University of Cape Town (UCT). For each subject we estimated their aversion to risk on the basis of procedures described below, and then assigned them randomly to one of two treatments. Subjects chose simulated investment funds modeled on products available in the South African market, and received payment based on the simulated performance of the fund they chose. Subjects in the control treatment received names of investment funds with basic information on each fund, specifically: fund objective, return history, and risk measures (standard deviation and maximum drawdown). Subjects in the treatment group were additionally provided with our client company's 'education

---

<sup>4</sup> Our not naming the company is part of a general policy observed here of censoring information that explicitly or implicitly reveals commercially valuable results of our research furnished to our client. This precludes our describing any results in terms of monetary magnitudes.

intervention' that allows users to engage interactively with the site in order to identify what the client regards as the normative fund type conditional on the user's demographic profile and financial situation and goals.

To avoid uncontrolled interaction between laboratory objects and subjects' varying knowledge of real-world objects, we designed simulated funds based on the principle that informed the original design of mutual funds available to retail consumers in South Africa. We coined names for the simulated funds that mimic those used by their providers. The expected performance of each simulated fund was based stochastically on the historical performance and volatility of the real funds that furnished their models.

Following an increasingly widely used methodology introduced by Hey & Orme (1994), each subject additionally made a series of choices between pairs of lotteries that had an expected (average) yield of 300 South African Rand (R300, which exchanged for about US\$27 at the time of the experiment) for the subject. The data generated by performance of this task allowed us to estimate the structures of risk preferences for each subject. For the simulated market choices, each subject received a pre-experiment endowment of R150. The simulated market was designed to be moderately bullish, such that the average take-home per subject from this part of the study would be R250.<sup>5</sup>

The analysis of subjects' choice data was aimed at investigating factors that were conjectured to influence subjects' preferences over passively received versus interactively explored information about funds, and their disposition to accept normative guidance. Possible factors to be examined included the structure of risk preferences and extent of risk aversion as revealed by the lottery choice experiment, clarity and confidence of subjects' beliefs about the risks associated with alternative investments and basic demographic factors: age, sex, 'race' (as defined by South African affirmative action policies), level of education, income and wealth.

The practical aim of the research was to furnish our client with information about the effect on average choice behaviour of furnishing subjects with their interactive 'education intervention,' and about varying responses to this information associated with the factors indicated above. We hoped that such information might help the client to present information and design interactive guidance initiatives that would effectively improve investor decision-making, and in so doing improve the savings achievements of South African households and individuals.

As our critical discussion in subsequent sections will depend on appreciation of the assumptions underlying our data analysis, we describe the Lottery and Investment tasks administered to subjects in detail sufficient for the reader to understand the structure of the data they yielded.

In the Lottery Task, 50 pairs of lotteries were chosen at random from a set of 100 pairs and presented to the subjects sequentially on computer screens in the form of pie

---

<sup>5</sup> Subjects also made predictions of future events, indicating their degrees of confidence in their predictions, and were rewarded with cash payments of up to R100 when their predictions were correct, with rewards reduced commensurately with subjects' confidence levels. As analysis of the results of this task will not figure in the discussion here, we will pass over the details of its design.

charts. Figure 1 provides a sample of the form of presentation shown to subjects. The lotteries in the choice pairs were composed of monetary outcomes, with 25 possible monetary values and 51 possible associated probabilities. The subjects were asked to choose one lottery from each pair by clicking on the corresponding button below their preferred lottery. Subjects were informed that one of the 50 choices they made would be selected at random for payment after all 50 choices had been made by using 2 ten-sided dice. The selected lottery would be played out by rolling the 2 ten-sided dice again, selecting the outcome of the lottery that corresponded to the roll of the dice.

[Figure 1 about here]

Lottery tasks similar to the ones employed here have been used in previous studies to estimate risk preferences for individuals, typically using maximum likelihood estimation (Harrison & Ng 2016).

In the Investment Task subjects were endowed with R65 and presented with 8 possible simulated funds in which they could invest their endowment. Each of these 8 funds represented some approximation to a financial product to which subjects could potentially have access through a brokerage. Each simulated fund can be thought of as a discretised lottery of the continuous distribution of historical returns associated with the real-life counterpart of the simulated fund in question. That is, each simulated fund had 50,000 possible outcomes, each with an equal probability of occurring. The 8 simulated funds were composed of 4 types: high equity, medium equity, low equity, and interest bearing. There were two simulated funds per group in the choice set, representing the existence of competing products in the actual marketplace.

Before the subjects made any choices in this task, it was explained to them that the task involved choosing an investment portfolio that would be played out against a simulated market. This market was represented by the 50,000 possible states of the world to which the real-world funds were mapped in discrete intervals. Subjects were told that, for practical reasons, one of these 50,000 states would be randomly selected to calculate their investment earnings for their experimental session before they had made the choices for this task. Subjects also were told that they could not view the 50,000 simulations before they made their choices because then they could simply choose the best-performing fund, but they were welcome to look through them after the experimental session if they wished to verify their earnings. No subject chose to perform this verification, suggesting that subjects trusted the fairness of the procedure. Die-rolling by subjects was used to select one of the simulated markets.

The task was conducted through a web browser that showed one of two displays, depending on whether a subject was in the control group or the treatment group. The first page for both groups was a screen informing them that they would be allocated a certain amount of money, and asked to invest that money in a fund or mix of funds. This page also described the different types of funds, but did not give details on their potential returns. Upon clicking a 'continue' button, subjects in the treatment group were presented with a page that contained the client's interactive 'education intervention' that allowed subjects to explore details about the histories of the funds, in formats hypothesized to be cognitively accessible. Subjects in both the treatment and control groups were then taken to the page that allowed them to allocate their endowments to funds. On this page, all 8 funds were listed along with some base level of

information about the potential returns for each fund. This information included the expected 3-year and 5-year returns, the standard deviation of yearly returns, and the maximum drawdown of each fund. Subjects were asked to select as many funds as they wished, and allocate either a Rand amount to each fund, or a percentage of their endowment to each fund. After the subjects were comfortable with their allocations, they clicked a 'submit' button and were taken to the final page showing their earnings for this task.

After each subject had completed all of their experimental tasks, a research assistant tallied their earnings on a record sheet and then paid them in cash in a secluded corner of the lab.

Subjects' risk preferences were analysed based on the Lottery Task. A subject's risk attitude (i.e., risk averse, risk neutral, or risk loving), or preference, can only be estimated from observed choices relative to the more general structure of his or her utility function where risk is concerned. For reasons indicated in Section 2, we conducted analysis based on the assumption that each subject's behaviour was either best characterized by EUT or by RDU. When a subject was estimated to be an RDU agent, based on his or her lottery choices, we tested further to determine which of several probability weighting functions best characterised the structure of his or her pessimism about probabilities.

We can use the results from a specific subject to illustrate the type of risk preferences estimated. Consider subject #22. We first have to determine if subject #22 should be classified as an EUT or RDU decision-maker. The log-likelihood value calculated for the best RDU model (-27.0) is better than the log-likelihood of the EUT model (-28.9), so the subject would be classified as RDU with Prelec probability weighting function by this metric. The difference in log-likelihoods, however, is numerically quite small. Once we test for the subject being EUT, the null hypothesis cannot be rejected at the 5% or 1% significance level, since the  $p$ -value is 0.099; it would be rejected at the 10% level. Thus the classification of this subject depends on the significance level used. Appendix B, available at [http://cear.gsu.edu/wp-2016\\_06-varieties-paternalism-heterogeneity-utility-structures/](http://cear.gsu.edu/wp-2016_06-varieties-paternalism-heterogeneity-utility-structures/) further documents the analysis used to determine whether a subject was better characterized as an EUT or an RDU agent.

If the sole metric for deciding if a subject was better characterised by EUT or RDU were the log-likelihood of the estimated model, then there would be virtually no subjects classified as EUT since RDU nests EUT. But if we use metrics of 10%, 5% or 1% significance levels on the test of the EUT hypothesis that  $\omega(p) = p$ , then we classify 50%, 57% or 67%, respectively, of our 193 subjects with valid estimates as being EUT-consistent. Figure 2 displays these results using the 5% significance level. The left panel shows a kernel density of the 193  $p$ -values estimated for each individual and the EUT hypothesis test that  $\omega(p) = p$ ; we use the best-fitting RDU variant for each subject. The vertical lines show the 1%, 5% and 10%  $p$ -values, so that one can see that subjects to the right of these lines would be classified as being EUT-consistent. The right panel shows the specific allocation using the representative 5% threshold. So 5% of the density in the left panel of Figure 2 corresponds to the right of the middle vertical line at 5%.

[Figure 2 about here]

We now turn to the data generated by the Investment Task. Our aim in the analysis of subjects' investment choices was to identify whether the information provided under the treatment, our client's education intervention, had a significant effect in reducing what we refer to, and described to our client as, subjects' 'welfare loss'. The significance of this interpretation of the analysis will be critically revisited below.

We made it explicit to our client that we viewed welfare loss as the difference between the certainty equivalents of the optimal portfolio conditional on risk preferences and the certainty equivalent of the actual portfolio chosen. The certainty equivalent (CE) is the certain, non-risky return that is equivalent in terms of a subject's subjective utility to the expected utility or (alternatively, depending on the subject) rank-dependent utility of the risky return. We used the estimated expected utility or rank-dependent functionals for each subject to calculate the CE. This approach to welfare evaluation follows Harrison and Ng [2016].

In estimating portfolio optima, we used a bootstrapping method, which we made less computationally intensive by optimizing over a grid of parameter values intended to map the range of feasible estimates, and then interpolating the bootstrapping procedure. Based on the distribution of point estimates of parameters, taking into account standard errors, we optimize portfolio allocations for the following parameter values: EUT:  $r = (0, 0.05, 0.1, \dots, 2, 2.5, 3, 3.5)$ ; RDU Power:  $r = (-10, -5, -3, -2, -1, 0, 0.1, 0.2, \dots, 1, 1.25, 1.5)$  and  $\gamma = (0.2, 0.7, 1.2, \dots, 3.2, 4, 5)$ ; RDU Inverse-S:  $r = (-10, -5, -3, -2, -1, 0, 0.2, 0.4, \dots, 1.6)$  and  $\gamma = (0.3, 0.4, 0.5, \dots, 1.1)$ ; RDU Prelec:  $r = (-10, -5, -3, -2, -1, 0, 0.25, 0.5, \dots, 2)$ ,  $\eta = (0.3, 0.8, 1.3, \dots, 2.8)$ , and  $\phi = (0.5, 0.7, 0.9, 1.1, 2, 3)$ .

Figure 3 displays the risk-return tradeoff from the simulated funds in the investment task. The return is the average of the annualized returns on the fund, and the risk is the standard deviation of the annualized returns on the fund. The returns here come from 50,000 simulations of fund performance, based on historical data on returns. We observe that for higher average returns the investor must be willing to take on greater risk, which is no surprise. But in some cases the extra return only entails a minimal increase in risk: for instance, compare the X123 Equity fund with the ABC Multi High fund. The evaluation of these increments in risk, exchanged for increments in return, depends on the attitude to risk of the investor, if we assume that the subjective risk perceptions of the investor match these historical returns.

For each of the high, medium and low equity asset classes, the historical performance of a mutual fund in each class was derived from returns for the whole asset class.<sup>6</sup> The second funds in each of the high and medium equity classes were simulations of real funds traded in the South African market. For the low equity fund, historical performance of the fund was equated to the historical inflation movement plus 5%. The interest bearing funds were derived from historical data using the interest

---

<sup>6</sup> We did not group asset classes based on subjective judgment. They were defined as per Association for Savings and Investment South Africa categories used by financial advisors.

bearing variable term funds and money market funds, respectively, also retailed in South Africa.

Month-end price data from June 2001 to August 2014 were used to determine the funds' performance parameters such as historical returns and standard deviation of returns. This period included the bull run of 2006/2007, the global financial crisis of 2007/2008, and the recovery period post-2008.

[Figure 3 about here]

Figure 4 shows the number of funds that received *some* allocations of the R65 subjects had available to invest. There is a clear mode at 2 funds, with very few subjects investing in more than 4 funds. Relatively few subjects chose to invest all of their money in one fund. Of course, this does not show us whether the funds invested in were optimal or how sub-optimal they were.

[Figure 4 about here]

The optimal allocation to equity funds was relatively easy to characterize. Using the relative risk aversion ( $r$ ) as a summary, descriptive measure of the risk premium, we found that 100% of the endowment of R65 would optimally have been allocated to the ABC Company Equity Fund for all values of  $r$  up to 0.62, and then that fraction declines to about 50% as  $r$  approaches 1. The residual is entirely the 123 Company Equity Fund. The vast bulk of estimates of relative risk aversion in the laboratory are around 0.65, with some variation of course (see Harrison and Rutström [2008] for a survey).

Figure 5 shows the average allocation of investment funds to each fund, where the total that could be invested was R65. We show a vertical red line at the 50% mark for reference.<sup>7</sup> In this display the funds are ordered in terms of smallest (average) allocation to largest, so one has to pay attention to the names of the funds. For the averages we see that the two equity funds received the highest average allocation, but that the 123 Company Equity Fund was only the third most popular in terms of median allocations.

Figure 5 also displays the average allocations to all funds in comparison to the optimal allocations. Since we find that all optimal allocations should be to the two equity funds, we aggregate these funds and show the optimal allocation as R65, or 100% of the portfolio. The remaining funds should always receive a zero allocation. Viewed in this light, and ignoring the optimality of the allocation within equity funds, we can see that the average investor was making a qualitatively optimal investment, with the majority of allocations to the equity funds. However, the level of allocations falls short of the optimal amount of R65. The distance between the average observed allocations and the optimal allocations is what generates the welfare losses we reported. These distances only tell us that there will be a welfare loss on average: to evaluate the significance of that loss we evaluated the foregone CE from the observed portfolios compared to the CE of the optimal portfolio.

[Figure 5 about here]

---

<sup>7</sup> The median allocations are close to the average allocation except for the Equity Fund. In that case the median is exactly R32.5, or 50% of the portfolio.

Each CE calculation uses 50,000 draws from the multivariate normal distribution underlying the simulated funds. These CE are conditional on estimates of the parameters defining risk preferences, and the uncertainty of the estimates is allowed for by sampling 500 draws from the joint parameter distribution. The means of these 500 draws are the parameter point estimates based on the winning risk preference structure model for the individual at the 5% significance level, and the covariance matrix between the parameter estimates.

Multivariate normality of the joint parameter distribution is assumed, which is potentially problematic with large standard errors for some subjects: very high or low estimates of probability weighting parameters give rise to implausible decision weight schemes, and very high or low estimates of the relative risk aversion coefficient give rise to numerical overflow. Simulated values of risk preference parameters were accordingly constrained within the following bounds: EUT:  $r \in [-5,5]$ ; RDU Power:  $r \in [-10,10]$ ,  $\gamma \in [0.2,5]$ ; RDU Inverse-S:  $r \in [-10,10]$ ,  $\gamma \in [0.3,3]$ ; RDU Prelec:  $r \in [-10,10]$ ,  $\eta \in [0.3,3]$ ,  $\phi \in [0.3,3]$ .

Welfare loss calculations could be performed for 174 of the 193 subjects. The remaining 19 were those for whom a winning model could not be assigned because the estimated coefficient of relative risk aversion was arbitrarily close to one. Negative welfare losses are calculated in several instances, because of the inaccuracies of the multilinear interpolation method, giving rise to a portfolio which is sub-optimal and yielding a lower CE than the actual allocation chosen.

Each of the 500 simulations presents a set of risk preference parameters, conditional on which welfare loss can be calculated. For each of these simulations, a  $t$ -test can reveal whether the mean welfare loss is significantly lower for the treatment group than for the control group. We allow for the error with which risk preference parameters are estimated by performing the test for each simulation and examining the distribution of test results.

Figure 6 displays the average welfare loss, in Rand, for each subject for which we could generate valid estimates of risk preferences and optimal portfolios conditional on those risk preferences. Truncating a small fraction of welfare losses greater than R300, we observe that the density of welfare losses is much smaller under the Education Intervention Treatment than under the Control. Hence we conclude that the Education Intervention Treatment leads to better decisions being made about investment in this setting, designed to mimic, under controlled conditions, the natural setting in which the intervention will be applied.

[Figure 6 about here]

Figure 7 shows that the Education Intervention Treatment did not generate a greater dispersion in welfare losses. This is useful to know, since this might have mitigated the benefits of the reduction in the average of welfare losses.

[Figure 7 about here]

Figures 8 and 9 show that the Education Intervention Treatment had benefits for both EUT and RDU decision-makers, but that the benefits for the RDU decision-makers



are much larger. In part, this is because the RDU decision-makers suffered greater welfare losses even in the Control.

It is easier to evaluate the total and marginal effects of various demographics and treatments using descriptive statistical methods such as a regression of average welfare loss. When the right-hand-side covariate is just the demographic characteristic or treatment dummy variable we evaluate the 'total effect' of the covariate, which is the effect taking into account all of the correlated effects of covariates that also vary with the covariate of interest. For example, if women are younger than men in our sample, then the total effect of women will also include any effect of being a woman and being younger. When the right-hand-side covariates are all demographic characteristics and treatment dummy variables we evaluate the 'marginal effect' of the covariate. Both total effects and marginal effects are of interest, and answer different questions.

Figure 10 displays the total effect of each characteristic and treatment, sorted by the size of the effect. The Education Intervention Treatment is shown in bold. Figure 11 displays the marginal effect of each characteristic and treatment. In both cases we see a significant effect of the Education Intervention Treatment to reduce welfare losses. We also see, in both cases, a significant effect, to increase welfare losses, of the subject being classified as violating EUT.

[Figure 8 about here]

[Figure 9 about here]

[Figure 10 about here]

[Figure 11 here]

The average of the difference in mean welfare loss between control and treatment groups across the 500 simulations is R57.28 (median = R56.23) with standard deviation R17.98. Welfare loss was lower for the treatment group in all 500 simulations. A one-sided test, with the alternative hypothesis being that welfare loss is lower for the treatment group than for the control, yields a  $p$ -value  $< 0.05$  in 392 of the 500 simulations. The  $p$ -value is  $< 0.1$  for 460 simulations.

In our concluding advice to our client, we emphasized that the value of their Education Intervention, measured in terms of client welfare, would depend on the proportion of RDU agents in their customer population. As our experimental subject pool was not representative of this population, we suggested that they might wish to run the Lottery Task on a large, randomly selected sample drawn from their client demographic. Generalizing this advice, our policy-relevant opinion is that the expected presence of significant numbers of people in South Africa whose risk preference structure is well characterized by an RDU structure is a main source of scope for investments in education about comparative details of portfolio risk structures to raise the frequency with which South Africans reach retirement with savings that better approximate available potentials.

#### *4. Are we nudging or are we boosting?*

At first glance, the recommendation we made to our client concerning application of their Investor Education Intervention, based on our experimental results,

might look like a prime case of boosting. If our advice were followed, investors would be presented with information about historical fund performances, in a format that would increase the likelihood that their decisions will optimize their returns, reducing the probability that their savings goals will be frustrated. The intervention directly improves the decision-making resources of the investor, especially the investor with a RDU risk preference structure, and might plausibly create rationality spillovers as discussed earlier. In particular, people familiarized with the richer information might be motivated to seek it out when they make other financial decisions under risky conditions. The intervention does not manipulate the targets in the straightforward sense of altering their environments without their knowledge.

On deeper reflection, however, matters aren't so clear-cut. The first three columns of Table 1 are taken from Grüne-Yanoff & Hertwig's (2016) (GYH) discussion of the differences between nudging and boosting. In the fourth column we add our assessment of the fit of this taxonomy to the recommendation we made to our client concerning application of their Investor Education Intervention. If we were to treat GYH's table as providing eight (non-exclusive) criteria for distinguishing a nudge from a boost, then our recommended policy would emerge as an exact hybrid, matching a nudge on four criteria and a boost on the other four.

Our assessments in the fourth column require some explanation and justification. Where the first row is concerned, the investors have historically not been able to infer that they decided in error until, arguably, well after the fact. Even then, according to our client, most did not attribute their early selling of their funds to any error made by them, though they sometimes expressed disappointment in the provider or advisor. But in general our advice does not rest on the assumption that any investors are ever aware of any errors. The suggestion is rather that information about historical distributions of fund values make people who reveal RDU risk preference structures behave more like people with EUT risk preferences. With respect to the second row, clearly the intervention is motivated by the client's view that many investors choose in such a way as to undermine their own welfare, as attributed based on their observed behavior, but can be induced to alter their decisions in at least a significant proportion of instances. We take our assessment in the third and fourth rows to be obvious: the main point of the further experimental evidence we urged our client to obtain is to gain richer knowledge of the structure of their customers' preferences (i.e., RDU or EUT), and of the distribution of non-EUT preferences. Clearly this implies, as per the fifth row, that the experts are less error prone than the investors, and it is far from clear that it would be generally efficacious for the experts to try to explain the differences between RDU and EUT preference structures to investors. Where the sixth row is concerned, as discussed earlier we suspect that our client is benevolent about investors' welfare to some extent, but this motivation is not necessary, as it is in the investment house's interest for customers to maintain their investments through market downturns. Finally, the intervention is only efficacious to the extent that investors are able and motivated to be influenced by carefully designed representations of more complete information to choose in ways that better approximate what they would choose were they expected utility optimizers.

**Table 1** Eight assumptions of the nudge and boost approaches

	Nudge	Boost	Investor Education Intervention
<i>Cognitive error awareness</i> Must the decision maker be able to detect the influence of error?	No	Yes	No
<i>Cognitive error controllability</i> Must the decision maker be able to stop or override the influence of the error?	No	Yes	Yes
<i>Information about goals</i> Must the designer know the specific goals of the target audience?	Yes	No	Yes
<i>Information about the goals' distribution</i> Must the designer know the distribution of goals in the target audience?	Yes	No	Yes
<i>Policy designer and cognitive error</i> Must experts be less error-prone than decision makers?	Yes	No	Yes
<i>Policy designer and benevolence</i> Must the designer be benevolent?	Yes	No	No
<i>Decision maker and minimal competence</i> Must the decision maker be able to acquire trained skills?	No	Yes	Yes
<i>Decision maker and sufficient motivation</i> Must the decision maker be motivated to use trained skills?	No	Yes	Yes

Our assessments in the fourth column require some explanation and justification. Where the first row is concerned, the investors have historically not been able to infer that they decided in error until, arguably, well after the fact. Even then, according to our client, most did not attribute their early selling of their funds to any error made by them, though they sometimes expressed disappointment in the provider or advisor. But in general our advice does not rest on the assumption that any investors are ever aware of any errors. The suggestion is rather that information about historical distributions of fund values make people who reveal RDU risk preference structures behave more like people with EUT risk preferences. With respect to the second row, clearly the intervention is motivated by the client's view that many investors choose in such a way as to undermine their own welfare, as attributed based on their observed behavior, but can be induced to alter their decisions in at least a significant proportion of instances. We take our assessment in the third and fourth rows to be obvious: the main point of the further experimental evidence we urged our client to obtain is to gain richer knowledge of the structure of their customers' preferences (i.e., RDU or EUT), and of the distribution of non-EUT preferences. Clearly this implies, as per the fifth row, that the experts are less error prone than the investors, and it is far from clear that it would be generally efficacious for the experts to try to explain the differences between RDU and EUT preference structures to investors. Where the sixth row is concerned, as discussed earlier we suspect that our client is benevolent about investors' welfare to some extent, but this motivation is not necessary, as it is in the investment house's interest for customers to maintain their investments through market downturns. Finally, the intervention is only efficacious to the extent that investors are able and motivated to be influenced by carefully designed representations of more complete information to choose in ways that better approximate what they would choose were they expected utility optimizers.

The general diagnosis of the hybrid nature of the intervention as between nudging and boosting lies in the epistemic status and the normative presuppositions of the economic experts (i.e., us). With respect to the former, we have technical knowledge about the relationship between objective risk and subjective preference structures that investors lack, and that would be difficult to directly explain to most of them, let alone to directly inspire through exhortation. Concerning normative presuppositions, we assume that by revealing preferences in relatively simple decision contexts, choices between risky lotteries, people provide an informational basis for assessing the implications for their own welfare of decisions in more complicated circumstances.

This follows the approach exemplified and promoted by Harrison and Ng (2016), when they evaluate the welfare gain 'introduced into the world' by a standard type of indemnity insurance product. If we can reliably estimate the distribution of risk preferences among individuals, and the distribution of their subjective beliefs about loss contingencies and likelihood of payout, there is a certainty equivalent of a risky insurance policy that can be compared to the certain insurance premium. This simple logic extends to non-standard models of risk preferences, such as RDU, in which some people exhibit 'optimism' or 'pessimism' about loss contingencies in their evaluation of the risky insurance policy.

Harrison and Ng (2016) illustrate the application of these basic ideas about the welfare evaluation of insurance policies in a controlled laboratory experiment, just as we do in the case study reviewed here. They estimate the risk preferences of individuals

from one task, and separately present each individual with a number of insurance policies in which loss contingencies are objective, so there is no issue about subjective beliefs being biased. They then estimate the expected consumer surplus gained or foregone from observed take-up decisions. There is striking evidence of foregone expected consumer surplus from incorrect take-up decisions. This motivates a highly relevant and general policy conclusion, namely, that the metric of take-up itself, widely used in welfare evaluations of insurance products, provides a qualitatively incorrect guide to the expected welfare effects of insurance.

There is a crude revealed preference argument to the effect that if the product is (not) taken up it must have been perceived to be a positive (negative) net benefit. But that is only the starting point of any serious welfare evaluation, particularly if one wants to *quantify* the size of the welfare effect. What if the subjective beliefs were biased, in the sense that the individual would revise them if given certain information? What if the evaluation of the product used some criteria other than EUT? What if the individual simply made a mistaken decision, given beliefs and risk preferences? As noted previously, invoking crude revealed preference implies that one could *never* find a negative welfare from *any* insurance decision. But the alternative to crude revealed preference is a *sophisticated* investigation and application of revealed preference, not abandonment of that approach.

Laboratory experiments provide the ideal environment for setting out all of the information and behavior we need to observe in order to draw inferences about welfare. Once we move to the field and consider naturally occurring data, we then realize what information is missing – in the present case study, the distribution of RDU preferences among investment fund customers – if we want to make interesting welfare evaluations. Laboratory and field experiments are complements, as stressed by Harrison and List (2004).

This general methodological approach allows the economist to draw useful conclusions about what types of decisions led to welfare losses, and to identify demographics that are more likely to make those types of decisions. To illustrate, again from the insurance policy choices considered by Harrison and Ng (2016): out of all purchase decisions made the subjects in their experiment, 60% were associated with a welfare loss. Notably, female subjects had a 9.8 pp higher chance than men of making such excess purchase errors, with a 95% confidence interval between 0 pp and 20 pp. When Harrison & Ng consider the marginal effect of gender, controlling for other demographics, this estimated effect was 11.8 pp with a 95% confidence interval between 1 pp and 23 pp. This type of information allows the economist to recommend structured interventions to improve decisions by targeting certain demographic groups and certain types of errors.

A further potential knowledge gain from welfare assessment based on sophisticated revealed preference experiments in lab and field is that one can rigorously identify which axioms of a normative model of risk preferences fail when one observes expected welfare losses. For instance, are the subjects that suffer losses when faced with an index insurance product those for whom the Reduction of Compound Lotteries axiom fails behaviorally? Precise characterizations of such failures can be identified in experiments (e.g., Harrison, Martínez-Correa and Swarthout 2015), just as the lottery

battery employed in the Investor Education Intervention study allows us to *structurally* identify behavioral failures of the Compound Independence axiom.

There are philosophical lessons to be gained in moving from typical consideration of CPT as the standard non-EUT preference structure, with its emphasis on loss aversion, to RDU, with its emphasis on probability optimism and pessimism. When paternalism, whether 'hard' or 'soft,' is based on the conviction that it is ethically appropriate to help people to be more rational, focus on CPT can raise difficult quandaries. Only very strong normative models pronounce loss aversion *per se* to be *irrational*. By contrast, it is more immediately plausible to suggest that if people make risky financial decisions on the basis of subjective probability assignments that are distortions of objective probabilities, the dominant prediction is that they will obtain an outcome they don't prefer, namely, sub-optimal expected monetary wealth that is not traded off for anything else that they prefer. Le Grand & New (2015, p. 91) note the same distinction, and interpret its implications as we do.

It is this gap between what subjects understand about their behavior in more complex decisions and what they reveal by their behavior in the simpler lottery choices that gives our apparent boosting intervention a nudging aspect. Most people choosing investment funds don't attempt to compute internally represented optima – either from EUT or RDU bases – and then make computational errors that could be pointed out to them. Following Dennett (1987), we assume the intentional stance to make sense of people's overall behavioral patterns, and use the lottery choice experiment as a *relatively* direct source of constraint on the *virtual* preference structures we assign when we perform welfare assessment of their investment fund choices. This *externalism* about preference content, according to which it is less plausible to regard preferences as being literally 'in people's heads,' as their attributed propositional content becomes more tightly refined, blurs the distinction between 'treating' the subject and 'treating' the subject's environment. Our technical tools allow us to identify virtual intentions that most subjects are not able to identify, and that they could not *deliberately* use to evaluate their own decisions. On the other hand, our experiment provides evidence that attention to certain informational patterns induces a significant number of subjects to act as if they were stochastically closer to expected value optimizers. We *boost* their informational access in a way that *nudges* their sub-deliberative cognition.

One might infer from the foregoing that we are urging strong welfarism, the view that well being is generally best promoted by favoring policies that would be chosen if everyone were modeled as having a virtual preference for optimizing expected value. That is *not* what we are arguing. Harrison & Ng (2016) take as given the type of risk preferences each individual in their data set employs, and use that as the basis for evaluating welfare effects of insurance decisions: *periculum habitus non est disputandum*. Given the economist's anti-paternalistic bias as characterized earlier, this is a natural starting point in their particular policy application. Even though the alternatives to EUT were originally developed to relax one of the axioms of EUT that some consider attractive normatively, it does not follow that one is unable to write down axioms that make those alternatives attractive normatively.

In the Investor Education Intervention case, this seems to us less persuasive because there is less distance, from the intentional stance, between the lottery choice task and the investment fund choice task. But in the insurance policy take-up case, some

might argue that RDU is not normatively attractive. Given such heterogeneity of ethical attitudes, it would be a useful exercise to do a calculation of consumer surplus in which we *only* assume EUT parameters for subjects: we could estimate the EUT model and get the corresponding CRRA coefficient estimate (i.e., not just use the CRRA coefficient estimate from the RDU specification). Then we would repeat the calculations as in Harrison & Ng (2016).<sup>8</sup>

### 5. Conclusion and future work

We have argued that there is no short road to a general ethical preference for boosting over nudging, and that part of what blocks the short road are methodological considerations. The existing literature on behavioral welfare economics, including Bernheim (2009), Bernheim & Rangel (2009), Manzini & Mariotti (2012, 2014), Rubinstein & Salant (2012), Salant & Rubinstein (2008) and Sugden (2004) has tended to focus on simple dichotomies between frankly paternalistic interventions and scrupulous *laissez faire*. Economists are not typically called upon to make these kinds of sweeping judgments. We have illustrated, by appeal to a quite widespread consulting role economists are often called upon to perform, how their technical expertise implicates them in more subtle assessments. We urge philosophers who aim to contribute to a richly situational ethics of policy advocacy to attend to such realistic and practically important cases. As we have argued, Grüne-Yanoff & Hertwig's (2016) distinction between nudging and boosting is a helpful conceptual advance, but it represents only the beginning, and not the conclusion, of a more sophisticated and engagement-ready ethics of behavioral welfare economics. The road ahead will not be straight but the vehicles for traveling on it are being steadily refined.

We view the laboratory as the appropriate place to 'wind tunnel' the normative welfare evaluation of new decision scaffolds. Figures 12 and 13 stand as explicit, rigorous objects of 'target practice' for anyone proposing nudges or boosts to improve welfare from informational interventions in financial planning.

### References

- Abdellaoui, M., Bleichrodt, H., & Paraschiv, C. (2007). Measuring loss aversion under prospect theory: A parameter-free approach. *Management Science* 53: 1659-1674.
- Ashcroft, R. (2011). Personal financial incentives in health promotion: Where do they fit in an ethic of autonomy. *Health Expectations* 14: 191-200.
- Bernheim, B. D. (2009). Behavioral welfare economics, *Journal of the European Economic Association* 7: 267-319.
- Bernheim, B.D., & Rangel, A. (2008). Choice-theoretic foundations for behavioral welfare economics. In A. Caplin and A. Schotter, eds., *The Foundations of Positive and Normative Economics: A Handbook*, pp. 155-192. Oxford: Oxford University Press.
- Binmore, K. (2009). *Rational Decisions*. Princeton: Princeton University Press.

---

<sup>8</sup> For subjects best descriptively modeled as EUT choosers there would be no change in the inferred CS, of course.

- 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.
- Booij, A., & van de Kuilen, G. (2009). A parameter-free analysis of the utility of money for the general population under prospect theory. *Journal of Economic Psychology* 30: 651-666.
- Camerer, C., Issacaroff, S., Loewenstein, G., O'Donoghue, T., & Rabin, M. (2003). Regulation for conservatives: Behavioral economics and the case for asymmetric paternalism. *University of Pennsylvania Law Review* 151: 1211-1254.
- Camerer, C., Loewenstein, G., & Prelec, D. (2005). Neuroeconomics: How neuroscience can inform economics. *Journal of Economic Literature* 43: 9-64.
- Cherry, T., Crocker, T., & Shogren, J. (2003). Rationality spillovers. *Journal of Environmental Economics and Management* 45: 63-84.
- Conly, S. (2013). *Against Autonomy: Justifying Coercive Paternalism*. Cambridge: Cambridge University Press.
- Cox, J. & Sadiraj, V. (2008). Risky decisions in the large and in the small: Theory and experiment. In J. Cox & G. Harrison (Eds.). *Risk aversion in experiments*. Bingley, UK: Emerald.
- Dennett, D. (1987). *The Intentional Stance*. Cambridge, MA: MIT Press.
- Fishburn, P., & Kochenberger, G. (1979). Two-piece von Neumann-Morgenstern utility functions. *Decision Sciences* 10: 503-518.
- Friedman, M. (1953). *Essays in Positive Economics*. Chicago: University of Chicago Press.
- Gigerenzer, G., Todd, P., & the ABC Research Group (1999). *Simple Heuristics that Make Us Smart*. Oxford: Oxford University Press.
- Grüne-Yanoff, T., & Hertwig, R. (2016). Nudge versus boost: How coherent are policy and theory? *Minds and Machines*, forthcoming.
- Harless, D. (1992). Predictions about indifference curves inside the unit triangle: A test of variants of expected utility. *Journal of Economic Behavior and Organization* 18: 391-414.
- Harrison, G., & List, J. (2004). Field experiments. *Journal of Economic Literature*, 42: 1009-155.
- Harrison, G., Martínez-Correa, J., & Swarthout, J.T. (2015). Reduction of compound lotteries with objective probabilities: Theory and evidence. *Journal of Economic Behavior and Organization* 119: 32-55.
- Harrison, G., & Ng, J.M. (2016). Evaluating the expected welfare gain from insurance. *Journal of Risk and Insurance* 83: 91-120.
- Harrison, G., & Rutström, E. (2008). Risk aversion in the laboratory. In J. Cox & G. Harrison (Eds.), *Risk Aversion in Experiments*, pp. 41-196. Bingley: Emerald.



- 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 2016-05, Center for Economic Analysis of Risk, Robinson College of Business, Georgia State University.
- Hausman, D. (2011). *Preference, Value, Choice and Welfare*. Cambridge: Cambridge University Press.
- Hertwig, R., Hoffrage, U., & the ABC Research Group (2013). *Simple Heuristics in a Social World*. Oxford: Oxford University Press.
- Hey, J.D. & Orme, C. (1994). Investigating Generalizations of Expected Utility Theory Using Experimental Data. *Econometrica* 62(6): 1291–1326.
- John, P., Cotterill, S., Moseley, A., Richardson, L., Smith, G., Stoker, G., & Wales, C. (2011). *Nudge, Nudge, Think, Think: Experimenting With Ways to Change Civic Behaviour*. London: Bloomsbury Academic.
- John, P., Smith, G., & Stoker, G. (2009). Nudge nudge, think think: Two strategies for changing civic behaviour. *Political Quarterly* 80: 361-370.
- Kahneman, D. (2011). *Thinking Fast and Slow*. New York: Farrar, Straus and Giroux.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica* 47: 263-292.
- Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.
- Köbberling, V., & Wakker, P. (2005). An Index of Loss Aversion. *Journal of Economic Theory* 122: 119-131.
- Leamer, E. (2012). *The Craft of Economics*. Cambridge, MA: MIT Press.
- Le Grand, J., & New, B. (2015). *Government Paternalism: Nanny State or Helpful Friend?* Princeton: Princeton University Press.
- Loomes, G., & Sugden, R. (1998). Testing different stochastic specifications of risky choice. *Economica* 65: 581-598.
- Manzini, P., & Mariotti, M. (2012). Categorize then choose: Boundedly rational choice and welfare. *Journal of the European Economic Association* 10: 1141-1165.
- Manzini, P., and Mariotti, M. (2014). Welfare economics and bounded rationality: The case for model-based approaches. *Journal of Economic Methodology* 21: 343-360.
- Pennings, J., & Smidts, A. (2003). The shape of utility functions and organizational behavior. *Management Science* 24: 1251-1263.
- Prelec, D. (1998). The Probability Weighting Function. *Econometrica* 66: 95-113.

- Quiggin, J. (1982). A Theory of Anticipated Utility. *Journal of Economic Behavior and Organization* 3: 323-343.
- Ross, D. (2014). *Philosophy of Economics*. London: Palgrave Macmillan.
- Rubinstein, A., & Salant, Y. (2012). Eliciting welfare preferences from behavioral datasets. *Review of Economic Studies* 79: 375-387.
- Salant, Y., & Rubinstein, A. (2008). (A, f): Choice with frames. *Review of Economic Studies* 75: 1287-1296.
- Samuelson, P. (1937). A note on measurement of utility. *Review of Economic Studies* 4: 154-161.
- Samuelson, P. (1938). A note on the pure theory of consumer's behaviour. *Economica* 5: 61-72.
- Schmidt, U., & Traub, S. (2002). An experimental test of loss aversion. *Journal of Risk and Uncertainty* 25: 233-249.
- Schmidt, U., & Zank, H. (2008). Risk aversion in cumulative prospect theory. *Management Science* 54: 208-216.
- Sugden, R. (2004). The opportunity criterion: Consumer sovereignty without the assumption of coherent preferences. *American Economic Review* 2004: 1014-1033.
- 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. *University of Chicago Law Review* 70: 1159-1202.
- Todd, P., Gigerenzer, G., & the ABC Research Group (2012). *Ecological Rationality: Intelligence in the World*. 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: Cambridge University Press.
- Wilcox, N. (2011). Stochastically more risk averse: A contextual theory of stochastic discrete choice under risk. *Journal of Econometrics* 162: 89-104.
- Yaari, M. (1987). The Dual Theory of Choice under Risk. *Econometrica* 55: 95-115.

Figure 1: A sample lottery choice pair

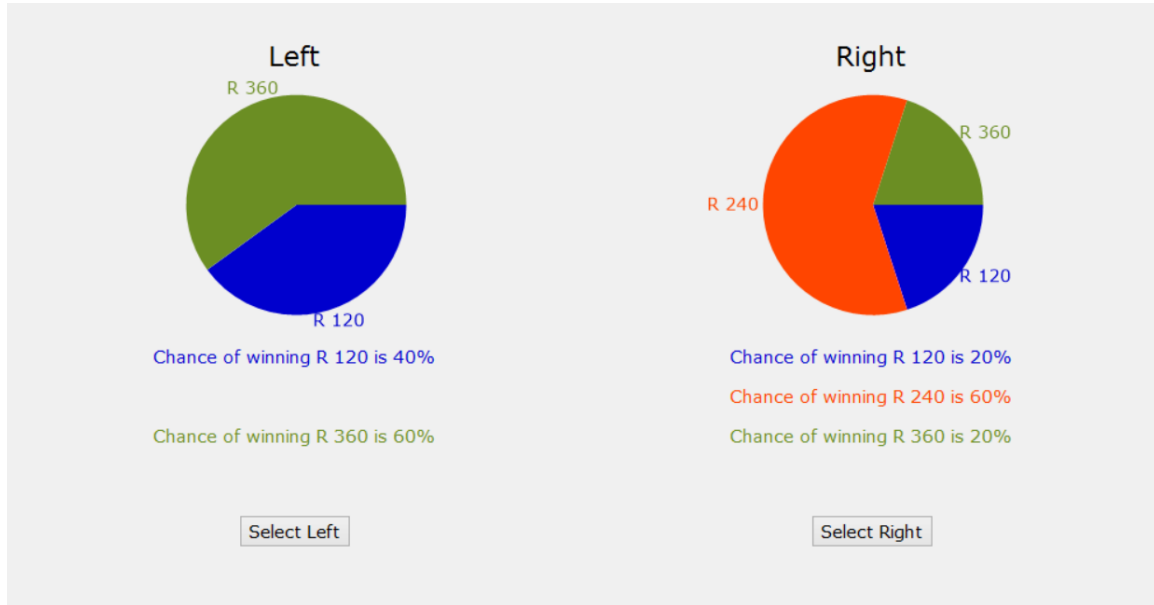


Figure 2: Classifying Subjects as EUT or RDU

N=193, one  $p$ -value per individual  
 Estimates for each individual of EUT and RDU specifications

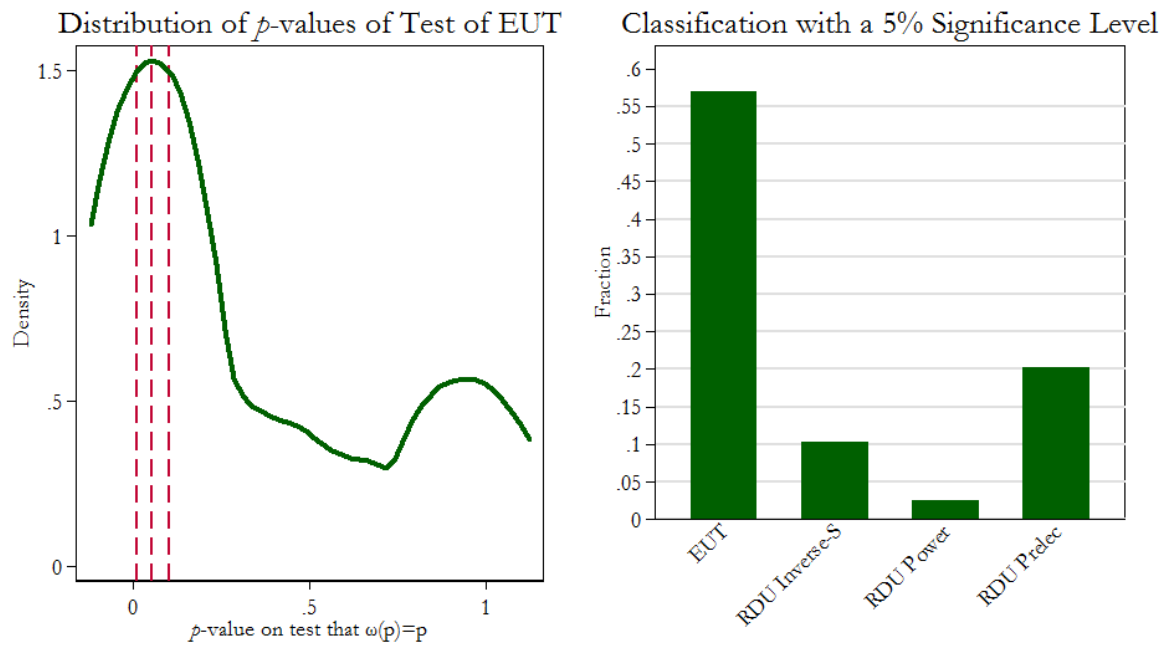
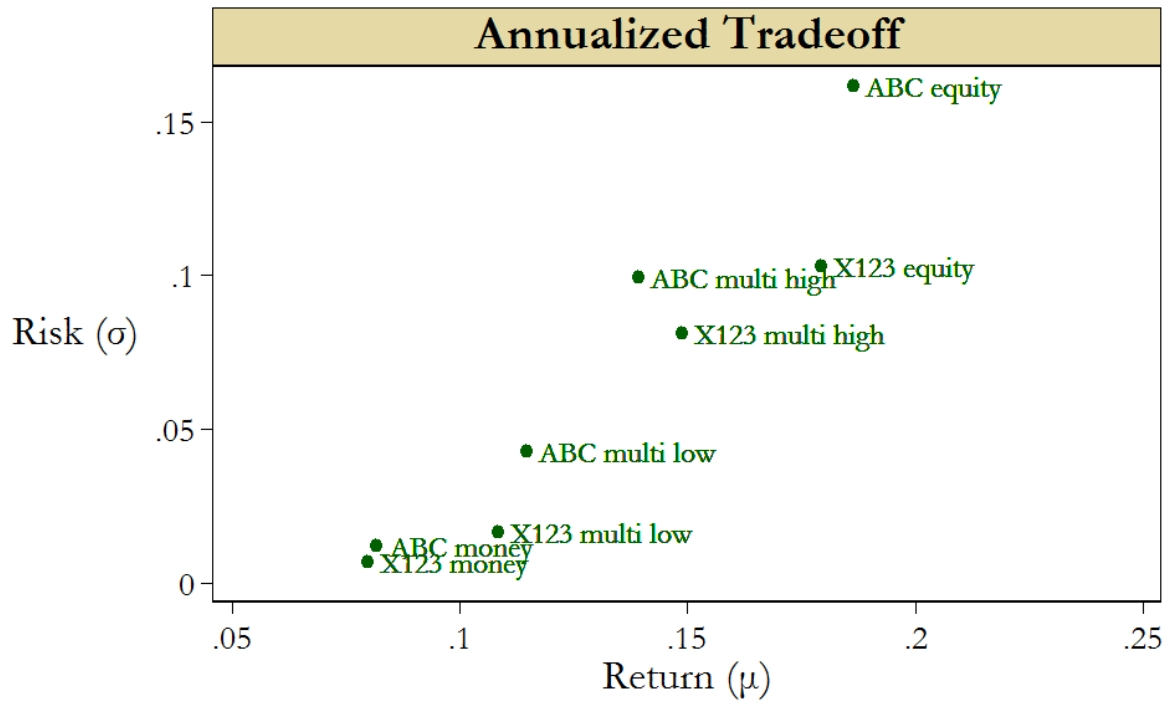
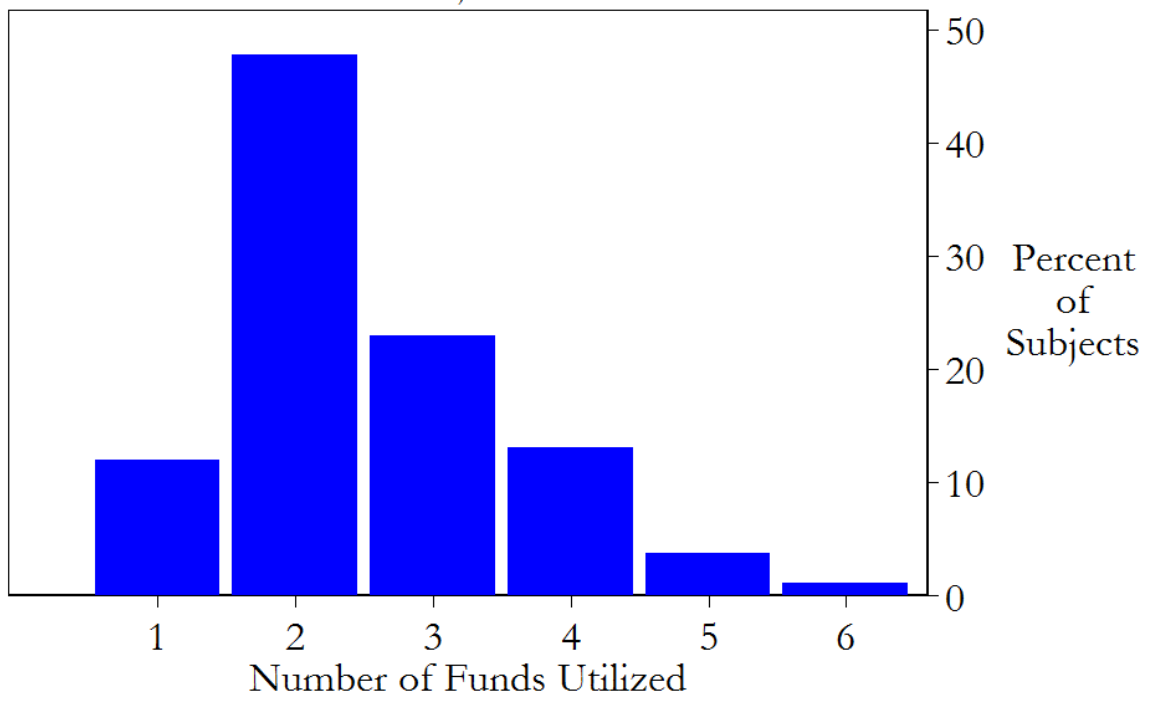


Figure 3: Annualized Risk-Return Tradeoffs in the Investment Task



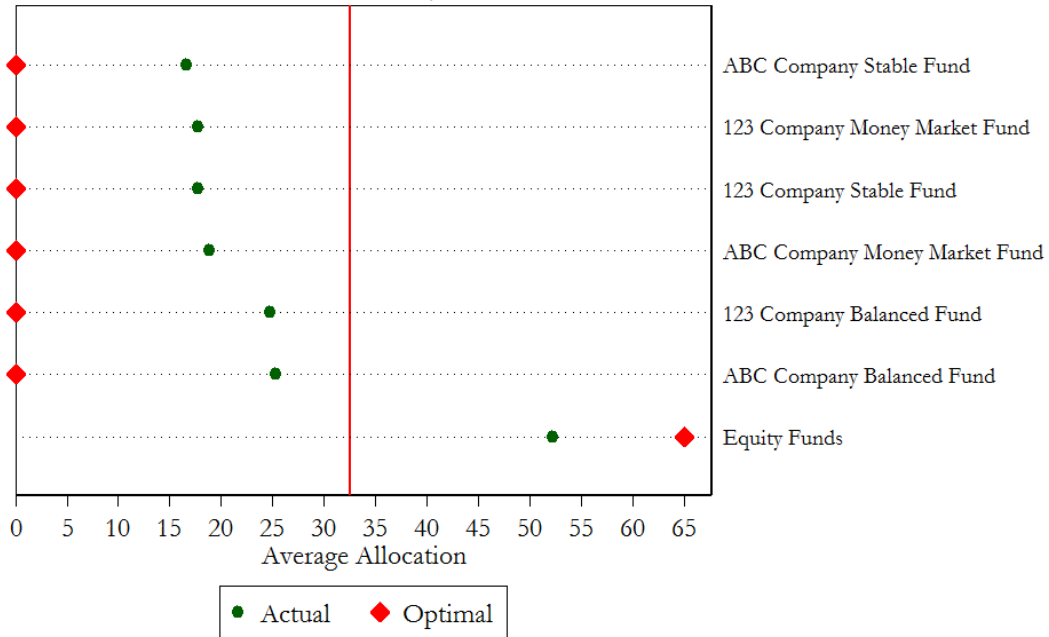
# Figure 4: Number of Funds Utilized

N = 193 subjects from the Investment Task



### Figure 5: Average and Optimal Allocations

N = 193 subjects from the Investment Task



### Figure 6: Average Welfare Loss in Control and Treatment Conditions

N = 174 subjects, with 85 in the Control and 89 in the Treatment

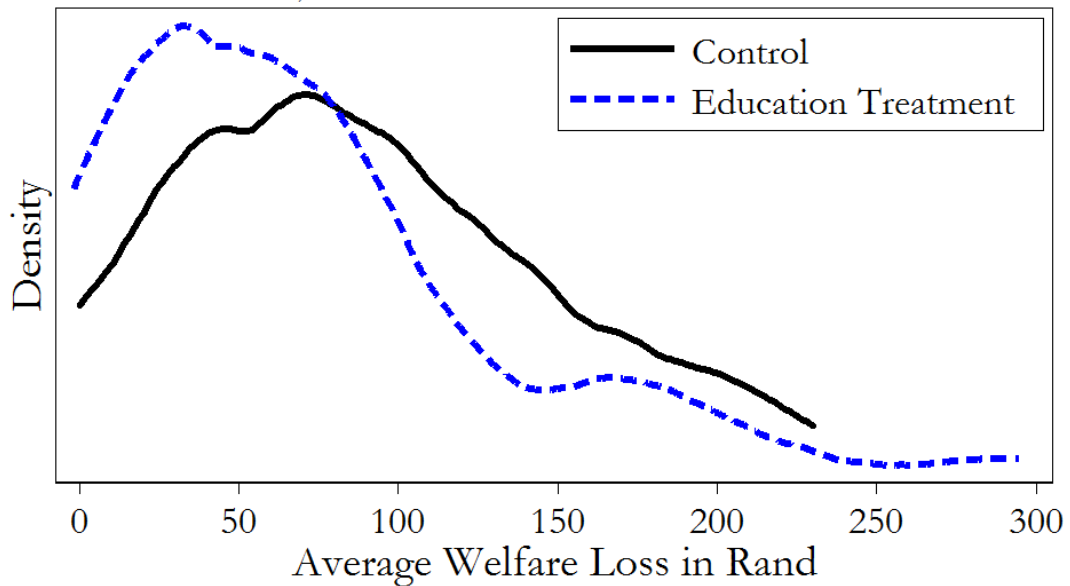


Figure 7: Standard Deviation of Welfare Loss in Control and Treatment Conditions

N = 174 subjects, with 85 in the Control and 89 in the Treatment

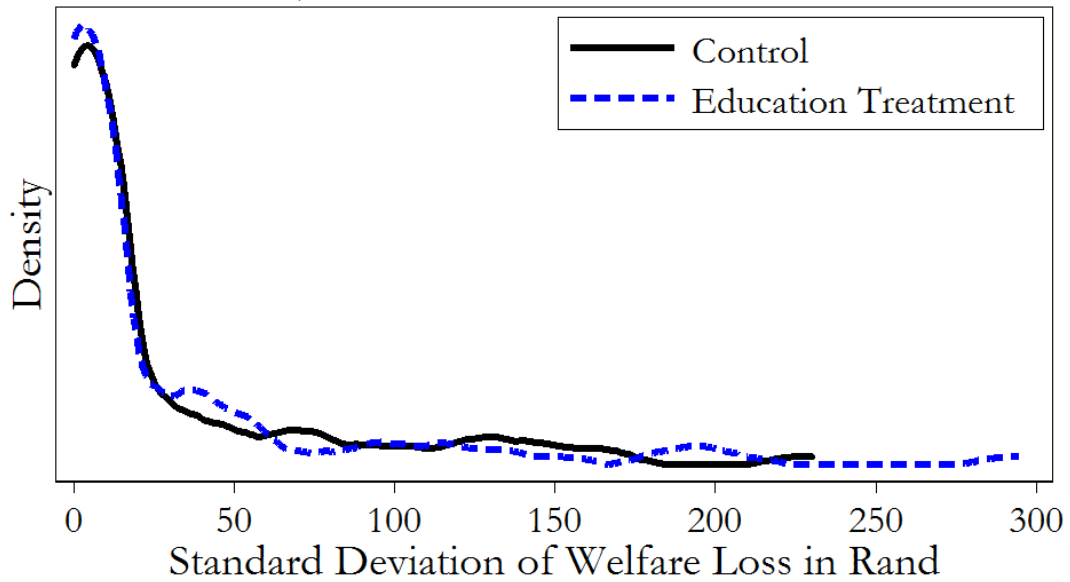


Figure 8: Average Welfare Loss in Control and Treatment Conditions for Expected Utility Theory Subjects

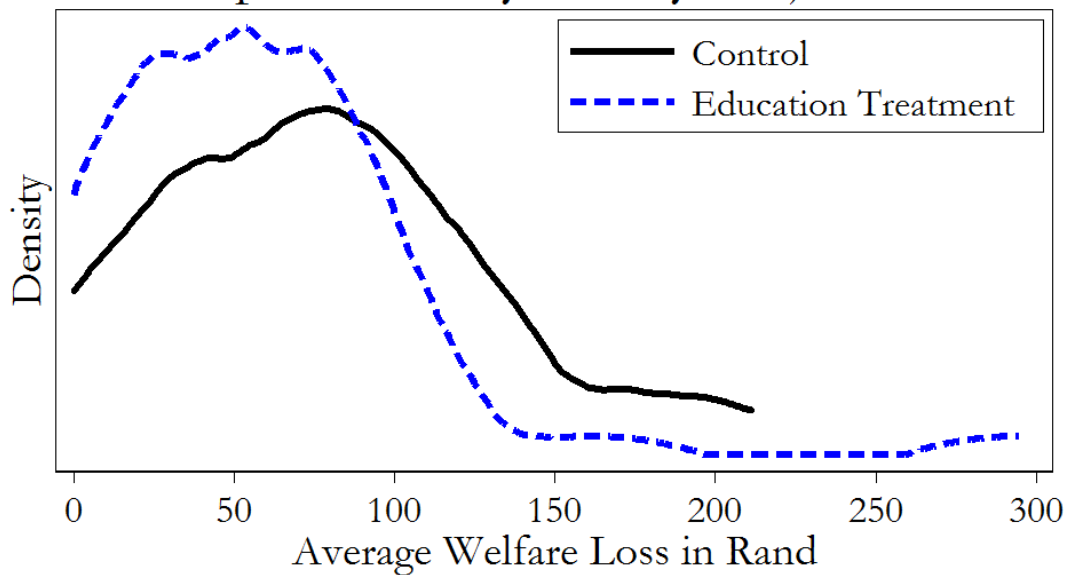


Figure 9: Average Welfare Loss in Control and Treatment Conditions for Rank Dependent Utility Subjects

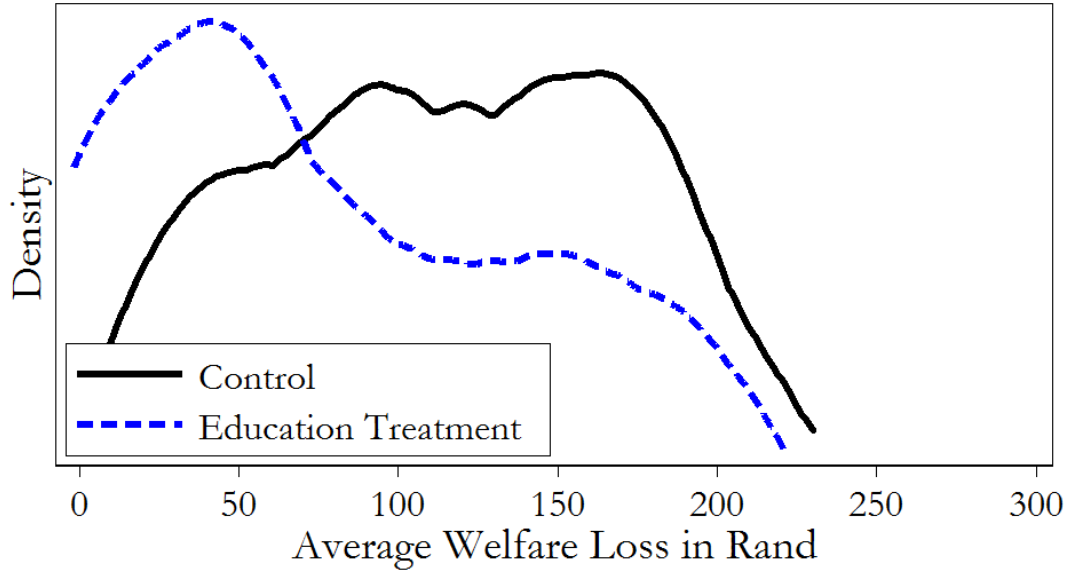


Figure 10: Total Effect on Welfare Loss

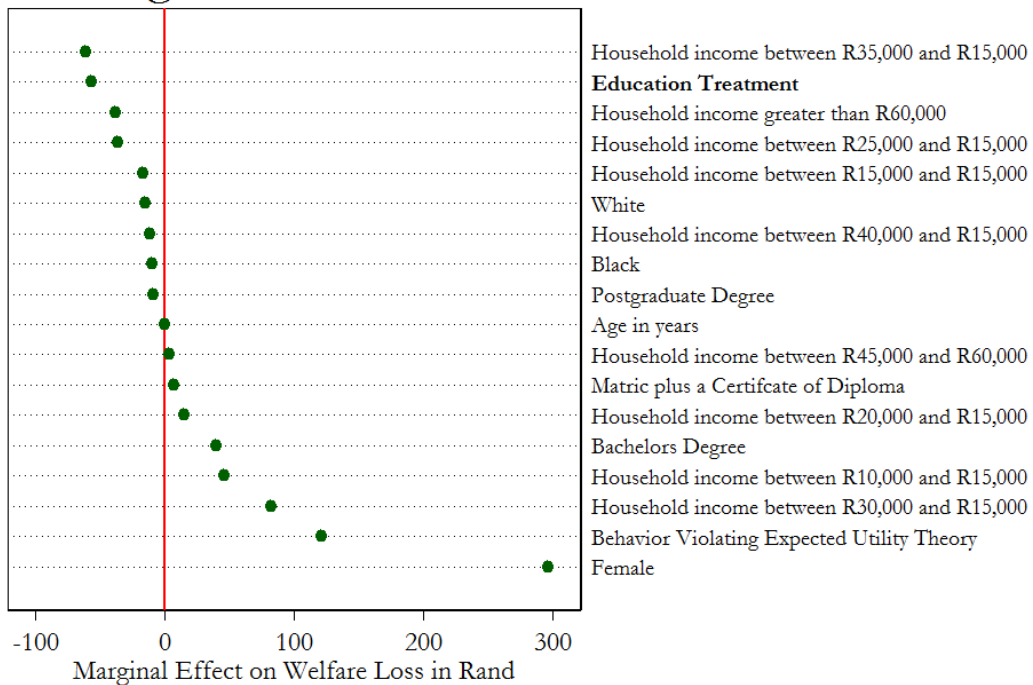
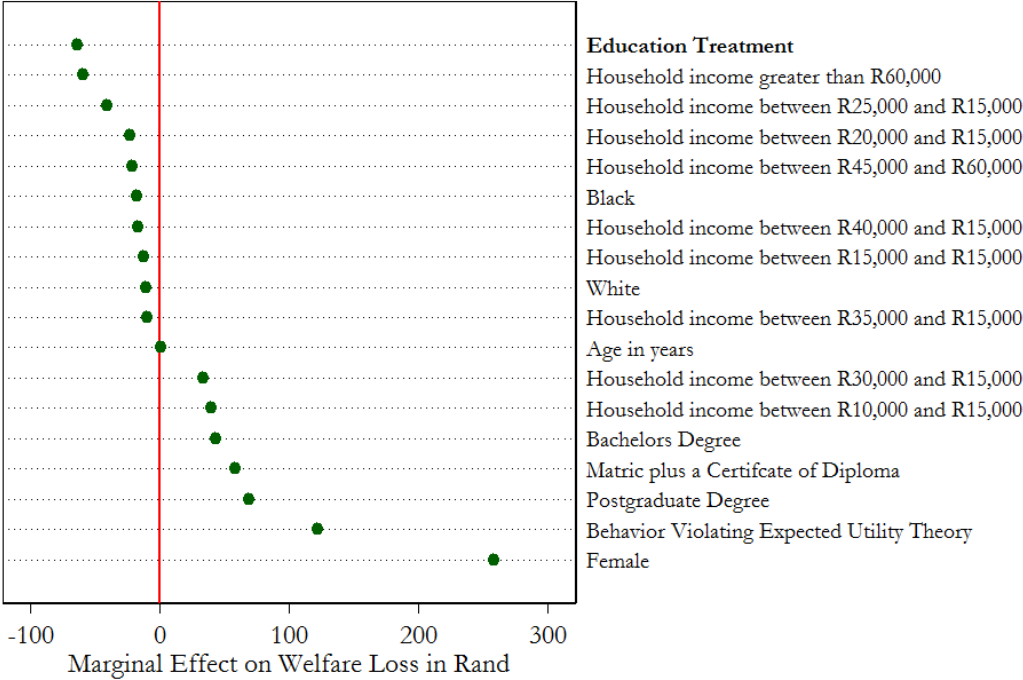




Figure 11: Marginal Effect on Welfare Loss



## Appendix A: The Intentional Stance and Sophisticated Revealed Preference Ascription

**[to be sent to referees, not to be published]**

Dennett (1987) has argued at length that ascribing preferences and beliefs involves taking the *intentional stance* toward an agent. One assumes that the agent's behavior is guided by goals and is sensitive to information about means to the goals, and about the relative probabilities of achieving the goals given the means. Such assumptions are not mere pretenses. Though goals, like preferences and beliefs, are not internal states of agents, but are rather relationships between agents, environments, and ascribers, there are nevertheless facts of the matter about what goals an agent has. It may be true that Carol goes to work because she believes that if she does she will get paid, and prefers having the paycheque to having the leisure she would gain if she bunked the job; but this truth status need not depend on there being discrete, recurring states of Carol's nervous system that realize the belief and, separately, the preferences. Beliefs and preferences are virtual states<sup>9</sup> of whole intentional systems rather than particular physical states of brains; but being virtual is a way of being real, not a way of being fictitious.

If a claim about intentional states is the sort of claim that can have a truth value, then it had better be possible to specify possible evidence that would undermine it. The holistic nature of intentional stance description allows for error, but also complicates it. Suppose we did not know, in setting out to explain Carol's behavior, that she has just won the lottery and so no longer needs the paycheque; but suppose further we also did not know that she would be ashamed to pass on a half-finished project to the colleague who will succeed her. On this hypothetical scenario, we predicted correctly that Carol would go to work because our two bits of ignorance cancelled one another out; but the error will reveal itself as we widen the sample of observations so that we include days beyond completion of Carol's current projects. It can also show up when we expand the range of behavior the intentional stance is called upon to rationalize – when we ask, for example, why Carol is no longer starting any new projects. Nevertheless, the holism of intentional attitude ascription *does* leave room for interpretive slack that we would not expect if we embraced naïve psychological realism associating beliefs and preferences with particular occurrent states in nervous systems. When we say that Carol prefers not to leave projects partly

---

<sup>9</sup> One way of understanding virtual states is as reaction potentials coupled with environmental affordances in the sense of Gibson (1977), except that the affordances in question will frequently be features of social events rather than (only) features detectable directly by sensory transducers. Because intentional states are propensities inferred from patterns of behavior, they approximately correspond to what some psychologists call 'latent' tendencies. However, psychologists often suppose that latent states have discrete neural realizations that might be discoverable by brain probes or functional neuroimaging. The use of 'virtual' expresses the widespread (though not unanimously endorsed) view among current philosophers that intentional states generally do not have such realizations because their semantic contents – what is believed or desired or preferred – vary partly with conditions external to the bodies of the agents whose states they are (Burge 1986; McClamrock 1995).

completed, do we refer to her conscientiousness, or to her fear of harm to her reputation? There might or might not be a fact of the matter here, and whether there is or isn't might not be relevant to the accuracy of the preference ascription.

Ross (2014) argues that this marks a main basis for the distinction between economics and psychology. Psychologists are professionally interested directly in how individuals process information, including information that influences decisions. Economists, by contrast, are concerned with this only derivatively. If a system of incentives will lead various people, through a heterogeneous set of psychological processes, to all make the same choice then the people form, at least for an analysis restricted to that choice, an equivalence class of economic agents. But it is a strictly empirical matter when this psychological heterogeneity will and won't matter economically. Economists, like all scientists, seek generalizations that support out-of-sample predictions. Different data-generating processes tend to produce, sooner or later, different data, including different economic data (that is, series of or patterns in incentivized choices). Economics is thus crucially informed by psychology in general, while not collapsing into the psychology of valuation as some behavioral economists have urged (Camerer, Loewenstein & Prelec 2005).

#### Appendix A References

- Burge, T. (1986). Individualism and psychology. *Philosophical Review* 95: 3-45.
- Camerer, C., Loewenstein, G., & Prelec, D. (2005). Neuroeconomics: How neuroscience can inform economics. *Journal of Economic Literature* 43: 9-64.
- Dennett, D. (1987). *The Intentional Stance*. Cambridge, MA: MIT Press.
- Gibson, J.J. (1977). The theory of affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*, pp. 67-82. Hillsdale, NJ: Lawrence Erlbaum.
- McClamrock, R. (1995). *Existential Cognition*. Chicago: University of Chicago Press.
- Ross, D. (2014). *Philosophy of Economics*. London: Palgrave Macmillan.

## **Appendix B: Procedure for Inferring Utility Function Structures from Experimental Choice Data**

We describe the analysis by which we determined whether a subject was better characterized as an EUT or an RDU agent. Assume that utility of income reflects constant relative risk aversion (CRRA), defined by

$$U(x) = x^{(1-r)}/(1-r) \quad (1)$$

where  $x$  is a lottery prize and  $r \neq 1$  is a parameter to be estimated. Then  $r$  is the coefficient of CRRA for an EUT individual:  $r=0$  corresponds to risk neutrality,  $r < 0$  to a risk loving attitude, and  $r > 0$  to risk aversion.

Let there be  $J$  possible outcomes in a lottery defined over objective probabilities. Under EUT the probabilities for each outcome  $x_j$ ,  $p(x_j)$ , are those induced by the experimenter, so expected utility (EU) is simply the probability weighted utility of each outcome in each lottery  $i$ :

$$EU_i = \sum_{j=1,J} [ p(x_j) \times U(x_j) ]. \quad (2)$$

The RDU model of Quiggin (1982) extends the EUT model by allowing for decision weights on lottery outcomes. The specification of the utility function is the same parametric specification (1) considered for EUT.<sup>10</sup> To calculate decision weights under RDU one replaces expected utility defined by (2) with RDU:

$$RDU_i = \sum_{j=1,J} [ \omega(p(x_j)) \times U(x_j) ] = \sum_{j=1,J} [ \omega_j \times U(x_j) ] \quad (3)$$

where

$$\omega_j = \omega(p_j + \dots + p_J) - \omega(p_{j+1} + \dots + p_J) \quad (4a)$$

for  $j=1, \dots, J-1$ , and

$$\omega_j = \omega(p_j) \quad (4b)$$

for  $j=J$ , with the subscript  $j$  ranking outcomes from worst to best, and  $\omega(p)$  is some probability weighting function.

We consider three popular probability weighting functions. The first is the 'power' probability weighting function considered by Quiggin (1982), with curvature parameter  $\gamma$ :

$$\gamma(p) = p^\gamma \quad (5)$$

So  $\gamma \neq 1$  is consistent with a deviation from the conventional EUT representation. Convexity of the probability weighting function, when  $\gamma > 1$ , is said to reflect 'pessimism' and generates, if one assumes, for simplicity, a 'linear' utility function, a risk premium since  $\omega(p) < p$  for all  $p$  and hence the RDU expected

---

<sup>10</sup> To ease complexity of notation we use the same parameter  $r$  because the context always make it clear if this refers to an EUT model or a RDU model.

value (EV) weighted by  $\omega(p)$  instead of  $p$  has to be less than the EV weighted by  $p$ .

The second probability weighting function is the ‘inverse-S’ function popularized by Tversky and Kahneman (1992):

$$\omega(p) = p^\gamma / (p^\gamma + (1-p)^\gamma)^{1/\gamma} \quad (6)$$

This function exhibits inverse-S probability weighting (optimism for small  $p$ , and pessimism for large  $p$ ) for  $\gamma < 1$ , and S-shaped probability weighting (pessimism for small  $p$ , and optimism for large  $p$ ) for  $\gamma > 1$ .

The third probability weighting function is a general functional form proposed by Prelec (1998) that exhibits considerable flexibility. This function is

$$\omega(p) = \exp\{-\eta(-\ln p)^\beta\}, \quad (7)$$

and is defined for  $0 < p \leq 1$ ,  $\eta > 0$  and  $\beta > 0$ .

In our experiment each subject made 50 binary choices between lotteries. After all decisions were made one of the 50 choices was chosen at random to be played out in accordance with the choices of the subject. Under EUT this experimental payment protocol provides incentives for truthful binary choices.<sup>11</sup>

The battery of lottery pairs was carefully selected for our purpose of identifying whether any given subject behaves more consistently under EUT or under RDU. Loomes and Sugden (1998) identify an important design feature for common ratio tests of EUT: variation in the gradient of the EUT-consistent indifference curves within a Marschak-Machina (MM) triangle. Respecting this advice generates choice patterns that are more powerful tests of EUT for any given risk attitude. Under EUT the slope of the indifference curve within a MM triangle is a measure of risk aversion. So there always exists some risk attitude such that the subject is indifferent, and evidence of common ratio violations has virtually zero power.<sup>12</sup> All of the lottery pairs implied by the battery have one or both lotteries on the ‘border’ of the MM triangle.

‘Border effects’ arise in tests of EUT when one nudges the lottery pairs in common ratio tests and common consequence tests into the interior of the MM triangle, or moves them significantly into the interior. The striking finding is that EUT often performs better when one does this. In fact the evidence is mixed in interesting ways. Camerer (1992) generated a series of experiments in which EUT did very well for interior lottery choices, but his data was unfortunately from hypothetical choices (i.e., subjects did not play out any lotteries for real money). These lotteries were well off the border. These lotteries can be contrasted with those used by Camerer (1989) that were on the border, and

---

<sup>11</sup> Harrison and Swarthout (2014) discuss the evidence for this experimental payment protocol, particularly when drawing inferences about RDU models. Their findings just make our classifications of subjects as EUT or RDU more conservative with respect to EUT (i.e., we are more likely with this payment protocol to classify subjects as RDU than if the protocol had no effect).

<sup>12</sup> EUT does not, then, predict 50:50 choices, as some claim.

where there were significant EUT violations. But Harless (1992) found that *just nudging the lotteries off the boundary did not improve behavior under EUT for real stakes*. So one natural question is whether the common ratio tests lead to EUT *not* being rejected when we are in the interior triangle, and to EUT being rejected when we are have choices on the boundary. This seems to be the conclusion from Camerer (1989, 1992), but it is not as clean as one would like.

The folk theorem on calibration of risk preferences for small stakes, originally stated by Hansson (1988) and popularized by others, is often raised to argue for the implausibility of using EUT to predict out-of-sample for larger stakes. But this argument depends on a premise that is false, at least for student subjects in the United States: such subjects do not universally exhibit small-stakes risk aversion for ‘all wealth’ (or for a large enough finite range of wealth levels). Cox and Sadiraj (2008; p. 33) proposed an elegant test of this premise. Give subjects choices between safe and risky lotteries, where the safe lotteries are certain amounts of money, and the risky lotteries are a 50:50 chance of  $+x/-y$  either side of the certain amount of money in the safe lottery. Hold  $x$  and  $y$  constant for choices that vary the safe prize level, and let  $x > y$  so that the expected value of this risky lottery is slightly above that of the safe lottery level. The idea here is to see the safe lottery as ‘lab wealth’  $w$ , and then see if subjects are risk averse as  $w$  varies. For instance, one might have  $+x/-y$  as  $+\$15/-\$10$ , then consider one binary choice in which the safe lottery is  $\$20$  and one binary choice in which the safe lottery is  $\$100$ . So the subject would make two choices: (a) take  $\$20$  for certain, or take a 50:50 chance of  $\$10$  or  $\$35$ , and (b) take  $\$100$  for certain, or take a 50:50 chance of  $\$90$  or  $\$115$ . Student subjects generally prefer the safe lottery in situation (a), and show indifference or even a slight preference for the risky lottery in situation (b). Whatever the fixed extra-lab wealth  $W$ , if we insist on perfect asset integration this evidence shows that the premise is false for  $W+w$ , as  $w$  is varied by the experimenter. We included 20 lottery pairs of this kind, to make our total number of lottery pairs 100. Our subjects then chose from amongst 50 lottery pairs, drawn randomly from this set of 100.

To evaluate RDU preferences we estimate an RDU model for each individual. We consider the CRRA utility function (1) and one of three possible probability weighting functions defined earlier by (5), (6) and (7). For our purposes of classifying subjects as EUT or RDU it does not matter which of these probability weighting functions characterize behavior: the only issue here is at what statistical confidence level we can reject the EUT hypothesis that  $\omega(p) = p$ .

#### Appendix B References

Camerer, C. (1989). An experimental test of several generalized utility theories. *Journal of Risk and Uncertainty* 2: 61-104.

Camerer, C. (1992). Recent tests of generalizations of expected utility theory. In W. Edwards (ed.), *Utility Theories: Measurements and Applications*. Boston: Kluwer.

- Cox, J. & Sadiraj, V. (2008). Risky decisions in the large and in the small: Theory and experiment. In J. Cox & G. Harrison (Eds.). *Risk aversion in experiments*. Bingley, UK: Emerald.
- Hansson, B. (1988). Risk aversion as a problem of conjoint measurement. In P. Gardenförs & N.-E. Sahlin, *Decision, Probability, and Utility*, pp. 136-158. Cambridge: Cambridge University Press.
- Harless, D. (1992), Predictions about indifference curves inside the unit triangle: A test of variants of expected utility. *Journal of Economic Behavior and Organization* 18: 391-414.
- Harrison, G., & Swarthout, J. T. (2014). Experimental payment protocols and the bipolar behaviorist. *Theory and Decision* 77: 423-438.
- Loomes, G., & Sugden, R. (1998). Testing different stochastic specifications of risky choice. *Economica* 65: 581-598.
- Prelec, D. (1998). The Probability Weighting Function. *Econometrica* 66: 95-113.
- Quiggin, J. (1982). A Theory of Anticipated Utility. *Journal of Economic Behavior and Organization* 3: 323-343.
- Tversky, A., & Kahneman, D. (1992). Advances in Prospect Theory: Cumulative Representations of Uncertainty. *Journal of Risk and Uncertainty* 5: 297-323.