CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Varieties of paternalism and the heterogeneity of utility structures
Authors	Harrison, Glenn W.;Ross, Don
Publication date	2017-10-05
Original Citation	Harrison, G. W. and Ross, D. (2017) 'Varieties of paternalism and the heterogeneity of utility structures', Journal of Economic Methodology, 25(1), pp. 42-67. doi:10.1080/1350178X.2017.1380896
Type of publication	Article (peer-reviewed)
Link to publisher's version	10.1080/1350178X.2017.1380896
Rights	© 2017, Informa UK Limited, trading as Taylor & Francis Group. This is an Accepted Manuscript of an article published by Taylor & Francis in Journal of Economic Methodology on 2017.10.05, available online: http://www.tandfonline.com/10.1080/1350178X.2017.1380896
Download date	2026-10-31 04:03:23
Item downloaded from	https://hdl.handle.net/10468/5427



UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Varieties of Paternalism and the Heterogeneity of Utility Structures

Glenn W. Harrison[1]

Don Ross[2]

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

2 – School of Sociology and Philosophy, University College Cork, Ireland,; 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

March 2017

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. Nudging Versus Boosting

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). Grüne-Yanoff & Hertwig

(2016) [GYH] have distinguished nudging, which is based on the heuristics-and-biases (H&B) branch of behavioral economics research associated with Kahneman & Tversky (1982) and Kahneman (2011), from policies aimed at 'boosting,' which apply the 'simple heuristics' (SH) research program of Gigerenzer *et al* (1999), Todd *et al* (2012) and 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: see Ashcroft (2011) and Conly (2012; 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[1], which they can choose to apply or not once they have been enlightened. Thus boosting, according to GYH, avoids manipulating the agents to whom the policies in question are applied, and is to that extent less paternalistic.[2]

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 'rationality spillovers' discussed by Cherry *et al* (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 the policy maker can make participation the outcome if the DM is passive, needing to take action only if the DM wants to act on a preference not to participate. The leading example of a boost discussed by GYH 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 evidence suggests that some 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

---

[1] GYH 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 a heuristic.

[2] This motivation for boosting is similar to reasons given by John *et al* (2009) and John *et al* (2011) in favour of what they call a 'think' strategy for correcting people's reasoning errors. Such 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 the concerns of Le Grand & New (2015, p. 142) concerning the general practicality and likely effectiveness of think strategies.

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 suspicion about 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, such as 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 become a minority stance within the discipline. Leamer (2012) stresses that 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. 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,[3] while any proposal for intervention that imposes normative

[3] Le Grand & New (2015) philosophically analyze government, as opposed to private, paternalism, and 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 (2015) defend the normative

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 some behavioral welfare theorists, such as Bernheim & Rangel (2008) and Bernheim (2016) 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. 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.[4] 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 concerned with the distinction between nudging and boosting as it applies to what we believe to be a 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. The behavioral welfare theorist's suggested meta-policy fails to give the economist helpful advice in the most common sorts of policy situations of practical interest.

We emphasize our *methodological* focus on practical issues that arise for applied economists, as opposed to philosophical issues that dominate abstract debates. Philosophical discussions, as in Hausman (2011), often proceed, for understandable

---

thesis that justification of paternalism requires identification of a correctible judgment. We 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.

[4] The idea here is not that preferences over options arising later in time should generally be regarded as dominating preferences over options arising earlier in time. The proposal of Bernheim and Rangel (2008) is that the welfare analyst should search for a choice environment in which the target agent's preferences are consistent. Earlier or later time slices of the agent's biography, drawn from environments in which consistency is violated, are treated as preferences of other agents. The welfare analyst then recommends any Pareto-consistent policies, applied to the community of sub-agents, that she can find. This of course allows for, indeed predicts, situations in which no recommendation between some alternative policies is favored. Bernheim (2016, p. 14) defends this unambitious program, with its *ad hoc* reliance on case-specific psychological hypotheses rather than general economic theory, on the grounds that structural models generally make overly strong assumptions that have "little basis." We submit that the case study we consider stands as a quite typical counter-example to this defeatist stance, as does Harrison and Ng (2016), which we also discuss below.

reasons, by considering the implications of conceptual distinctions for idealized, general, or hypothetical cases, set up so as to push pragmatic 'side issues' into the background. We are not directly engaging the debate at that level of abstraction. Thus we should not be interpreted as trying to argue that nudging and boosting are conceptually 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 an everyday domain of economic engineering. In this domain, normative and technical considerations are typically tightly entangled, as we illustrate. We argue that a meta-policy, according to which boosting is morally unproblematic, while nudging proposals must always be accompanied by responses to concerns about paternalism, is awkwardly adapted to the front line of applied economics. If we see economics as largely a policy science, a form of institutional engineering, then economists cannot simply refuse to engage with normative complexities. But Leamer (2012) also reminds us that philosophical distinctions developed *in vivo* need to be examined *in situ* if they are to be made fully relevant to economists.

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. Crucially, our exercise was not designed to be a test-bed for conceptual or normative issues. Equally importantly, the advice we based on it, if implemented by the client, will have real consequences for individuals and households.

In Section 2 we describe the commissioned experimental research that we conducted, and the advice we were asked to provide on the basis of it. Section 3 motivates the analyses we performed on the data, and the results we obtained. Section 4 pulls the preceding strands together and gives the argument for the main methodological conclusions.

Although we have stressed that we are not engaged in first-order philosophical investigation into the idealized concepts of nudging and boosting, we believe that debates drawn from the philosophy of mind and agency can shed diagnostic light on the difficulties encountered in translating welfare theory into policy-focused practice. This diagnosis is outlined in our concluding Section 5.

.

2. Helping Investment Product Retailers Give Better Customer Advice

In 2014 we accepted a commission for research from a major South African retailer[5] of household investment products, which are primarily mutual funds in

[5] 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.

American terminology. The company's motivation in commissioning the research began from its observation, 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 products, or compounding losses by churning their 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.[6] The need for us to guard our client's intellectual property limits the extent of detail with which we can describe this informational intervention. However, we can say enough to locate the intervention in terms of the distinction between nudging and boosting. The client's customers, when meeting with a broker to choose portfolios, were typically told only about options' probable long-run rates of return on initial investment, maximum expected 'drawdown' (lowest value likely to be visited by the asset's value walk), and historical standard deviation. This allowed for a crude, qualitative operationalization of 'risk aversion': if a customer indicated discomfort with the maximum expected drawdown, they would be advised to opt for a portfolio with lower variance at the expense of a more modest expected long-run return. The client's 'education intervention,' which we were asked to experimentally test, provided clients with online charts showing fill histories of portfolios under consideration. These showed historical variance in the strict sense, along with skew and kurtosis in distributions of returns. Furthermore, the information site was interactive so that the customer could retrieve definitions and brief explanations of the risk-related portfolio properties displayed. The intervention included no simple heuristics or motivating messages, of the kind which Ambuehl, Bernheim and Lusardi (2014) found under some circumstances can lead retail investors to choose less optimally (with respect to their subjective utility) than if they are provided with objective information only.

[6] 'Adaptedness' refers in evolutionary psychology to pre-adapted dispositions a subject brings to a task.

Our research consisted in designing, administering, and analyzing a controlled trial of a prototype of the intervention. The client believed that most customers they perceived as 'rational,' in the sense that they did not prematurely sell their portfolios or over-churn, would be annoyed and discouraged by the time involved in experiencing the intervention, and might find the explanatory notes condescending. The client therefore wanted to identify demographic characteristics of potential customers that could predict which subsets of the customer base were likely to benefit from the intervention. We brought to the client's attention that scientifically estimated risk preference structures, which we could elicit in an experiment, might prove to be at least as informative as demographic properties. The client agreed that our experiment should explore this aspect.

Our 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, and then assigned them randomly to one of two investment treatments.

Risk attitudes were measured by evaluating a series of choices by each subject between pairs of lotteries that had an average yield of 300 South African Rand (R300, which exchanged for about US$27 at the time of the experiment). In this **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 charts, illustrated in Figure 1. The subjects were asked to choose one lottery from each pair by clicking on the corresponding button below their preferred lottery. One of the 50 choices was selected at random for realization and payment.

[Figure 1 about here]

The data generated by performance of this task allowed us to estimate the structures of risk preferences for each subject. Lottery tasks similar to the ones employed here have been used to estimate risk preferences for individuals, typically using maximum likelihood estimation in the spirit of Hey & Orme (1994) and Harrison & Ng (2016) **.**

In the **Investment Task**, each subject 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: investment objective, return history, standard deviation, and maximum drawdown. Subjects in the treatment group were additionally provided with the 'education intervention.'

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.

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.[7]

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 an approximation to a financial product to which subjects could potentially have access through a brokerage. Each simulated fund was a discretized lottery of the continuous distribution of historical returns associated with the real-life counterpart of the simulated fund in question. 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 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. Die-rolling by subjects was used to select one of the simulated markets.

The task started with a screen explaining that a certain amount of money was to be invested in one or more funds. The different types of funds were explained, but without details on their potential returns. Those in the treatment group were then presented with the interactive 'education intervention' that allowed exploration of the histories of the funds, in formats hypothesized to be cognitively accessible. Subjects in both the treatment and control groups were then allowed to allocate their endowments to funds, and everyone saw some base level of information about the potential fund returns: 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 invest in as many funds as they wanted.

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

3. Analytical Methods and Results

Idealized discussions of welfare and of economic policy have, at least until the recent emergence of the behavioral literature, taken the Expected Utility Theory (EUT) of Savage (1954) to provide the basic technical apparatus for normatively comparing alternative states of the world for an agent. Binmore (2009) provides an authoritative updating, with suitable cautions against hubristic over-extension, of this theoretical landmark in the context of contemporary operationalization.

[7] 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. Analysis of the results of this task will not figure in the discussion here, so we pass over design details.

Behavioral economists often interpret their work as motivating revisions to, or, for those who favor rhetorics of disruption, paradigm replacements for, EUT. Among various formal models of choice under risk or uncertainty that are contrasted with EUT, the Cumulative Prospect Theory (CPT) of Tversky and Kahneman (1992) has received the most attention. A common strategy in both theoretical and applied behavioral economics has been to run 'horse races' between EUT and CPT or another alternative as rival models for estimating a specific data set, and urging that the winner of the race, that is, the model that yields the best fitting estimation, should then be used as the basis for empirical interpretation. Such horse races stack the deck against EUT when, as is almost always the case, the other horse has greater structural complexity and observed behavior is economically heterogeneous (Ross 2005, pp. 174-176). When an investigator following this approach concludes that EUT is the 'losing' contender with respect to empirical estimation, the question remains open about how to proceed to normative analysis. An economist who follows Savage (1954) in thinking that EUT is the normatively correct model of 'rational' decision, regardless of the extent to which real human choice conforms to it, might analyze agents' welfare against the outcomes they would have obtained had EUT correctly characterized their behavior. Alternatively, one might employ the latent utility function embedded in a more elaborate model of risk preferences, as proposed by Bleichrodt *et al* (2001).

Following recent theoretical advances summarized in Harrison and Rutström (2008), the technical apparatus used to analyze the experiment we discuss goes beyond this 'horse race' methodology. It reflects advances in understanding of the relationship between CPT and other alternatives to EUT as descriptive models of choice estimated at the level of individuals.

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 two core models of decision-making under objective risk. One is EUT,[8] and posits that the risk premium is explained solely by an aversion to variability of earnings from a prospect. The second is the Rank-Dependent Utility (RDU) model of Quiggin (1982), which 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)

[8] We consider decision making under objective risk because all of our methodological points can be made in that setting. An important extension would be to consider risk preferences under subjective risk, using either Subjective Expected Utility or some models that allow for uncertainty aversion when individuals do not apply the Reduction of Compound Lotteries axiom to subjective probability distributions. In that latter case some aspect of the distribution, other than the average, matters for decisions: see Harrison (2011; §4). Models that allow for ambiguity aversion when individuals do not even have well-formed subjective probability distributions could also be considered, but this would raise many additional issues of positive and normative methodology well beyond our immediate remit.

intentional stance;[9] that is, it might not refer 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.

We do not estimate our data using CPT. Our avoidance of CPT is based on analysis of its relationship to RDU, both theoretically and in application to empirical data. Harrison & Swarthout (2016) provide an extensive literature review, which finds that most reported evidence for 'loss aversion' is actually evidence for probability weighting. They also report evidence of (at least local) asset integration in the laboratory, which is fatal for empirical adequacy of CPT. Harrison and Ross (2017) review further evidence, and consider the implications for welfare assessment of the conjecture that the many reported 'horse race' victories of CPT over EUT were really wins for RDU in disguise, where CPT's successes stemmed from its allowance for probability weighting rather than 'utility' loss aversion relative to an idiosyncratic reference point. We thus focus on EUT and RDU.

We begin with EUT. 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 induced by the experimenter, so expected utility is the probability weighted utility of each outcome in each lottery. Once the utility function is estimated, risk aversion is measured. 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 loose, this can be viewed as characterizing individuals that are averse to mean-preserving increases in the variance of returns. We assume that utility of income reflects constant relative risk aversion (CRRA), defined by $U(x) = x^{(1-r)}/(1-r)$ 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.

The RDU model extends EUT 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.

Subjects' risk preferences were analysed based on the Lottery Task. Again, we conducted analysis based on the assumption that each subject's behavior was either best characterized by EUT or by RDU. When a subject was estimated to be an RDU agent, we tested further to determine which of several probability weighting functions best characterized the pessimism or optimism about probabilities.

We consider three popular probability weighting functions. The first is the 'power' probability weighting function with curvature parameter γ: $\gamma(p) = p^{\gamma}$. So $\gamma \neq 1$ is

[9] The intentional stance is discussed in Section 5.

consistent with a deviation from the conventional EUT representation.[10] The second probability weighting function is the 'inverse-S' function: $\omega(p) = p^{\gamma} / (p^{\gamma} + (1-p)^{\gamma})^{1/\gamma}$. 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 $\omega(p) = \exp\{-\eta(-\ln p)^{\varphi}\}$ and is defined for $0<p\leq1$, $\eta >0$ and $\varphi >0$. The RDU agent is also assumed to have a CRRA utility function with parameter r.

We can use the results from a specific subject to illustrate the type of risk preferences estimated. Consider subject #22. We first 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 that $\omega(p) = p$ 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, following Harrison and Ng (2016).

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, 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)

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

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).[11]

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, …, 2, 2.5, 3, 3.5); RDU Power: r = (-10, -5, -3, -2, -1, 0, 0.1, 0.2, …, 1, 1.25, 1.5) and γ = (0.2, 0.7, 1.2, …, 3.2, 4, 5); RDU Inverse-S: r = (-10, -5, -3, -2, -1, 0, 0.2, 0.4, …, 1.6) and γ = (0.3, 0.4, 0.5, …, 1.1); RDU Prelec: r = (-10, -5, -3, -2, -1, 0, 0.25, 0.5, …, 2), η = (0.3, 0.8, 1.3, …, 2.8), and ϕ = (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.[12] 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

---

[11] Bleichrodt *et al* (2001) maintain that EUT is the appropriate normative model, and correctly note that if an individual is an RDU or CPT decision-maker, then recovering the utility function from observed lottery choices requires allowing for probability weighting and/or sign-dependence. They then implicitly propose using that utility function to infer the certainty equivalent using EUT. These are radically different normative positions. Some notation will help make this clear. Let RDU(x) denote the evaluation of an insurance policy x in Harrison and Ng (2016) using the RDU risk preferences of the individual, including the probability weighting function. They calculate the certainty-equivalent CE by solving $U^{RDU}(CE) = RDU(x)$ for CE, where U^{RDU} is the utility function from the RDU model of risk preferences for that individual. But Bleichrodt *et al* (2001) evaluate the CE by solving $U^{RDU}(CE) = EUT(x)$ where EUT(x) uses that utility function in an EUT manner, assuming no probability weighting. This strikes us as normatively illogical. The logical approach here would be to estimate the "best fitting EUT risk preferences" for the individual from their observed lottery choices, and then use the utility function U^{EUT} as the basis for evaluating the CE using $U^{EUT}(CE) = EUT(x)$.

[12] 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.

plus 5%. The interest bearing funds were derived from historical data using the interest 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.[13] 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.

[13] 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.

[Figure 5 about here]

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 would optimize their returns, reducing the probability that their savings goals would be frustrated. The intervention is thus intended to directly improve 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 the 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.

Table 1 Eight assumptions of the nudge and boost approaches

	Nudge	Boost	Investor Education Intervention
Cognitive error awareness Must the decision maker be able to detect the influence of error?	No	Yes	No
Cognitive error controllability Must the decision maker be able to stop or override the influence of the error?	No	Yes	Yes
Information about goals Must the designer know the specific goals of the target audience?	Yes	No	Yes
Information about the goals' distribution Must the designer know the distribution of goals in the target audience?	Yes	No	Yes
Policy designer and cognitive error Must experts be less error-prone than decision makers?	Yes	No	Yes
Policy designer and benevolence Must the designer be benevolent?	Yes	No	No
Decision maker and minimal competence Must the decision maker be able to acquire trained skills?	No	Yes	Yes
Decision maker and sufficient motivation 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. Concerning our assessments in the third and fourth rows, 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.[14] 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 (Ambuehl, Bernheim and Lusardi 2014). 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, in part, an approach exemplified and promoted in a similar problem context by Harrison and Ng (2016), when they evaluate the welfare gain 'introduced

[14] A referee objected that our client would not need to track specific goals of any customers once the intervention had been administered, but would simply leave it to the 'educated' customers to reflect their new information in their choices or not. Thus the referee suggested that our assessments should be "no" in column 4 of rows 3 and 4. This suggestion depends on equivocation over what the intervention is: the client viewed administration of the education intervention as burdensome to customers. If the client company follows our advice, then, it will be selecting certain customers to be burdened on the basis of identifications made by it, not on the basis of self-identifications by customers of their own needs in light of enhanced knowledge.

into the world' by a standard type of indemnity insurance product. They aim to reliably estimate the distribution of risk preferences among individuals, and the distribution of their subjective beliefs about loss contingencies and likelihood of payout, so as to identify 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.

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, upon which we elaborate in Section 5 below. According to that thesis, such attitudes, which include 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 the insistence of Hausman (2011) 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 information 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.

Let us apply this methodological point to the normative analysis given by Harrison and 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 by crude revealed preference dogma to say that the choice of the policy necessarily revealed a preference for having the policy over not having the policy, 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.

Lottery choices made under controlled experimental conditions, as in our case study, arguably provide 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. This is often, though not always, a reasonable assumption in policy contexts. [15]

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 by 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 (2016) 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.

It might seem that all of this amounts only to a modest, practical point that should be of limited interest to theorists. That is, we might seem to be saying only that, although the concepts of nudging and boosting are as clear as can reasonably be expected at the abstract level, consulting clients often frame the questions they assign to economists in terms that force the distinction to be elided in practice. In that case, it might be thought that the sole upshot is that economists could usefully bring the nudging/boosting distinction to clients' attention while research briefs are being

[15] Sugden (2004)(2009) denies, at least, that the assumption is viable generally *enough* to provide a sound methodology for normative economics. We take up his objection in Section 5.

negotiated, so that clients will at least appreciate that presuppositions they bring to the framing of their policy options may embed normative blind spots.

In fact, however, we think that lessons of deeper methodological, and indeed philosophical, significance can be taken from the main case study we have presented, and from its relationship to the Harrison and Ng (2016) case. We draw out these implications in the concluding section.

5. Welfare Analysis From the Intentional Stance

In our case study, although we recommended additional cognitive preparation for RDU choosers before they selected investment products, we did not recommend trying to teach them the concept of probability weighting so they could then apply this characterization to themselves. This is only partly motivated by the questionable practicality of the pedagogical task that would be required. It also reflects wariness about telling subjects a story about themselves they would surely interpret as telling them that they possess a kind of internal psychological 'defect' when such a story would outrun our available data and is in any case doubtful according to sophisticated philosophy of mind.

It is unlikely that most people choosing investment funds attempt to compute internally represented optima – either from EUT or RDU bases – and then make computational errors that could be pointed out to them. This echoes a point made by Infante, Lecouteux and Sugden (2016) (ILS) when they complain that behavioral welfare economists typically follow Hausman (2011) in 'purifying' empirically observed preferences. ILS argue that purification reflects an implicit philosophy according to which an 'inner' Savage-rational agent is 'trapped within' a psychological, irrational 'shell' from which best policy should try to rescue her. ILS provide no general philosophical framework within which they motivate their skepticism about 'inner rational agents'. However, such a framework is available.

Dennett (1987) provides a rich account of the ontology of beliefs, preferences and other 'propositional attitude' that relate behavioral and cognitive dispositions to different states of the world and to different representations of those states. Dennett (1987) argues at length that ascribing preferences and beliefs involves taking the *intentional stance* toward an agent. This consists in assuming 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 available means. Goals, like preferences and beliefs, are not internal states of agents, but are rather relationships between agents, environments, and ascribers, The baseline case for understanding such ascription is effort by a third party to interpret and predict the agent's actions by means of controlled speculation about an agent's overall behavioral ecology and information-processing capacities. Crucially, people are socially obliged, and trained during socialization while growing up, to adopt the intentional stance toward themselves. For the sake of coordination in both action and communication, agents' self-ascriptions are made under constraint of at least approximate alignment with ascriptions of others.

These ascriptions and self-ascriptions are not guesses about 'true' beliefs and preferences hidden from direct view in people's heads. Rather, constructed rationalizations of agents' behavioral and cognitive ecologies is what beliefs and preferences *are.* Critics have sometimes misinterpreted this view as *instrumentalism,* a

doctrine according to which beliefs and preferences are mere useful fictions. Dennett has consistently maintained, however, that there are facts of the matter about agents' goals and access to information, and hence also facts about their propositional attitudes. 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[16] 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 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

[16] 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 view among many 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).

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 *et al* 2005).

Applying this philosophy of mind and agency to our main case study,, we assume the intentional stance to make sense of our experimental subjects' 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. *Externalism about preference content blurs the distinction between 'treating' the subject and 'treating' the subject's environment.* Furthermore, the more precisely we specify the contents of propositional attitudes, especially in quantitative terms, the less weight in identification will rest on 'inboard' elements of data generating processes relative to external aspects of the agents' overall behavioral ecologies.[17] Our technical tools allow us to identify virtual intentions that most subjects are not able to identify when they take the intentional stance to themselves, 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. These patterns therefore enter into a fully informed analyst's specification of the subjects' beliefs and preferences. In this philosophical framework, it makes sense to say that we *boost* the subjects' informational access in a way that *nudges* their (sub-deliberative) cognition.

It helps to contextualize our approach to normative analysis to contrast it with the more radical revisionism advocated by Sugden (2004)(2009). He develops an insightful framework for normatively evaluating agents' outcomes under alternative institutional arrangements in a way that privileges their autonomy as choosers (i.e., their consumer sovereignty) without depending on their specific preference orderings, and thus without requiring their preferences to even be consistently ordered, let alone fully EUT-compliant. According to Sugden (2004)(2009), agents are made better off to the extent that their opportunity sets are expanded, and worse off to the extent that their opportunity sets are contracted. Against this standard, 'pure' boosts will typically make agents better off and 'pure' nudges will typically make them worse off. We find this idea, which Sugden (2004)(2009) elegantly formalizes, attractive as a way of addressing normative questions in circumstances where welfare analysis in the technical sense is not possible due to preference reversals. Thus, for example, this approach can generate recommendations in cases where the method of Bernheim and Rangel (2008) and Bernheim (2016) would find Pareto indifference and therefore yield no guidance. But we should not abjure *ever* doing standard welfare analysis merely because it can't be undertaken in *every* context. In both the Harrison and Ng (2016) case and in the situation presented to us by our consulting client, the complications arise

[17] Clark (1998) refers to these external elements as 'cognitive scaffolding.' Ross (2005)(2014) develops the role of scaffolding in specifying and identifying utility functions using sophisticated revealed preference theory.

from the existence of preferences that violate EUT but are nevertheless well-ordered. We suggest that this is the standard situation where relevant utilities are expected monetary values.[18]

To summarize, the claimed normative advantage of boosting over nudging relies 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 generally 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 complicates, though it does not vitiate, attempts to apply the nudging/boosting distinction to practical economic welfare assessments.

References

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

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

Ainslie, G. (1992). *Picoeconomics*. Cambridge: Cambridge University Press.

Andersen, S., Harrison, G.W., Lau, M. I. and Rutström, E. E. (2014). Discounting Behavior: A Reconsideration. *European Economic Review*, 71, 15-33.

Ambuehl, S; Bernheim, B. D., and Lusardi, A. (2014). The effect of financial education on the quality of decision making. NBER Working Paper 20618. http://www.nber.org/papers/w20618

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. (2016). The good, the bad, and the ugly: A unified approach to behavioral welfare analysis. *Journal of Benefit-Cost Analysis* 7: 12-68.

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.

[18] Someone who thinks that even expected monetary payoffs are typically hyperbolically discounted by people would quarrel with this suggestion. Sugden (2004)(2009) implies this concern. But that hypothesis is rejected by the leading psychological theorist of hyperbolic discounting, Ainslie (1992), and is contrary to empirical findings reported by Andersen et al (2014).

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

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

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.

Burge, T. (1986). Individualism and psychology. *Philosophical Review* 95: 3-45.

Camerer, C., Issacaroff, S., Loewenstein, G., O'Donaghue, 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.

Clark, A. (1998). *Being There.* Cambridge, MA: MIT Press.

Conly, S. (2012). *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.

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.

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. (2011). Experimetal methods and the welfare evaluation of policy lotteries. *European Review of Agricultural Economics* 38: 335-360.

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., & Ross, D. (2017). The empirical adequacy of cumulatve prospect theory and its implications for normative assessment. CEAR Working Paper 2017-01, Center for Economic Analysis of Risk, Robinson College of Business, Georgia State University.

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.

Infante, G.; Lecouteux, G,, and Sugden, R. (2016). Preference purification and the inner rational agent: A critique of the conventional wisdom of behavioral welfare economics. *Journal of Economic Methodology* 23: 1-25.

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 behavior. *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: Heursitics 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.

McClamrock, R. (1995). *Existential Cognition.* Chicago: University of Chicago Press.

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. (2005). *Economic Theory and Cognitive Science: Microexplanation.* Cambridge, MA: MIT Press.

Ross, D. (2014). *Philosophy of Economics*. London: Palgrave Macmillan.

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 behavior. *Economica* 5: 61-72.

Savage, L. (1954). *The Foundations of Statistics*. New York: Wiley.

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.

Sugden, R. (2009). Market simulation and the provision of public goods: A non-paternalistic response to anomalies in environmental evaluation. *Journal of Environmental Economics and Management* 57: 87-103.

Sunstein, C., & Thaler, R. (2003a). Libertarian paternalism. *American Economic Review, Papers and Proceedings* 93: 175-179.

Sunstein, C., & Thaler, R. (2003b). Libertarian paternalism is not an oxymoron. *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.

Figure 1: A Sample Lottery Choice Pair

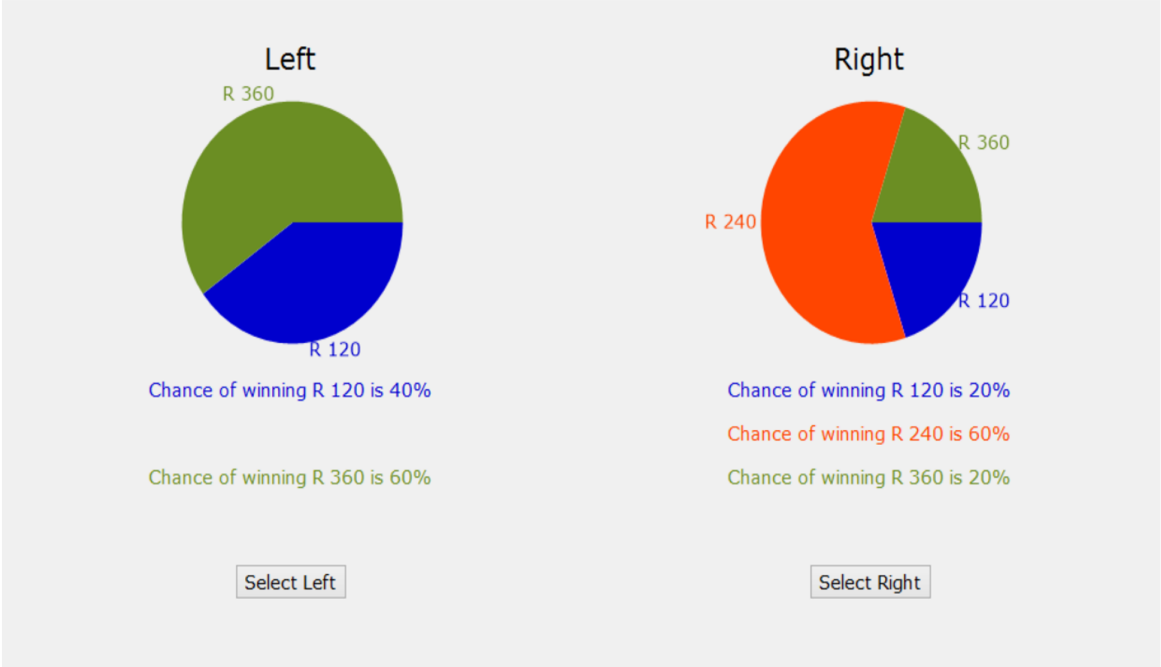


Figure 2: Classifying Subjects as EUT or RDU

N=194, one *p*-value per individual
Estimates for each individual of EUT and RDU specifications

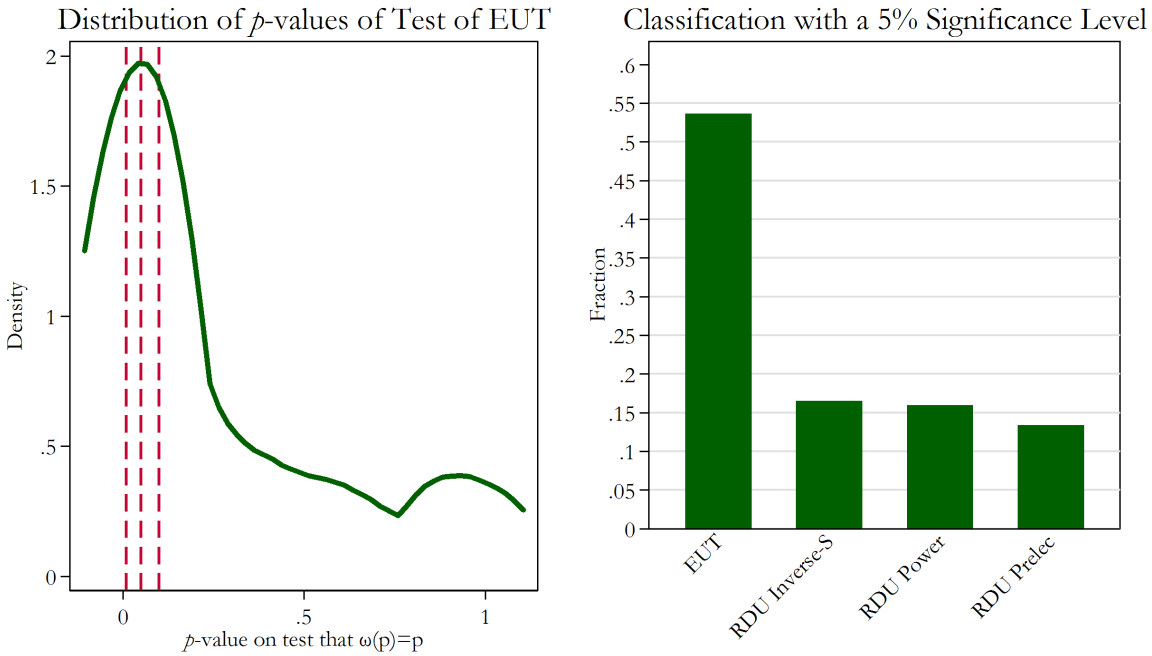


Figure 3: Annualized Risk-Return Tradeoffs in the Investment Task

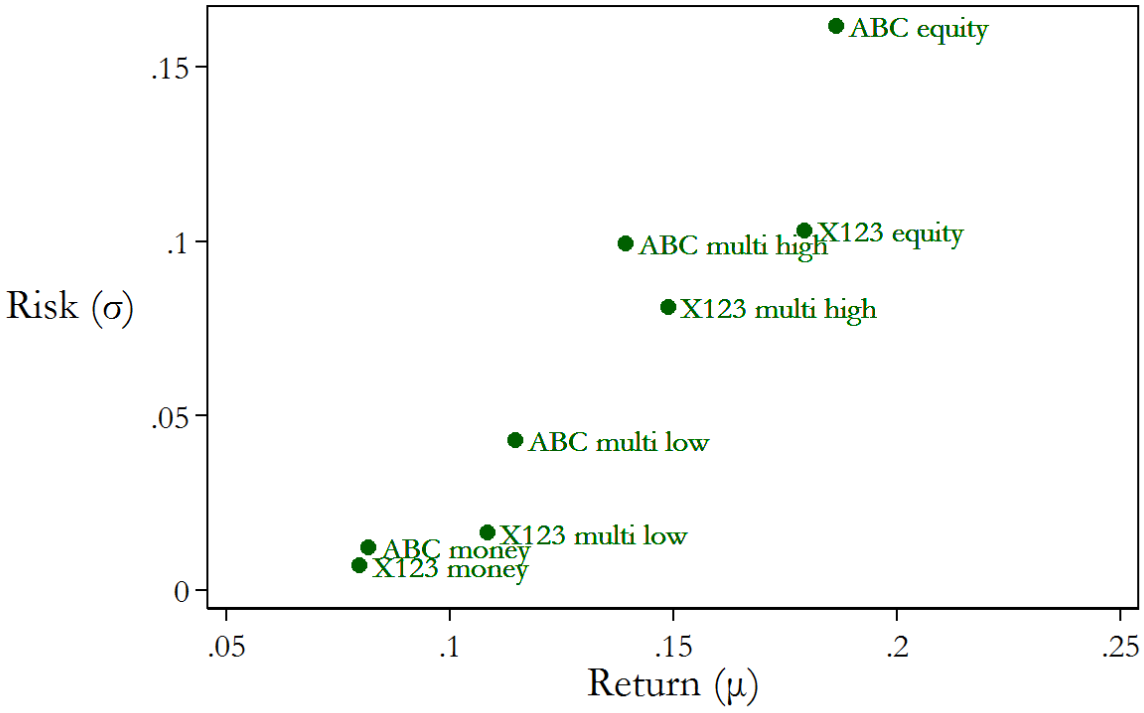


Figure 4: Number of Funds Utilized

N = 193 subjects from the Investment Task

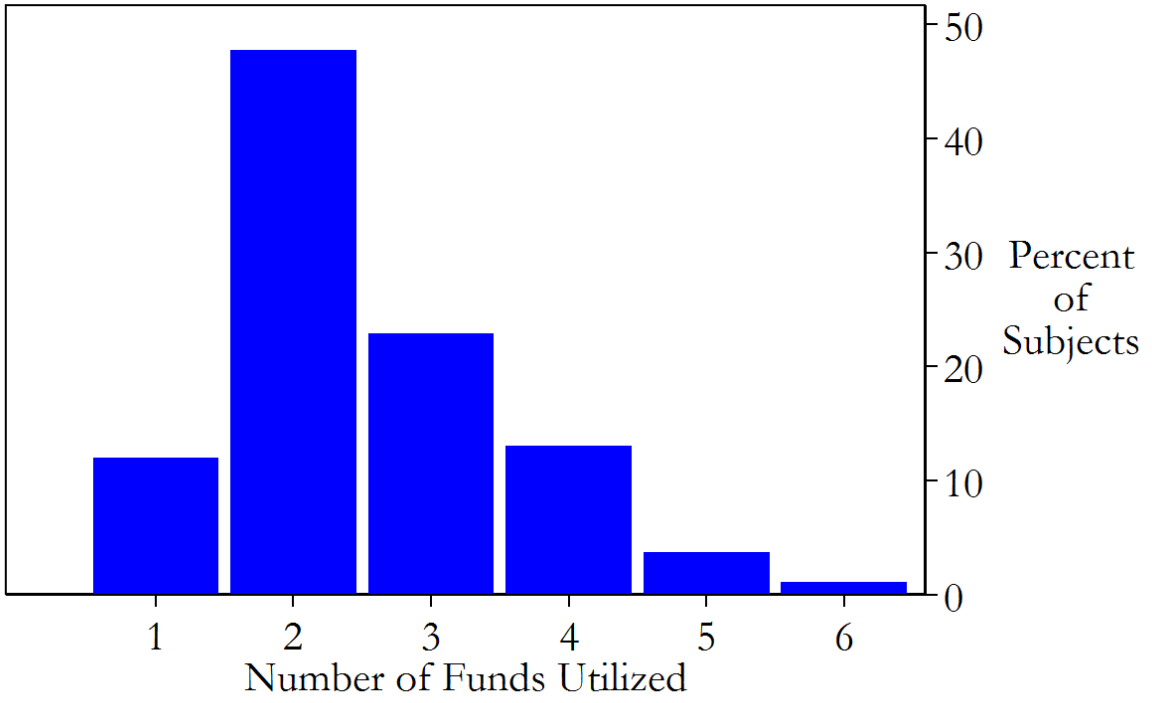


Figure 5: Average and Optimal Allocations

N = 193 subjects from the Investment Task

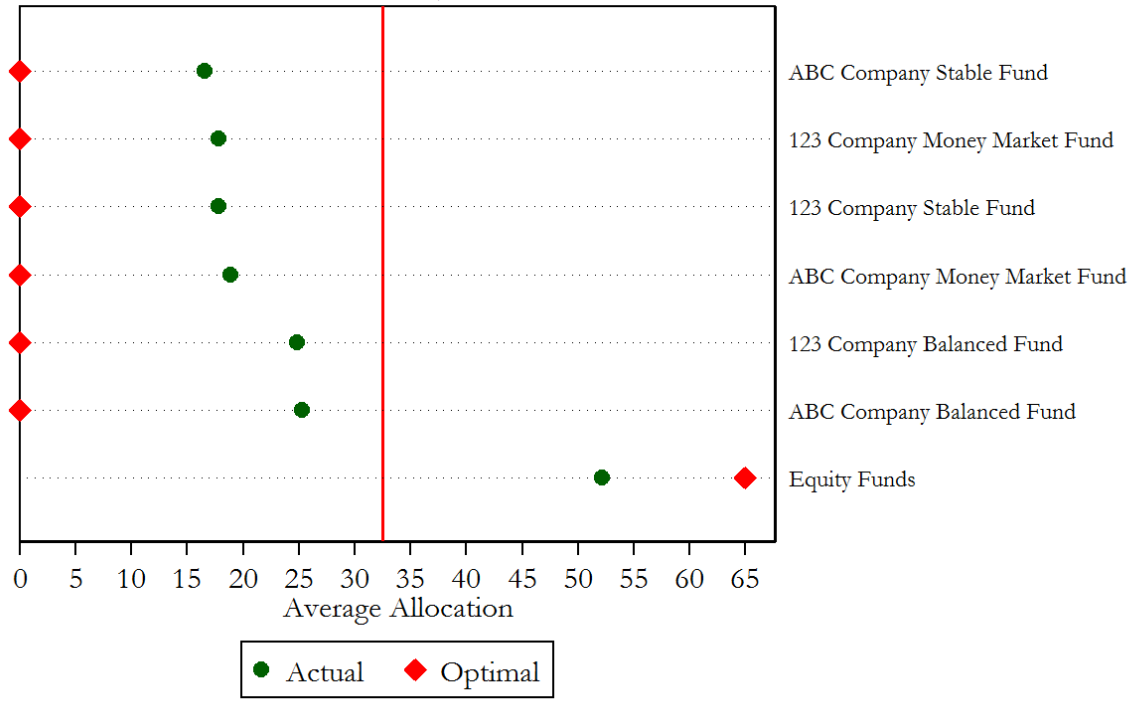


Figure 6: Average Welfare Loss in Control and Treatment Conditions

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

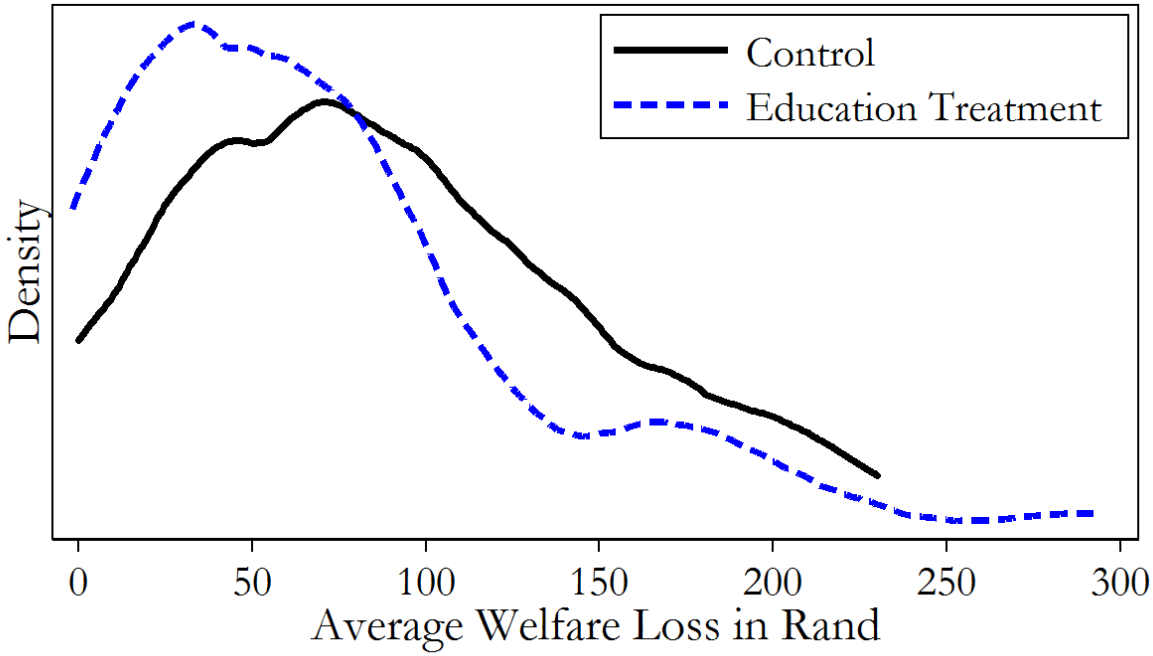


Figure 7: Standard Deviation of Welfare Loss in Control and Treatment Conditions

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

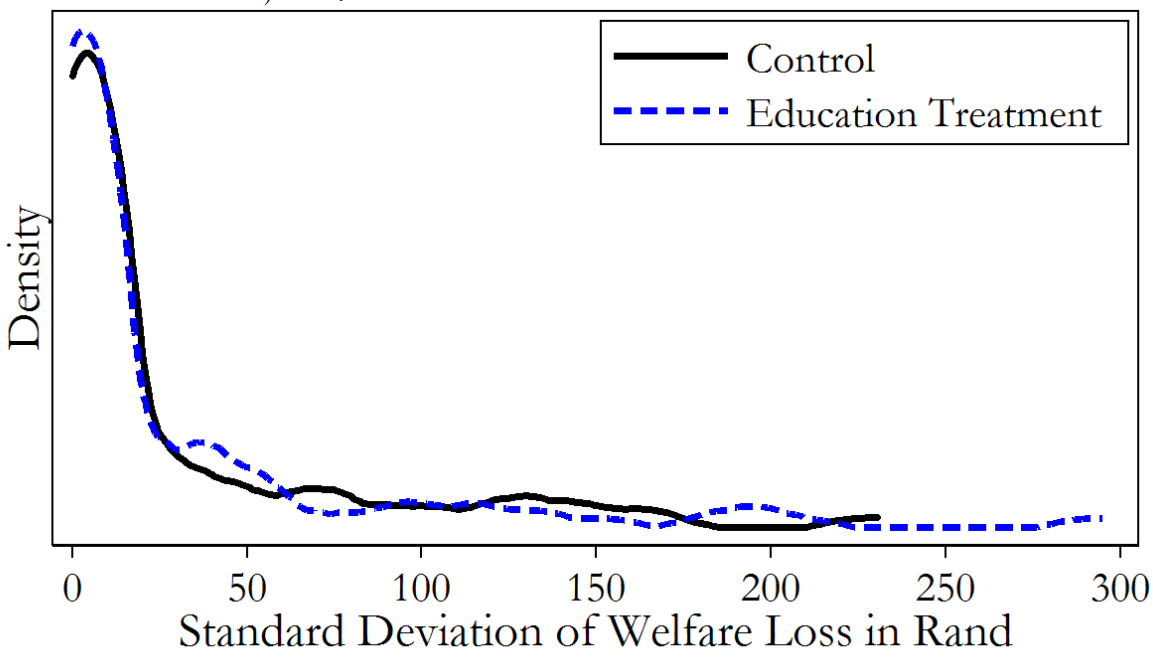


Figure 8: Average Welfare Loss in Control and Treatment Conditions for Expected Utility Theory Subjects

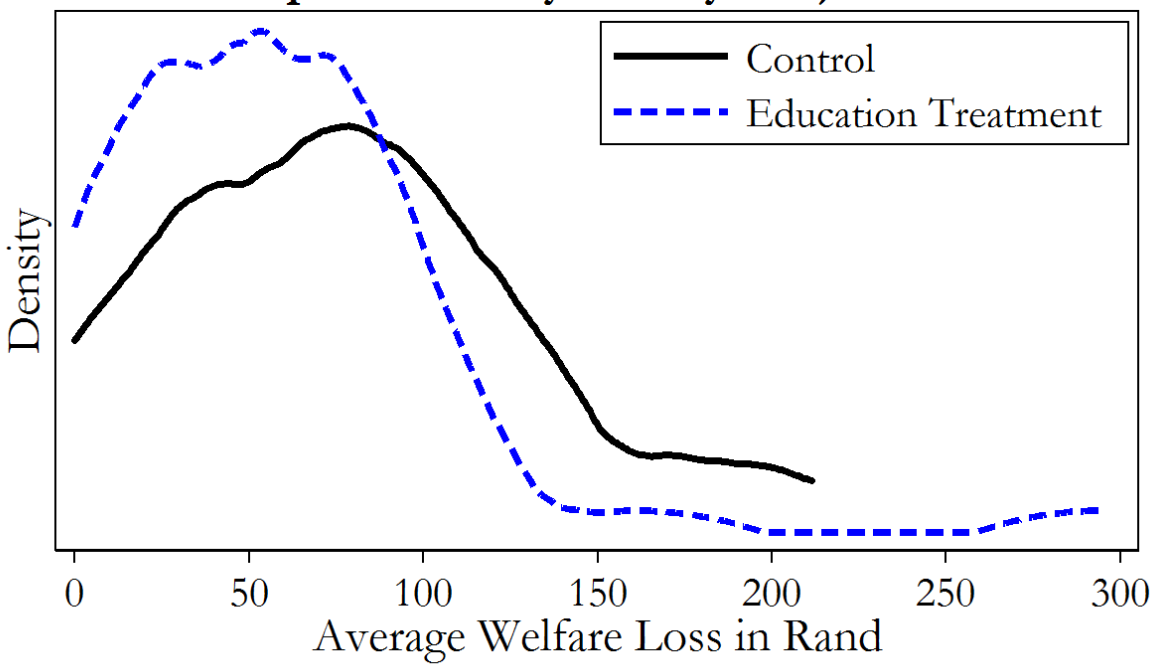


Figure 9: Average Welfare Loss in Control and Treatment Conditions for Rank Dependent Utility Subjects

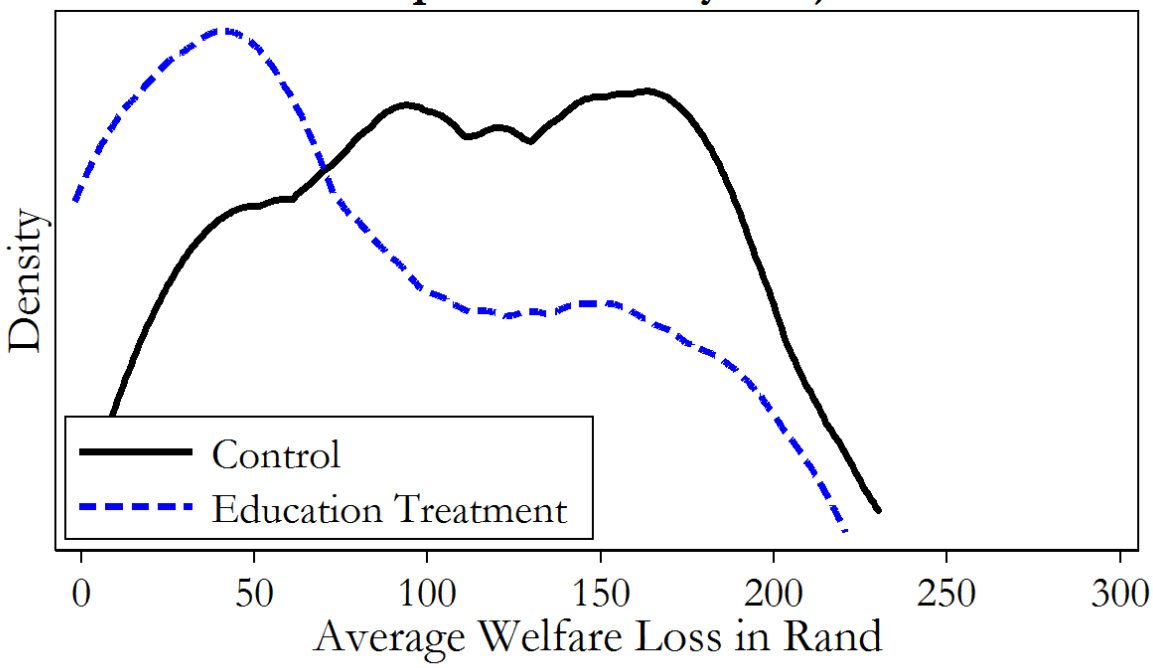


Figure 10: Total Effect on Welfare Loss

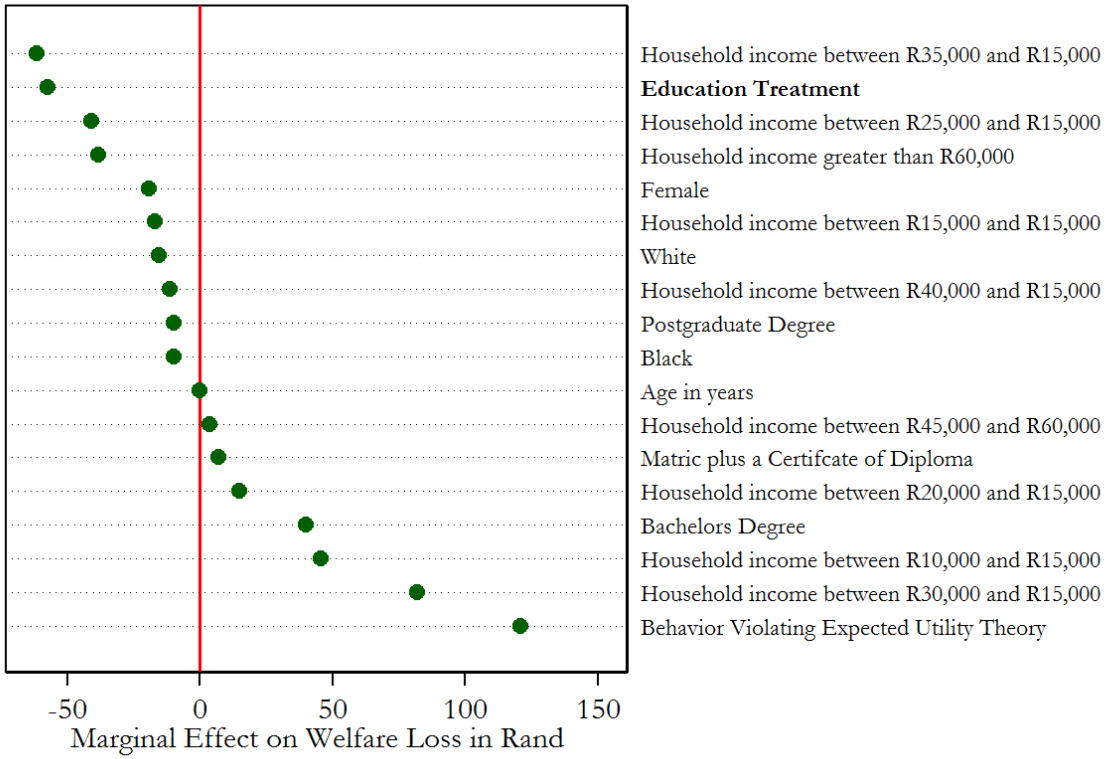


Figure 11: Marginal Effect on Welfare Loss

