Unusual Estimates of Probability Weighting Functions

by

Nathaniel T. Wilcox*

Abstract: I present new estimates of the probability weighting functions found in rank-dependent theories of choice under risk. These estimates are unusual in two senses. First, they are free of functional form assumptions about both utility and weighting functions, and they are entirely based on binary discrete choices and not on matching or valuation tasks, though they depend on assumptions concerning the nature of probabilistic choice under risk. Second, estimated weighting functions contradict widely held priors of an inverse-s shape with fixed point well in the interior of the (0,1) interval: Instead I usually find populations dominated by "optimists" who uniformly overweight best outcomes in risky options. The choice pairs I use here mostly do not provoke similarity-based simplifications. In a third experiment, I show that the presence of choice pairs that provoke similarity-based computational shortcuts does indeed flatten estimated probability weighting functions.

JEL Classification Codes: C25, C91, D81

Keywords: prospect theory, probability weighting, probabilistic choice, rank-dependent utility, risk, similarity

September 2019

*Economic Science Institute (Chapman University) and Center for the Economic Analysis of Risk (Georgia State University). Phone 714-628-7212, email nwilcox@chapman.edu. After December 31 2019, Department of Economics, Appalachian State University, email wilcoxnt@appstate.edu. I have benefitted from conversations with Pavlo Blavatskyy, Jerome Busemeyer, Chew Soo Hong, Jim Cox, Glenn Harrison, John Hey, Stefan Hoderlein, Jonathan Leland, Graham Loomes, Mark Machina, John Quiggin, Michel Regenwetter, Joerg Stoye and an anonymous referee. I thank Stacey Joldersma for her excellent research assistance. This work was financially supported by the University of Houston and Chapman University.

The probability weighting function of the rank-dependent family of choice theories is widely believed to follow an "inverse-s" shape on the unit square—rising steeply at first but concave enough so that it crosses the 45 degree line from above (usually around a third or so though some report estimates near a half) and thereafter becoming convex and accelerating upward to the unit point (e.g. Tversky and Kahneman 1992, Prelec 1998). This belief is so widely and strongly held that many contemporary empirical scholars impose this shape a priori on their estimates of the weighting function (e.g. Nilsson et al. 2011, Scheibehenne and Pachur 2015). Occasionally contrary evidence appears in the empirical literature (e.g. Harrison and Swarthout, this volume), but the vast majority of researchers strongly believe the inverse-s shape holds with a fixed point well inside the interior of the (0,1) interval.

This inverse-s shape explains the well-known Allais phenomena, but so does a strictly concave "optimist" shape (Quiggin 1993). Moreover, there has long been an alternative explanation for the same phenomena based on similarity judgments, frequently associated with Rubinstein (1988) and Leland (1994) but harking back to well-known work by Tversky (1969). Some widely accepted aspects of the probability weighting function might be due to similarity-induced flattening of apparent probability weighting. If so, we might expect to estimate markedly different weighting functions when we confine decision makers' choices to option pairs less likely to bring similarity judgments into play. I explore this and find support for it.

I perform three new risky choice experiments in which the chance device is a single roll of either a six-sided, four-sided or twelve-sided die. In the first two experiments this confines outcome probabilities to a relatively coarse grid (sixths in the first experiment, fourths in the second), so that option pairs rarely present subjects with easy opportunities to exploit similarity-based procedures that ignore small probability differences and bypass a full judgment that weights utilities of outcomes by probability weights. The twelve-sided die used for the third experiment helps to clinch this interpretation of the results. I usually find that the plurality type of decision maker is an optimist—a person whose probability weighting function exceeds all true probabilities of the high outcome used in the experiment (in the second experiment, using the four-sided die, they are an outright majority). This is not the received shape of the probability weighting function.

Note that the data from these three experiments are wholly discrete choices from pairs of risky options. In particular, no valuation tasks or certainty equivalent elicitations are used here. Most evidence for the inverse-s shape and its associated fourfold pattern of risk aversion comes from elicited certainty equivalents (e.g. Tversky and Kahneman 1992; Gonzales and Wu 1999; Abdellaoui 2000; Abdellaoui, Bleichrodt and Paraschiv 2007; Bruhin, Fehr-Duda and Epper 2010), and Wilcox (2017) puts the interpretation of all such evidence in doubt. Moreover, the estimates of probability weights I report here are wholly free of functional form assumptions concerning both outcome utilities and probability weights, using an approach like that of Hey et al. (2010) and Blavatskyy (2013). As mentioned earlier, many studies based on discrete choice from pairs of options impose the inverse-s shape of the weighting function a priori. When we remove studies based on certainty equivalents and do not consider studies that impose it a priori, there is remarkably little consistent evidence for the inverse-s shape that does not also support the optimist shape.

My estimates do depend on assumptions about the probabilistic nature of discrete choice under risk. To guard against the possibility that the results crucially depend on those assumptions, I perform the estimations with three recent but different models of probabilistic choice under risk, all of which have been shown to perform better than older models. The results are, for the most part, insensitive to the choice of one of these models or another. Optimistic decision weights appear to be the norm in my experiments that are relatively free of similarity-based opportunities for choice simplification.

1. Preliminaries

In general, the notation (q_l, q_m, q_h) denotes an option's probability distribution on a vector $\langle l, m, h \rangle$ of three outcomes which I call the <u>context</u> of a choice pair. In the first experiment, each choice pair is a set $\{risky, safe\} \equiv \{(1-q_h, 0, q_h), (0,1,0)\}$ of two options on a context $\langle l, m, h \rangle$. The option $safe = (0,1,0)$ pays m dollars with certainty, while the option $risky = (1-q_h, 0, q_h)$ pays h dollars with probability q_h and l dollars with probability $1-q_h$, where $h > m > l \geq$ US\$40. Subjects choose between $risky$ and $safe$ in each pair presented to them.

The instructions to subjects in Appendix III show a pair where $\{risky, safe\}$ is $\{(5/6,0,1/6), (0,1,0)\}$ on the context $\langle 40,50,90\rangle$. Table 1 shows the 100 choice pairs used in the first experiment, organized into groups of four pairs (the rows of the table) by their shared context. All $risky$ lotteries are chances q_h and $1 - q_h$ (in sixths, generated by a six-sided die) of receiving high and low outcomes h and l on the context, respectively: Four values of q_h, shown on each row in Table 1 (q_h^a, q_h^b, q_h^c and q_h^d) create four $risky$ lotteries on each context, and each of these is paired with $safe$ (the middle outcome m of the context, paid with certainty) to create four pairs on the context. There are twenty-five contexts built from the nine positive money outcomes \$40, \$50,…,\$120.

Table 1: The 100 option pairs of the first experiment.

contexts $\langle l,m,h\rangle$		four pairs			
		q_h^a	q_h^b	q_h^c	q_h^d
1	⟨40,50,60⟩	5/6	4/6	3/6	2/6
2	⟨40,50,70⟩	5/6	4/6	3/6	2/6
3	⟨40,50,80⟩	4/6	3/6	2/6	1/6
4	⟨40,50,90⟩	4/6	3/6	2/6	1/6
5	⟨40,60,100⟩	4/6	3/6	2/6	1/6
6	⟨40,60,110⟩	4/6	3/6	2/6	1/6
7	⟨40,60,120⟩	4/6	3/6	2/6	1/6
8	⟨50,60,90⟩	4/6	3/6	2/6	1/6
9	⟨50,70,100⟩	5/6	4/6	3/6	2/6
10	⟨50,70,110⟩	4/6	3/6	2/6	1/6
11	⟨50,70,120⟩	4/6	3/6	2/6	1/6
12	⟨60,70,90⟩	5/6	4/6	3/6	2/6
13	⟨60,80,110⟩	5/6	4/6	3/6	2/6
14	⟨60,80,120⟩	4/6	3/6	2/6	1/6
15	⟨70,80,100⟩	5/6	4/6	3/6	2/6
16	⟨70,80,110⟩	4/6	3/6	2/6	1/6
17	⟨70,80,120⟩	4/6	3/6	2/6	1/6
18	⟨70,90,110⟩	5/6	4/6	3/6	2/6
19	⟨80,90,100⟩	5/6	4/6	3/6	2/6
20	⟨80,90,110⟩	5/6	4/6	3/6	2/6
21	⟨80,90,120⟩	4/6	3/6	2/6	1/6
22	⟨80,100,120⟩	5/6	4/6	3/6	2/6
23	⟨90,100,110⟩	5/6	4/6	3/6	2/6
24	⟨90,100,120⟩	5/6	4/6	3/6	2/6
25	⟨100,110,120⟩	5/6	4/6	3/6	2/6

The subjects in the first experiment were 80 undergraduates at the University of Houston, recruited widely by means of a single e-mail to all undergraduates. Each subject was individually scheduled for three separate sessions on three separate days of their own choosing, almost always finishing all three sessions within a week. Only rarely did any day's session last more than an hour, and most sessions were substantially shorter than this. On each day, each subject made choices from the 100 choice pairs shown in Table 1, so that each made 300 choices in all by the end of their third day. On each day, for each subject, the 100 choice pairs were randomly ordered into two halves of 50 pairs each, separated by about ten to fifteen minutes of other tasks (a survey and tests of arithmetic and problem-solving ability). A computer presented each choice pair, one pair at a time on each screen: Call this separated decisions or SED.

To conclude each subject's third day, one of their 300 chosen options was selected at random (by means of the subject drawing a ticket from a bag of 300 numbered tickets) to determine the subject's payment: Call this random problem selection or RPS. If the subject's selected option was $risky$, the subject picked a six-sided die from a box of six-sided dice (rolling them until satisfied if they wished), and that die was then rolled by the attendant to determine the payment. A detailed explanation of this protocol, as well as instructions to subjects, appears in Appendix III.

Quiggin (1982) originally developed rank-dependent utility or RDU; later, Quiggin's rank-dependent probability weighting function became a part of cumulative prospect theory or CPT (Tversky and Kahneman 1992). Under RDU (or CPT for pure gain options), the value of an option (q_l, q_m, q_h) is

$$(1)\quad RDU(q_l, q_m, q_h) = \quad w(q_h)u(h) + [w(1-q_l) - w(q_h)]u(m) + [1 - w(1-q_l)]u(l),$$

where $u(z)$ is the utility of outcome z and $w(q)$ is the probability weighting function at q. In the first experiment, the RDU value difference between $risky$ and $safe$ is simply

$$(2)\quad \Delta RDU = RDU(risky) - RDU(safe) = w(q_h)u(h) \ + \ [1 - w(q_h)]u(l) - u(m).$$

I wish to estimate the utilities $u(z)$ and weights $w(q_h)$ of RDU (or CPT for pure gains) with no assumptions concerning their functional form, and using only binary choice data. To do this, I need assumptions about the nature of probabilistic discrete choice from option pairs.

2. The probabilistic choice models

Beginning with Mosteller and Nogee (1951), many experiments on discrete choice under risk suggest that these choices have a strong probabilistic component. Repeated trials of choice from pairs of risky options reveal high rates of choice switching by the same subject between trials of the same pair. In some cases, the repeated trials span days (e.g. Tversky 1969; Hey and Orme 1994; Hey 2001) and decision-relevant conditions might have changed between trials. Yet switching occurs even between trials separated by bare minutes, with no intervening change in wealth, background risk, or any other obviously decision-relevant variable (Camerer 1989; Starmer and Sugden 1989; Ballinger and Wilcox 1997; Loomes and Sugden 1998).

To construct observation likelihoods, assumptions about the probabilistic nature of these choices are needed. I use three different probabilistic choice models of the form

$$(3)\ \ P \equiv Prob(risky\ chosen\ from\ \{risky, safe\}) = F\left(\lambda \frac{\Delta RDU}{D(risky,safe)}\right),$$

where λ is a scale (or inverse standard deviation) parameter, $D(risky, safe)$ adjusts the scale parameter, and $F\colon X \to [0,1]$ is an increasing function with $F(0) = 0.5$ and $F(x) = 1 - F(-x)$, where $X \subseteq \mathbb{R}$. The probabilistic models are my own "contextual utility" or CU model (Wilcox 2011), the "decision field theory" or DFT model of Busemeyer and Townsend (1992, 1993) and the "stronger utility" or SU model of Blavatskyy (2014). Respectively, these models are:

$$(4)\ \ P^{cu} = Prob(risky) = F\left(\lambda \frac{\Delta RDU}{u(h)-u(l)}\right), \text{ contextual utility;}$$

$$(5)\ \ P^{dft} = Prob(risky) = F\left(\lambda \frac{\Delta RDU}{[u(h)-u(l)]\sqrt{w(q_h)[1-w(q_h)]}}\right), \text{ decision field theory; and}$$

$$(6)\ \ P^{su} = Prob(risky) = H_\lambda\left(\frac{\Delta RDU}{w(q_h)[u(h)-u(m)]+[1-w(q_h)][u(m)-u(l)]}\right), \text{ stronger utility.}$$

In the contextual utility and decision field theory models, $X = \mathbb{R}$, while in stronger utility $X = (-1,1)$. However, by way of a suitable choice of H_λ, the stronger utility model can be rewritten in a form with $F\colon \mathbb{R} \to [0,1]$ as well (see Appendix I):

(7) $P^{su} = Prob(risky) = F\left[\lambda \ln\left(\frac{w(q_h)[u(h)-u(m)]}{[1-w(q_h)][u(m)-u(l)]}\right)\right]$.

This means that all three of these probabilistic models may be estimated using a common choice for the function F. Busemeyer and Townsend (1993) give theoretical reasons for choosing the logistic c.d.f. $\Lambda(x) = [1 + \exp(-x)]^{-1}$ for use with decision field theory (see Appendix I) so I use it as the function F in all estimations for all three models.

Until recently (e.g. Anderson et al. 2008), most applied econometric estimations would have been done with the simple homoscedastic latent variable model $P^h = Prob(risky) = F(\lambda\Delta RDU)$: I call this the homoscedastic model. For many reasons, much professional opinion has turned against the homoscedastic model for discrete choice under risk. Long ago Luce (1959) remarked that the ratio scale nature of probabilistic models satisfying the Choice Axiom (for instance the homoscedastic binary logit) is deeply inconsistent with interval scale theories (such as EU and RDU). Since then Loomes and Sugden (1995) noted that the homoscedastic model does not respect stochastic dominance. Still more damaging, Blavatskyy (2011), Wilcox (2011), and Apesteguia and Ballester 2018) all show that this model cannot coherently represent comparative risk aversion across agents in different choice contexts. The laboratory evidence against the homoscedastic model, for choice under risk, is now extensive (Busemeyer and Townsend 1993; Loomes and Sugden 1998; Rieskamp 2008; Wilcox 2008, 2011, 2015; Butler, Isoni and Loomes 2012; Blavatskyy 2014). Previous applied econometric users of the simple homoscedastic model have put it aside in favor of the newer models (e.g. Anderson et al. 2013). Appendix I presents more information on contextual utility, decision field theory and stronger utility.

There are other ways to introduce probabilistic choice into models of decision under risk. One of these is random preferences (Loomes and Sugden 1995; Gul and Pesendorfer 2006): This approach treats vectors of outcome utilities and/or probability weights as random draws from a fixed distribution of these vectors. Random preference models also exhibit context dependence (Wilcox 2011, p. 101). There is, however, a difficult problem with considering a random preference RDU specification for the experimental data considered here: It is not possible to generalize an RDU random preference specification across more than three outcome contexts without changing estimation techniques in fundamental ways (Wilcox 2008 pp. 252-256; Wilcox 2011 pp.101-102). The first experiment has 25 distinct outcome contexts while the second and third experiments have 10 each. Therefore, I do not consider random preferences here.

3. Estimation

To discuss the estimation, it is helpful to define indices for pairs, trials (days) and subjects, as well as some important sets of indices:

$i = 1,2, \dots I$, indexing I distinct pairs. Here $I = 100$.

Pairs i are $\{risky_i, safe_i\} \equiv \{(1-q_{hi}, 0, q_{hi}), (0,1,0)\}$ on context $\langle l_i, m_i, h_i \rangle$.

t= 1,2,…τ_i, indexing τ_i distinct trials of each pair i. Here $\tau_i = 3$ (three days).

s= 1,2,…S, indexing the S distinct subjects. Here $S = 80$.

it: A double subscript indicating the tth trial of pair i.

$r_{it}^s = 1$ if subject s chose $risky_i$ in her tth trial of pair i, and zero otherwise.

$\mathbf{r}^s$ = the observed choice vector of subject s over all pairs and trials it.

Let $u^s(z)$ and $w^s(q)$ denote utilities of outcomes z and weights associated with probabilities q, respectively, of subject s. The first experiment involves nine distinct outcomes $z \in \{\$40, \$50, \dots, \$120\}$ across its 100 choice pairs, but because of the affine transformation invariance property of RDU and EU utilities, we can choose $u^s(40) = 0$ and $u^s(120) = 1$ for all subjects s. With this done, the unique estimable utility vector $\mathbf{u}^s$ of the seven remaining outcomes is $\mathbf{u}^s = \langle u^s(50), u^s(60), \dots, u^s(110) \rangle$. Function-free estimations make each of those seven utilities a separate parameter to be estimated.

The first experiment involves five distinct probabilities $q_h \in \left\{\frac{1}{6}, \frac{2}{6}, \dots, \frac{5}{6}\right\}$, so there is a vector $\mathbf{w}^s = \langle w^s\left(\frac{1}{6}\right), w^s\left(\frac{2}{6}\right), \dots, w^s\left(\frac{5}{6}\right) \rangle$ of five weights to estimate for each subject. Function-free estimations make each of those five weights a separate parameter to be estimated. To summarize, the function-free latent index of the RDU representation, for subject s and pair i, is

(8) $\Delta RDU_i(\mathbf{u}^s, \mathbf{w}^s) = w^s(q_{hi})u^s(h_i) + [1 - w^s(q_{hi})]u^s(l_i) - u^s(m_i)$, where

$\mathbf{w}^s = \langle w^s\left(\frac{1}{6}\right), w^s\left(\frac{2}{6}\right), \ldots, w^s\left(\frac{5}{6}\right)\rangle$, and

$\mathbf{u}^s = \langle u^s(50), u^s(60), \ldots, u^s(110)\rangle$, with $u^s(40) = 0$ and $u^s(120) = 1\ \forall\ s$.

Combine eq. 8 with eqs. 4, 5 and 7, let $\boldsymbol{\theta}^s \equiv (\mathbf{u}^s, \mathbf{w}^s, \lambda^s)$ and choose the logistic c.d.f. as $F(x)$, and we have the following choice probability specifications:

(9) $P_i^{rdcu}(\boldsymbol{\theta}^s) = \Lambda\left[\lambda^s \frac{\Delta RDU_i(\mathbf{u}^s,\mathbf{w}^s)}{u^s(h_i)-u^s(l_i)}\right]$;

(10) $P_i^{rddft}(\boldsymbol{\theta}^s) = \Lambda\left[\lambda^s \frac{\Delta RDU_i(\mathbf{u}^s,\mathbf{w}^s)}{[u^s(h_i)-u^s(l_i)]\sqrt{w^s(q_{hi})[1-w^s(q_{hi})]}}\right]$; and

(11) $P_i^{rdsu}(\boldsymbol{\theta}^s) = \Lambda\left[\lambda^s \ln\left(\frac{w^s(q_{hi})[u^s(h_i)-u^s(m_i)]}{[1-w^s(q_{hi})][u^s(m_i)-u^s(l_i)]}\right)\right]$.

Equations 9-11 give the probability of events $r_{it}^s = 1$ (subject s chose $risky$ in the tth trial of pair i). Letting $P_i^{spec}(\boldsymbol{\theta}^s)$ denote any of those probabilities, the log likelihood of $\mathbf{r}^s$ is

(12) $\mathcal{L}^{spec}(\mathbf{r}^s|\boldsymbol{\theta}^s) = \sum_{it} r_{it}^s \ln\left[P_i^{spec}(\boldsymbol{\theta}^s)\right] + (1 - r_{it}^s)\ln\left[1 - P_i^{spec}(\boldsymbol{\theta}^s)\right]$.

I estimate $\boldsymbol{\theta}^s$ by a penalized maximum likelihood procedure, for each subject s; Appendix II contains details of this estimation.

4. Some Monte Carlo results

The 100 choice pairs in Table 1 were in part chosen through Monte Carlo simulations exploring estimation performance with alternative sets of choice pairs. To gain confidence in the estimations reported here—and to understand their limitations—it helps to see some Monte Carlo results. Consider a data generating process or DGP based on one of the choice probability models in equations 9-11, combined with well-known parametric estimates of utility and weighting functions. For the utility function, I use the CRRA utility of money given by $u^s(z|\rho^s) = z^{1-\rho^s}/(1-\rho^s)$, normalized[1] so that $u^s(40) = 0$ and $u^s(120) = 1$, and begin with the parameter value $\rho^s = 0.12$ (very mild concavity of utility)

[1] This normalized version of CRRA utility is simply $u^s(z|\rho^s) = \left(z^{1-\rho^s} - 40^{1-\rho^s}\right)/\left(120^{1-\rho^s} - 40^{1-\rho^s}\right)$.

reported by Tversky and Kahneman (1992). For the weighting function, I use Prelec's (1998) two-parameter function, given by $w^s(q|\beta^s,\gamma^s) = \exp\left(-\beta^s[-\ln(q)]^{\gamma^s}\right) \;\forall\; q \in (0,1)$, $w(0) = 0$ and $w(1) = 1$, and begin with parameter values $\beta^s = 1$ and $\gamma^s = 0.65$ which, according to Prelec, match earlier estimations of weights using other weighting functions. Express the parameters of this first DGP as $(\rho,\gamma,\beta) = (0.12,0.65,1)$: These parameters are cumulative prospect theory as first conceived a quarter century ago, and I call this "Prospector I" for short. I take a more recent parametric version of cumulative prospect theory from Bruhin, Fehr-Duda and Epper (2010), using a second DGP $(\rho,\gamma,\beta) = (0.043,0.45,0.8)$. I call this "Prospector II" for short: It closely resembles Bruhin, Fehr-Duda and Epper's most common subject type (that they estimated with a finite mixture model using all of their data).

For contrast, and anticipating later results, I examine two other DGPs. One of these DGPs is $(\rho,\gamma,\beta) = (3,1.5,0.4)$: I call this DGP "Optimist" since its weighting function is such that $w^s(q) > q$ for all q in the first experiment: The decision maker overweights probabilities of highest outcomes. By itself such probability weighting would imply risk-seeking, but this DGP also has a highly concave utility function which, by itself, would imply risk aversion. The last DGP for Monte Carlo study is $(\rho,\gamma,\beta) = (1.5,3,2)$: The implied weighting function in this case is s-shaped—opposite of the inverse s-shape of received Cumulative Prospect Theory. This weighting function may represent a decision maker who sometimes rounds low probabilities to zero and high probabilities to unity, so I call this DGP "Rounder."

Figures 1, 2, 3 and 4 show results of function-free estimations of utilities (the left panels) and weights (the right panels) for 80 simulated subjects, using the contextual utility specification of eq. 9 for the estimation. These simulated subjects all have true (DGP) choice probabilities given by the contextual utility model in eq. 9 with $\lambda^s = 12$. In Figure 1, the 80 simulated subjects have the "Prospector I" DGP; in Figure 2 they have the "Prospector II" DGP; in Figure 3 they have the "Optimist" DGP; and in Figure 4 they have the "Rounder" DGP. On all panels, the true (DGP) utility functions or weighting functions appear as a bold black curve, while the 80 functions estimated using the function-free method appear as thinner curves of varying greys.

Figures 1, 2, 3 and 4 show that the function-free estimates cluster around the DGP curve with little in the way of strong biases, except occasionally near the endpoints of the functions where true utilities and/or weights are close to zero or one (this is expected for maximum likelihood estimates of parameters lying near a boundary of an allowed parameter space). The variability of the estimated curves (not small) is due both to the inherent variability of (simulated) observed choices that is consequent to probabilistic choice in the DGP, and to the burden of the function-free estimation.[2] Yet comparison of these four figures shows that collected function-free estimations track different DGPs: The collective impression made by the "cloud" of individual estimates matches different amounts of utility concavity and different weighting function shapes quite well.

Tables 2-A, 2-B and 2-C show distributions of five estimated weighting function shapes for 1000 simulated subjects, using each of the four DGPs:

(1) prospector— there is a $q^* \in \left(\frac{1}{6},\frac{5}{6}\right)$ such that $\hat{w}^s(q) \gtreqless q$ as $q \lesseqgtr q^*$;

(2) pessimist— $\hat{w}^s(q) < q$ for all q;

(3) optimist— $\hat{w}^s(q) > q$ for all q;

(4) rounder— there is a $q^* \in \left(\frac{1}{6},\frac{5}{6}\right)$ such that $\hat{w}^s(q) \lesseqgtr q$ as $q \lesseqgtr q^*$; and

(5) unclassified— estimated weights cross the identity line more than once.

The tables show that most estimated weighting function shapes match the shape of the DGP (usually more than 80%, but a bit less for Prospector I). These tables also bear on later results. First, Cumulative Prospect Theory DGPs (that is, Prospector I and Prospector II) produce estimated optimist or rounder shapes less than about 8% of the time: If a sample of 80 subjects comes from a population composed solely of Prospector I and Prospector II, we expect that function-free estimation will produce about 7 subjects having estimated optimist or rounder shapes. Second, Cumulative Prospect Theory DGPs produce estimated pessimist shapes about 11% of the time: If we see 8 or so estimated pessimist shapes, these

[2] Parametric estimations produce estimates with about eighty to fifty percent of the variability of these function-free estimates (around the true DGP weighting function).

Figure 1: 80 function-free estimates from Monte Carlo data with 'Prospector I' DGP $(\rho, \gamma, \beta) = (0.12, 0.65, 1)$ and contextual utility.

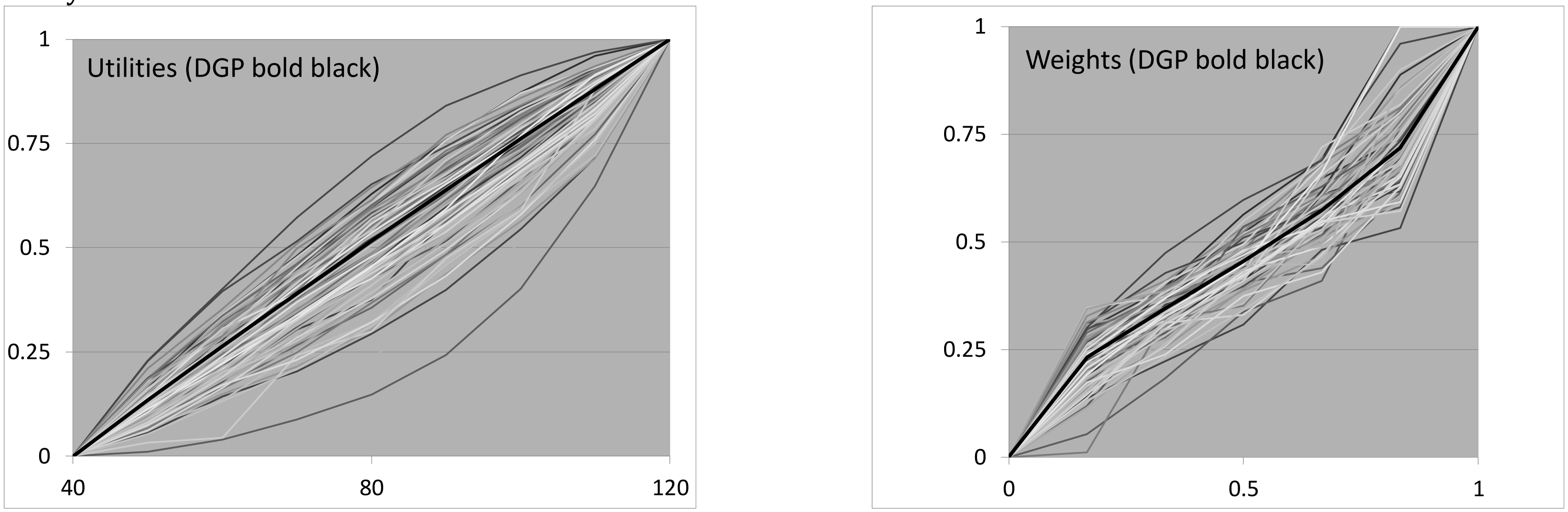


Figure 2: 80 function-free estimates from Monte Carlo data with 'Prospector II" DGP $(\rho, \gamma, \beta) = (0.043, 0.45, 0.8)$ and contextual utility.

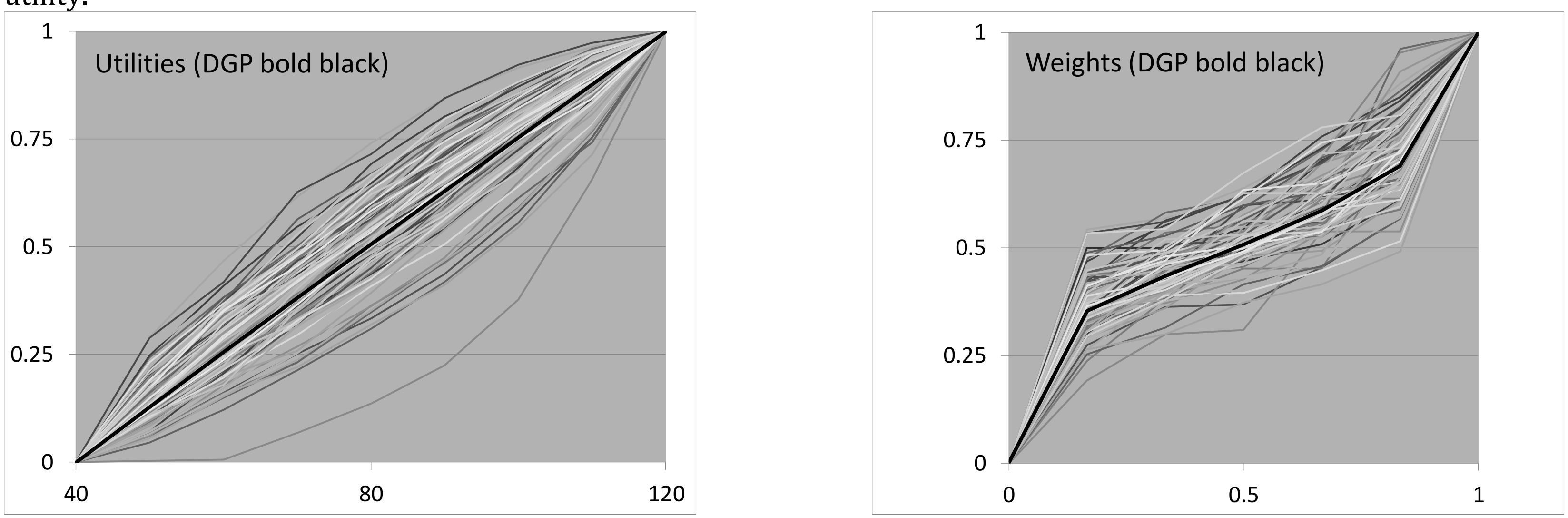


Figure 3: 80 function-free estimates from Monte Carlo data with 'Optimist' DGP $(\rho, \gamma, \beta) = (3, 1.5, 0.4)$ and contextual utility.

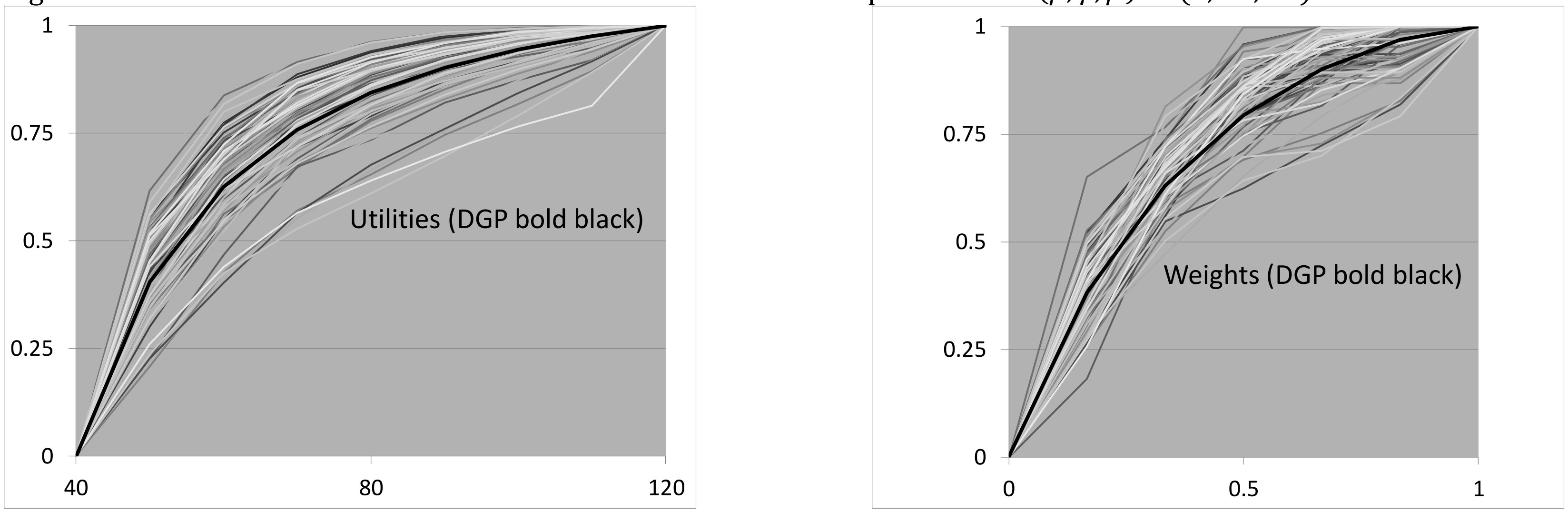
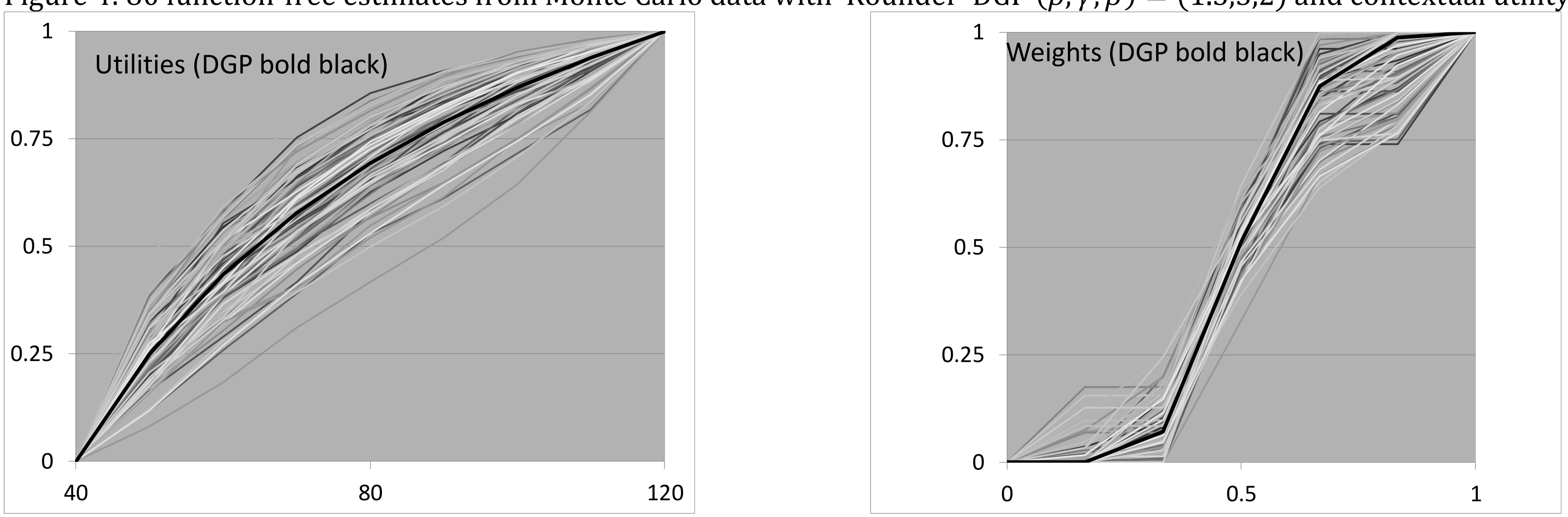


Figure 4: 80 function-free estimates from Monte Carlo data with 'Rounder' DGP $(\rho, \gamma, \beta) = (1.5, 3, 2)$ and contextual utility.



Tables 2. Monte Carlo results: Distribution of 1000 function-free estimations of weighting function shapes, using 1000 simulated subjects from four different DGPs.

Table 2-A. Contextual utility ($\lambda^s = 12$) is in the DGP and also used for estimations.

estimated weighting function shapes	DGP utility and weighting function parameters (ρ, γ, β)			
	Prospector I, as in Figure 1 (0.12,0.65,1)	Prospector II, as in Figure 2 (0.043,0.45,0.8)	Optimist, as in Figure 3 (3,1.5,0.4)	Rounder, as in Figure 4 (1.5,3,2)
Prospector	66.7%	84.1%	8.4%	0.0%
Pessimist	10.9%	0.4%	0.0%	2.4%
Optimist	4.6%	8.1%	91.0%	0.0%
Rounder	1.8%	0.0%	0.6%	81.0%
Unclassifiable	16.0%	7.4%	0.0%	16.6%

Table 2-B. Decision field theory ($\lambda^s = 5$) is in the DGP and also used for estimations.

estimated weighting function shapes	DGP utility and weighting function parameters (ρ, γ, β)			
	Prospector I, as in Figure 1 (0.12,0.65,1)	Prospector II, as in Figure 2 (0.043,0.45,0.8)	Optimist, as in Figure 3 (3,1.5,0.4)	Rounder, as in Figure 4 (1.5,3,2)
Prospector	74.4%	89.8%	20.1%	0.0%
Pessimist	11.7%	2.0%	0.0%	0.0%
Optimist	3.2%	4.9%	79.1%	0.0%
Rounder	1.4%	0.1%	0.5%	88.0%
Unclassifiable	9.3%	3.2%	0.3%	12.0%

Table 2-C. Stronger utility ($\lambda^s = 2$) is in the DGP and also used for estimations.

estimated weighting function shapes	DGP utility and weighting function parameters (ρ, γ, β)			
	Prospector I, as in Figure 1 (0.12,0.65,1)	Prospector II, as in Figure 2 (0.043,0.45,0.8)	Optimist, as in Figure 3 (3,1.5,0.4)	Rounder, as in Figure 4 (1.5,3,2)
Prospector	67.4%	88.5%	4.0%	0.0%
Pessimist	12.6%	1.1%	0.0%	0.4%
Optimist	4.2%	5.9%	94.2%	0.0%
Rounder	3.3%	0.2%	1.8%	95.6%
Unclassifiable	12.5%	4.3%	0.0%	4.0%

may be simply the result of sampling variability and true Cumulative Prospect Theory types in the sampled population. Finally, Prospector I DGPs produce unclassified shapes about 10% to 15% of the time, so we should not be surprised to estimate a small number of unclassified shapes if Prospector I DGPs (or other DGPs relatively close to identity weights) are common in the sampled population.

I noted that until recently, the simple homoscedastic latent index model $P^h = Prob(risky) = F(\lambda \Delta RDU)$ was commonly used for such estimations. Figure 5 shows some consequences of such a homoscedastic latent index estimation when the DGP in fact features any of the three probabilistic models I use here. The DGP utility and weighting function is in all cases the Optimist DGP, that is $(\rho, \gamma, \beta) = (3,1.5,0.4)$, which also features pronounced concavity of utility. Figure 5 shows that for all three DGPs, this results in reliable underestimation of both utility concavity and weighting optimism: Almost all of the 80 estimates lie below the bold black DGP curves. The newer heteroscedastic probabilistic models are consequential for estimation and inferences concerning utility and weighting functions and, as mentioned earlier, there is now much evidence against the homoscedastic model.

5. Results of the first experiment

Figures 6, 7 and 8 show most of the results of the function-free individual estimations: Figure 6 shows contextual utility estimations; Figure 7 shows decision field theory estimations; and Figure 8 shows stronger utility estimations. In each figure, the upper left panel shows 80 estimated utility functions while the remaining three panels show most (at least 68 of 80) estimated weighting functions, divided into the three most commonly estimated shapes—optimists, rounders and prospectors, generally in that order (except with decision field theory). The remaining 12 subjects (whose estimated weighting functions are not shown) break almost evenly between pessimists and unclassified,[3] certainly consistent with the sampling variability considerations of the previous section and not strong evidence that these types even exist in the sampled population. Overall, by individual-level

[3] 6 of each with contextual utility, 4 pessimists and 7 unclassified with decision field theory, and 7 pessimists and 5 unclassified with stronger utility.

Figure 5: 80 function-free estimates from Monte Carlo data with ‘Optimist’ DGP $(\rho, \gamma, \beta) = (3,1.5,0.4)$, estimated with the homoscedastic model, when the true DGP uses one of the three heteroscedastic models:

DGP is contextual utility (CU) | DGP is decision field theory (DFT) | DGP is stronger utility (SU)

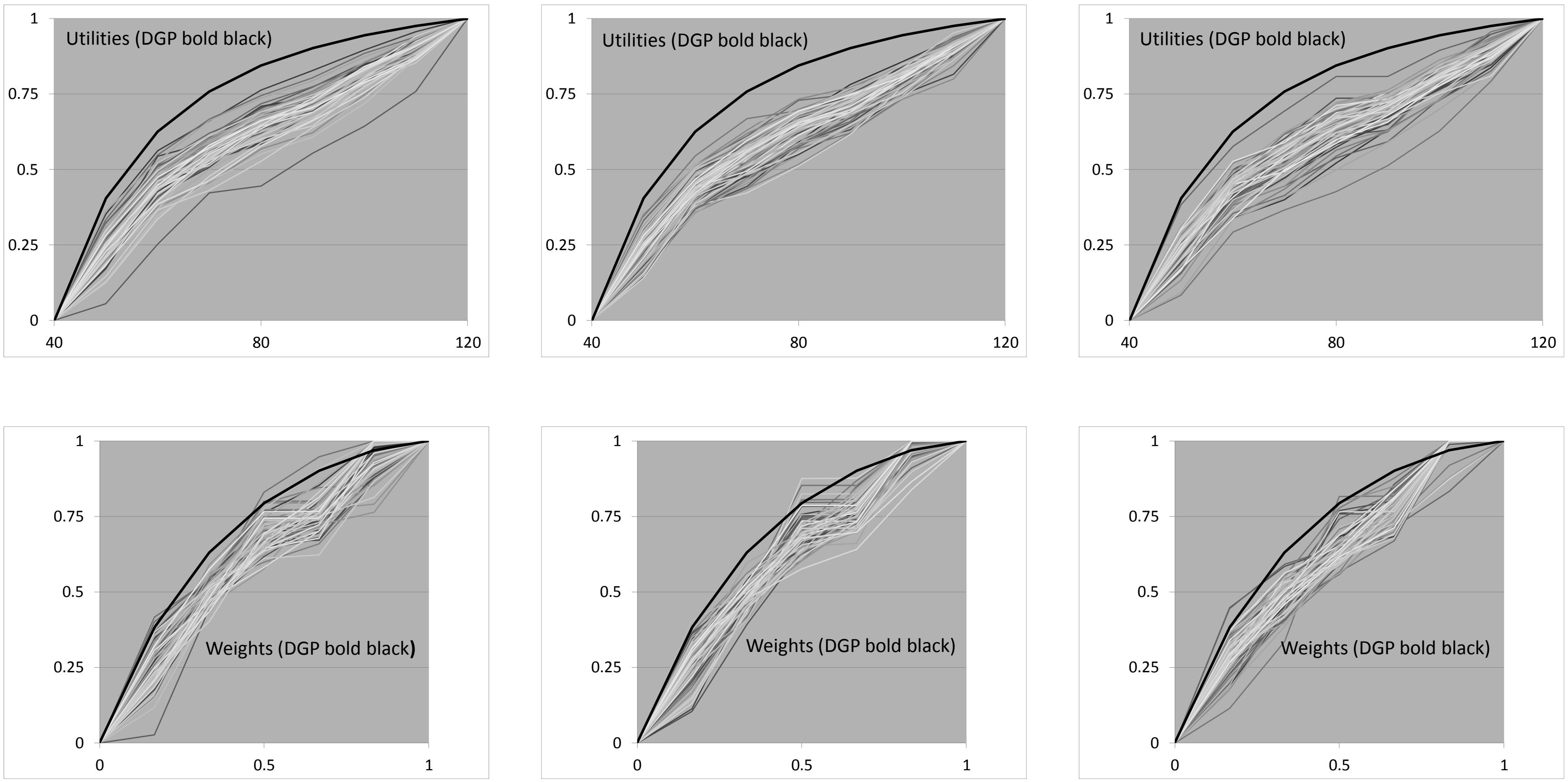


Figure 6: 80 function-free individual estimates, estimated with the contextual utility model using data from the first experiment. The first panel shows estimated utility functions together; the next three panels show estimated weighting functions of the three most common estimated shapes (68 of the 80 subjects). The median estimated λ^s is about 11.3.

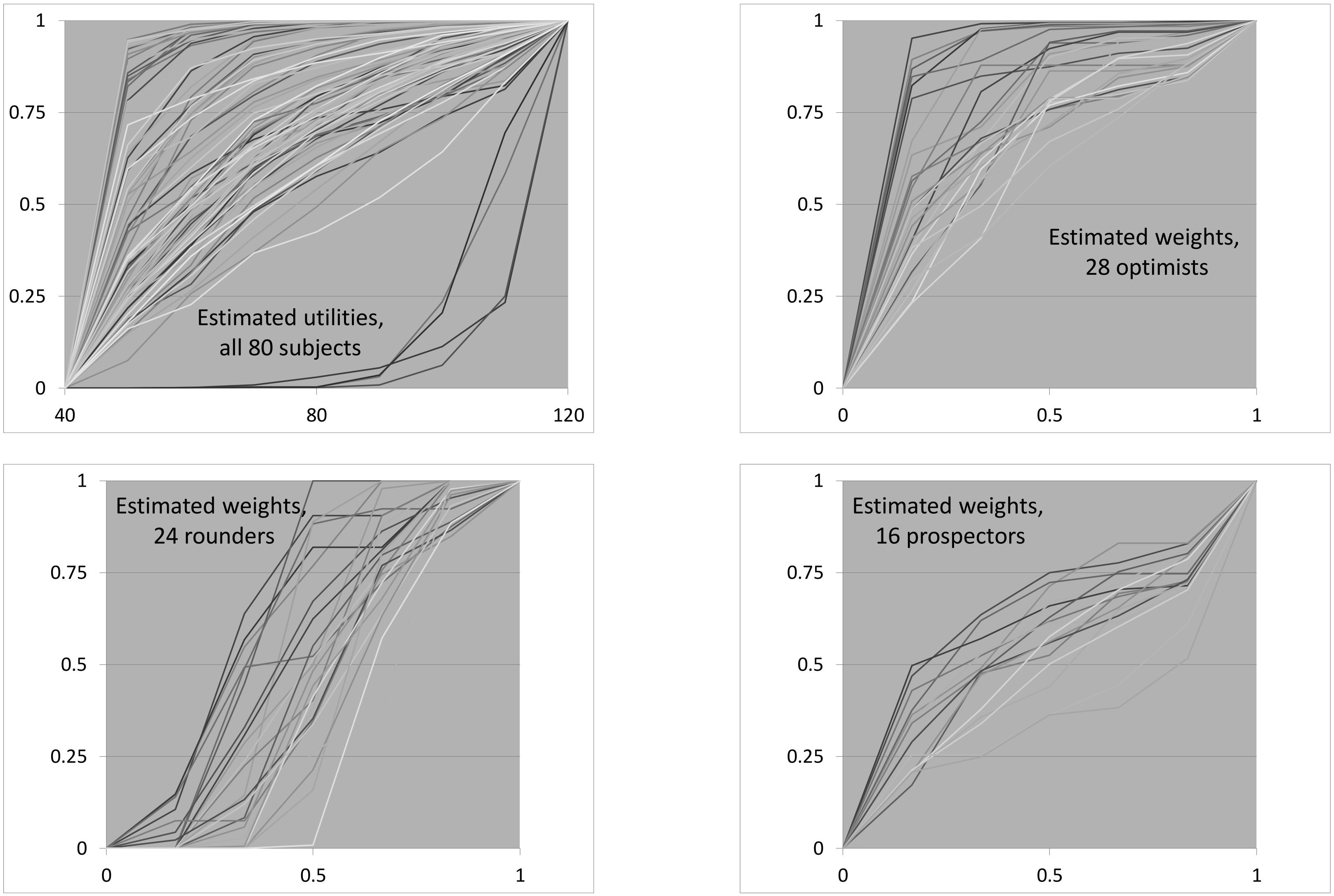


Figure 7: 80 function-free individual estimates, estimated with the decision field theory model using data from the first experiment. The first panel shows estimated utility functions together; the next three panels show estimated weighting functions for the most commonly estimated shapes (69 of the 80 subjects). The median estimated λ^s is about 5.15.

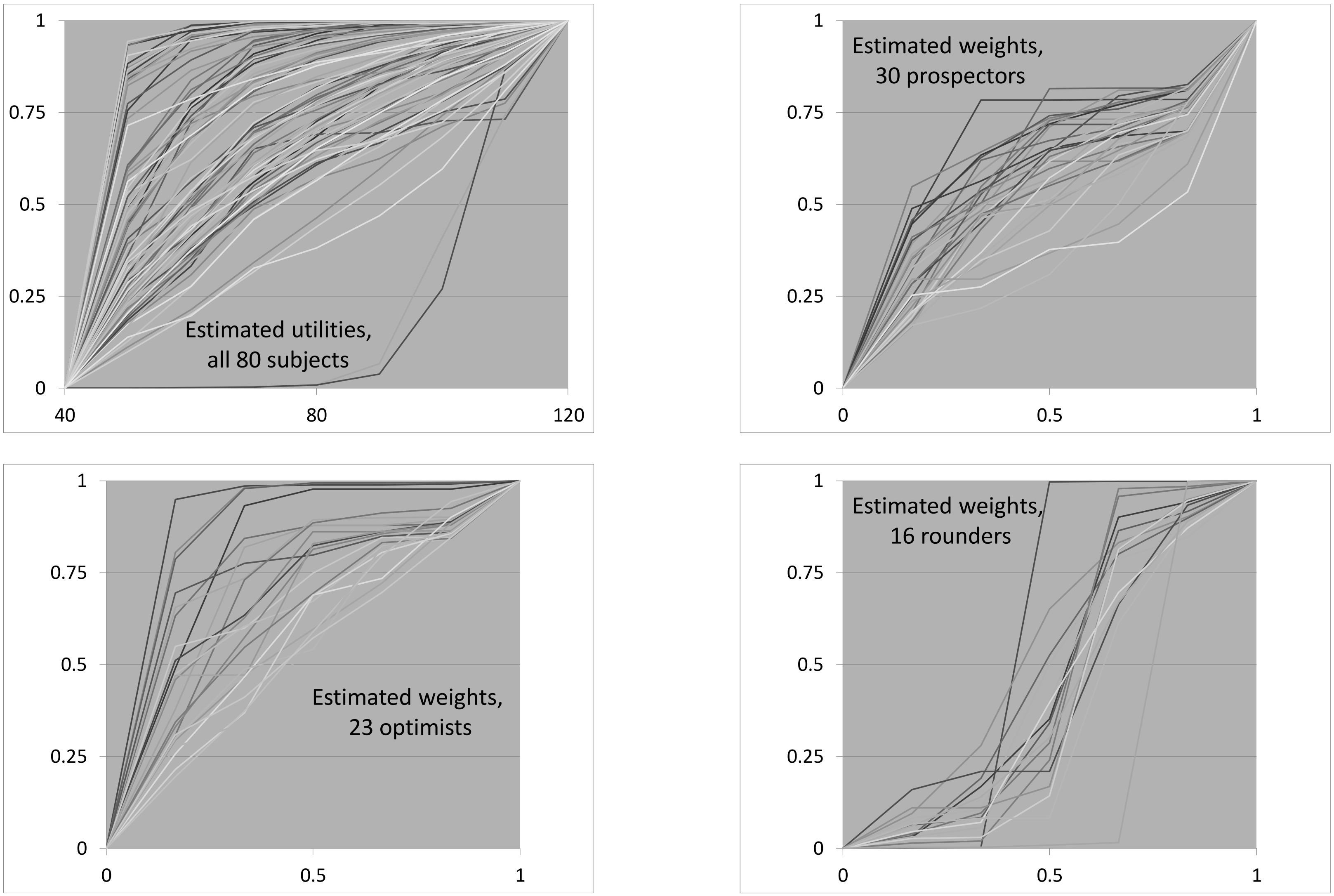
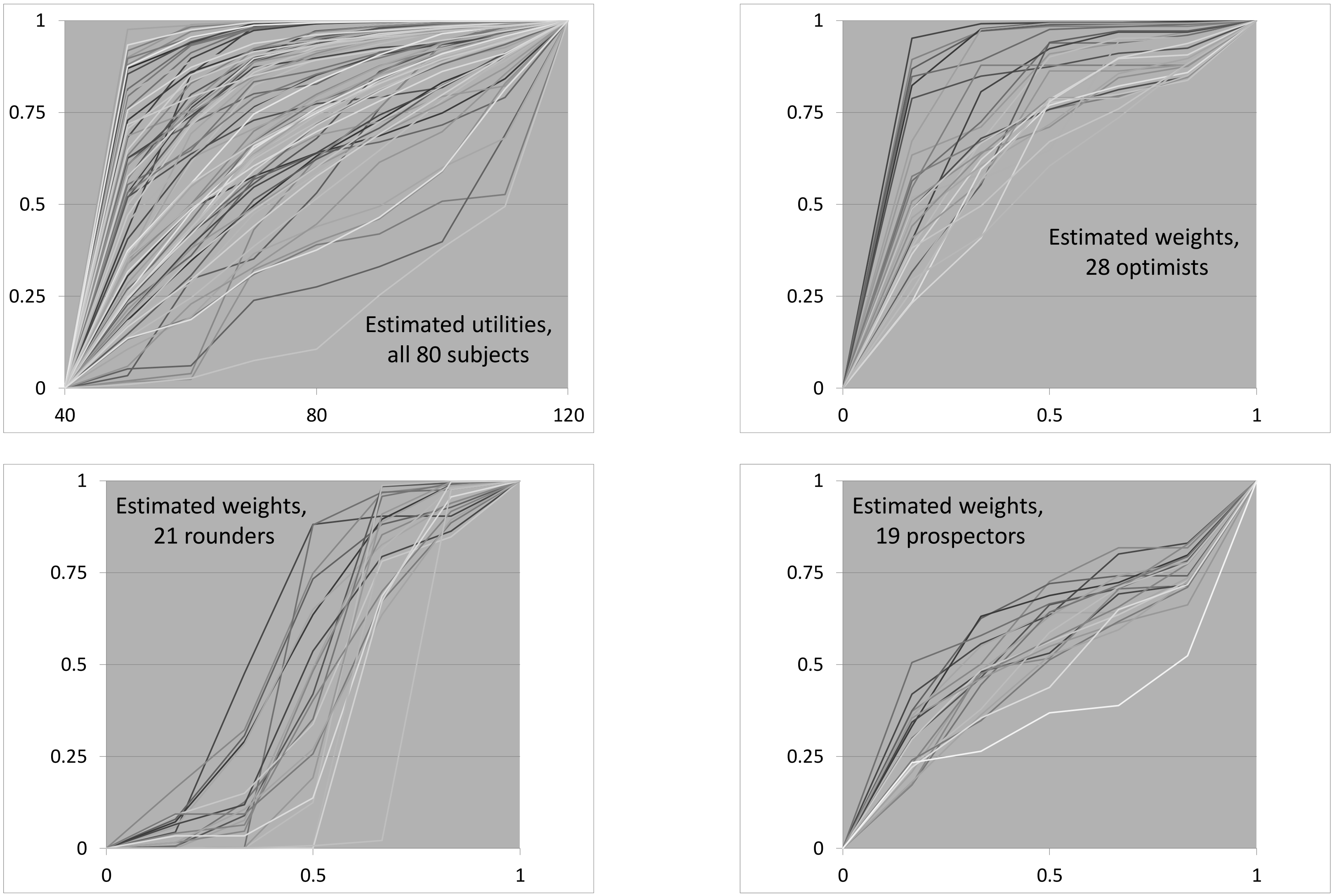


Figure 8: 80 function-free individual estimates, estimated with the stronger utility model using data from the first experiment. The first panel shows estimated utility functions together; the next three panels show estimated weighting functions for the most commonly estimated shapes (68 of the 80 subjects). The median estimated λ^s is about 2.13.



likelihood ratio tests, about 85% to 95% of these estimated weighting functions significantly differ from identity weighting at the five percent level of significance, depending somewhat on the probabilistic model used.

Figures 6 and 8, however, are clearly unusual given widely held priors concerning weighting functions: Both the contextual utility and stronger utility estimations suggest a sampled population where the plurality type of decision maker is an optimist rather than a prospector, and where even rounders outnumber prospectors. Figure 7 is an exception to this, but not a very convincing one: Decision field theory estimations do produce prospectors as the plurality type, but inspection of the upper right panel (the prospector shapes) reveals that a large number of these estimated weighting functions might be optimists aside from the estimated weight at $q = 5/6$. Camerer and Ho (1994) and Wu and Gonzalez (1996) estimate that "the" weighting function crosses the identity line at some $q < 1/2$: Very few of the "prospector" shapes in the upper right panel of Figure 7 do this. These previous estimations used the homoscedastic latent index model which, as shown in Figure 5, tends to bias estimated weights downward—which could account for the discrepancy I point out here. The prospector shapes produced by decision field theory don't fit received Cumulative Prospect Theory priors.

6. First discussion

Consider two options $safe = (1 - p_m, p_m, 0)$ and $risky = (1 - q_h, 0, q_h)$ where $p_m > q_h$ and as usual $h > m > l$. Tversky (1969), Rubinstein (1988) and Leland (1994) have all noted that if $h - m$ is large but $p_m - q_h$ is sufficiently small, so that p_m and q_h are deemed "similar" but h and m are not, a decision procedure might not bother with computing and comparing overall values of $safe$ and $risky$, but instead simply ignore the similar probabilities and choose the option $risky$ with the noticeably larger "prize" h. Tversky showed that such decision procedures produce intransitive choices, and both Rubinstein and Leland showed that such decision procedures account for many of the Allais phenomena. This kind of decision procedure would reveal "apparent weights" ω such that $\omega(p_m) = \omega(q_h)$. If an experiment contains many option pairs of this kind, with many paired "similar" probabilities p_m and q_h in some interval $[a,b] \subset [0,1]$, a straightforward estimation of RDU or CPT will result in an estimated probability weighting function that is too flat on $[a,b]$—reflecting, to some extent, similarity-based computational shortcuts rather than the true difference $w(b) - w(a)$. The first experiment contains no option pairs like these, and does not often produce estimated weighting functions that are relatively flat on the range of low interior probabilities—a marker of prospector shapes.

However, the first experiment does contain option pairs which may lead to the commonly observed rounder shape. Consider the pair $safe = (0,1,0)$ and $risky = (1/6,0,5/6)$ on the context $\langle 40,50,60 \rangle$. A decision maker might sometimes regard $risky$ as (0,0,1) and choose it over $safe$. Likewise for a pair such as $safe = (0,1,0)$ and $risky = (5/6,0,1/6)$ on the context $\langle 50,70,110 \rangle$, a decision maker might sometimes regard $risky$ as (1,0,0) and choose $safe$ instead. I conjecture that this kind of decision maker produces the rounder shape. The coarse probability grid of the six-sided die wasn't coarse enough to make such behavior rare, given the frequency of estimated rounder shapes that is apparent in Figures 6, 7 and 8: or, at least, this is one interpretation of the rounder shape. These considerations suggest a second experiment that uses fourths as a very coarse probability grid: Perhaps these rounding shortcuts can be made rarer still with the help of a 4-sided die.

Andreoni and Sprenger (2012, p. 3373) have suggested that "Subjects exhibit a preference for certainty when it is available…" This could have an effect on estimated probability weighting. Because all relatively safe options in the first experiment are sure outcomes, this data is not well-suited to seeing whether this is an issue or not in the function-free estimations of probability weights. For now I observe that Cheung (2013) fails to replicate this finding when using a choice list method rather than the budget allocation method of Andreoni and Sprenger. In the second experiment, I also fail to replicate it using the choice pairs method.

7. Design of the second experiment

In this second experiment, I mainly seek a replication of the prevalence of estimated optimism. This experiment is done with a sampled population from a different university, with different option pairs and a different random device, the 4-sided die. As suggested in the previous section, the 4-sided die is an attempt to limit the prevalence of estimated rounder shapes. The second experiment also uses option pairs going beyond the sure things versus two-outcome risks of the first experiment. Let $safe = (p_l, p_m, p_h)$ and $risky = (q_l, q_m, q_h)$ denote vectors of outcome probabilities on the context $\langle l, m, h \rangle$. In all pairs, $p_m > q_m$ while $p_l < q_l$ and $p_h < q_h$. As before, subjects choose between $safe$ and $risky$ in each pair presented to them. Table 3 shows the 69 option pairs used in the experiment: Some are repeated up to four times as indicated in the "trials" column, for a total of 100 choice tasks in the experiment. There are ten distinct 3-outcome contexts, all created from the five positive money outcomes $15, $20, $30, $45 and $80. There is now plenty of variation in whether the option $safe$ is a sure thing (0,1,0) or not, which allows a check on concerns raised by Andreoni and Sprenger (2012).

Constraining all probabilities to the set of fourths (0, 1/4, 1/2, 3/4 or 1), the option pairs (and number of trials of each pair) were selected by way of iterated Monte Carlo simulation. The iterative procedure aimed at approximately maximizing the average determinant of the function-free estimator's information matrix for the worst 10% (lowest decile of information matrix determinants) of estimated parameters in a simulated population of decision makers whose distribution of DGPs resembled what had been previously estimated using past experimental data at Chapman University.

The subjects for the second experiment were 98 undergraduate students at Chapman University. Each subject participated in a single session, making choices from the choice tasks shown in Table 3. Sessions commenced with computerized instructions, including tests of understanding that returned subjects to relevant instruction sections in the event of test mistakes. Subjects had to correctly answer

Table 3. The 69 option pairs used in the second experiment.

pair #	trials	context $\langle l, m, h \rangle$	*safe* option outcome probabilities			*risky* option outcome probabilities		
			p_l	p_m	p_h	q_l	q_m	q_h
1	4	⟨15,20,30⟩	0	1	0	0.75	0	0.25
2	1		0	1	0	0.25	0.5	0.25
3	3		0	1	0	0.5	0	0.5
4	1		0	1	0	0.25	0	0.75
5	1		0.25	0.75	0	0.75	0	0.25
6	1		0.25	0.75	0	0.5	0	0.5
7	4		0	0.75	0.25	0.5	0	0.5
8	1		0.5	0.5	0	0.75	0	0.25
9	1		0.25	0.5	0.25	0.5	0	0.5
10	1		0	0.5	0.5	0.25	0	0.75
11	1	⟨15,20,45⟩	0	1	0	0.75	0	0.25
12	1		0	1	0	0.5	0	0.5
13	1		0.25	0.75	0	0.75	0	0.25
14	1		0	0.75	0.25	0.5	0	0.5
15	1		0	0.75	0.25	0.25	0	0.75
16	1		0	0.5	0.5	0.25	0	0.75
17	1	⟨15,20,80⟩	0	1	0	0.75	0	0.25
18	1		0	1	0	0.5	0	0.5
19	1		0.25	0.75	0	0.5	0	0.5
20	2		0.5	0.5	0	0.75	0	0.25
21	1	⟨15,30,45⟩	0	1	0	0.5	0	0.5
22	1		0	1	0	0.25	0	0.75
23	1		0.25	0.75	0	0.75	0	0.25
24	2		0	0.75	0.25	0.5	0	0.5
25	1		0	0.75	0.25	0.25	0	0.75
26	1		0.5	0.5	0	0.75	0	0.25
27	1		0	0.5	0.5	0.25	0	0.75
28	3	⟨15,30,80⟩	0.25	0.75	0	0.75	0	0.25
29	1		0.25	0.75	0	0.5	0	0.5
30	4		0	0.75	0.25	0.25	0	0.75
31	1		0.5	0.5	0	0.75	0	0.25
32	4		0.25	0.5	0.25	0.5	0	0.5
33	1		0	0.5	0.5	0.25	0	0.75
34	1	⟨15,45,80⟩	0	1	0	0.75	0	0.25
35	1		0	1	0	0.25	0	0.75
36	1		0.25	0.75	0	0.75	0	0.25
37	1		0.25	0.75	0	0.5	0	0.5
38	2		0	0.75	0.25	0.5	0	0.5
39	1		0	0.75	0.25	0.25	0	0.75
40	1		0.5	0.5	0	0.75	0	0.25
41	1		0	0.5	0.5	0.25	0	0.75

Table 3 (continued). The 69 option pairs used in the second experiment.

pair #	trials	context $\langle l, m, h\rangle$	*safe* option outcome probabilities			*risky* option outcome probabilities		
			p_l	p_m	p_h	q_l	q_m	q_h
42	2	⟨20,30,45⟩	0	1	0	0.75	0	0.25
43	1		0	1	0	0.5	0.25	0.25
44	1		0	1	0	0.25	0.5	0.25
45	4		0	1	0	0.5	0	0.5
46	1		0	1	0	0.25	0.25	0.5
47	2		0	1	0	0.25	0	0.75
48	1		0.25	0.75	0	0.75	0	0.25
49	1		0.25	0.75	0	0.5	0	0.5
50	1		0	0.75	0.25	0.5	0	0.5
51	1	⟨20,30,80⟩	0.25	0.75	0	0.75	0	0.25
52	1		0.25	0.75	0	0.5	0	0.5
53	1		0.5	0.5	0	0.75	0	0.25
54	1	⟨20,45,80⟩	0	1	0	0.25	0	0.75
55	1		0.25	0.75	0	0.5	0	0.5
56	1		0	0.75	0.25	0.5	0	0.5
57	1		0	0.5	0.5	0.25	0	0.75
58	4	⟨30,45,80⟩	0	1	0	0.75	0	0.25
59	1		0	1	0	0.5	0.25	0.25
60	1		0	1	0	0.25	0.5	0.25
61	3		0	1	0	0.5	0	0.5
62	1		0	1	0	0.25	0	0.75
63	1		0.25	0.75	0	0.75	0	0.25
64	3		0.25	0.75	0	0.5	0	0.5
65	1		0	0.75	0.25	0.5	0	0.5
66	1		0	0.75	0.25	0.25	0	0.75
67	1		0.5	0.5	0	0.75	0	0.25
68	1		0.25	0.5	0.25	0.5	0	0.5
69	1		0	0.5	0.5	0.25	0	0.75

all questions before proceeding. The 100 choice pairs were divided into two parts (a first part of 60 pairs and a second part of 40 pairs), separated by about ten to fifteen minutes of other tasks (again, a survey and tests of arithmetic and problem-solving ability). At the conclusion of a session, one of each subject's 100 choice pairs was selected at random (by means of the subject rolling two ten-sided dice) and the subject was paid according to their choice in that pair. If the subject's choice in the selected pair involved chance, the subject rolled a four-sided die (using a dice cup) to resolve payment. Sessions rarely lasted more than 70 minutes.

The second experiment involves five distinct outcomes but as before we can choose $u^s(15) = 0$ and $u^s(80) = 1$ for all subjects s. The unique estimable utility vector $\mathbf{u}^s$ for each subject s is the utilities of the three other outcomes $\mathbf{u}^s = \langle u^s(20), u^s(30), u^s(45)\rangle$, and function-free estimation makes those three utilities separate parameters to estimate. The experiment also involves three distinct probabilities $q \in \left\{\frac{1}{4}, \frac{2}{4}, \frac{3}{4}\right\}$, and so a vector $\mathbf{w}^s = \langle w^s\left(\frac{1}{4}\right), w^s\left(\frac{2}{4}\right), w^s\left(\frac{3}{4}\right)\rangle$ of three weights to be estimated for each subject. Function-free estimation makes those three weights separate parameters to estimate. Including the scale parameter λ^s, this makes seven total parameters for the function-free estimation. I use the same penalized maximum likelihood procedure for this estimation (see Appendix II).

8. Results of the second experiment

Figures 10, 11 and 12 show estimation results using the data from the second experiment. Figure 9 shows contextual utility estimations; Figure 10 shows decision field theory estimations; and Figure 11 shows stronger utility estimations. In each figure, the upper left panel shows 98 estimated utility functions while the remaining three panels show most (at least 84 of the 98) estimated weighting functions, divided into the three most commonly estimated shapes—optimists, rounders and pessimists or prospectors, generally in that order (except with contextual utility). Remaining subjects (estimated weighting functions not shown) include 11 or 13 prospectors and 1 or 2 unclassified.[4] Overall, by individual-level likelihood ratio tests, about 65% to 70% of the estimated weighting functions significantly differ from identity weighting at the five percent level of significance, depending somewhat on the probabilistic model used.

[4] 11 prospectors and 1 unclassified with contextual utility, 11 prospectors and 2 unclassified with stronger utility, and 13 prospectors and 1 unclassified with decision field theory.

Figure 9: 98 function-free individual estimates, estimated with the contextual utility model using data from the second experiment. The first panel shows estimated utility functions together; the next three panels show estimated weighting functions for the most commonly estimated shapes (86 of the 98 subjects). The median estimated λ^s is about 14.0.

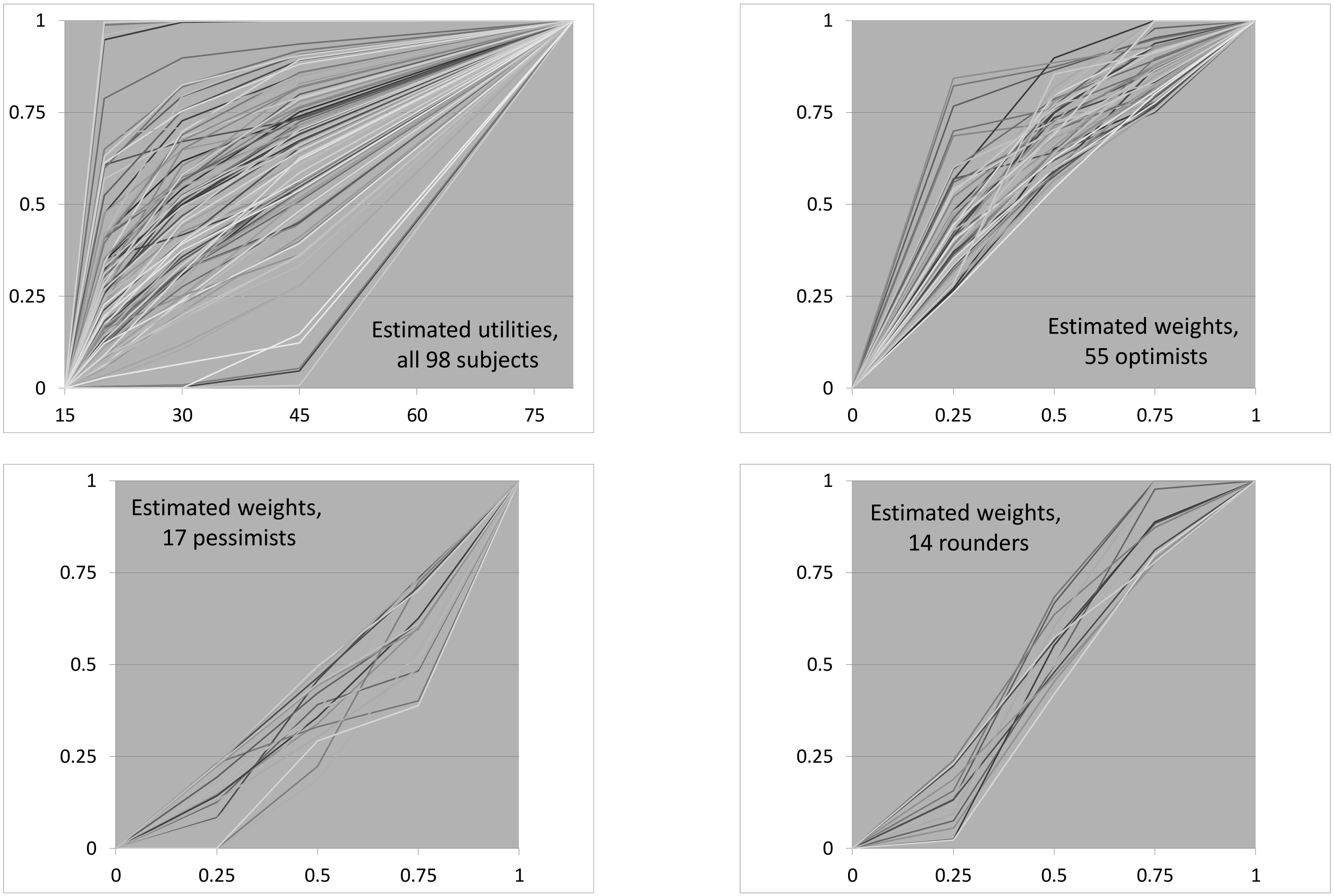


Figure 10: 98 function-free individual estimates, estimated with the decision field theory model using data from the second experiment. The first panel shows estimated utility functions together; the next three panels show estimated weighting functions for the most commonly estimated shapes (84 of the 98 subjects). The median estimated λ^s is about 5.65.

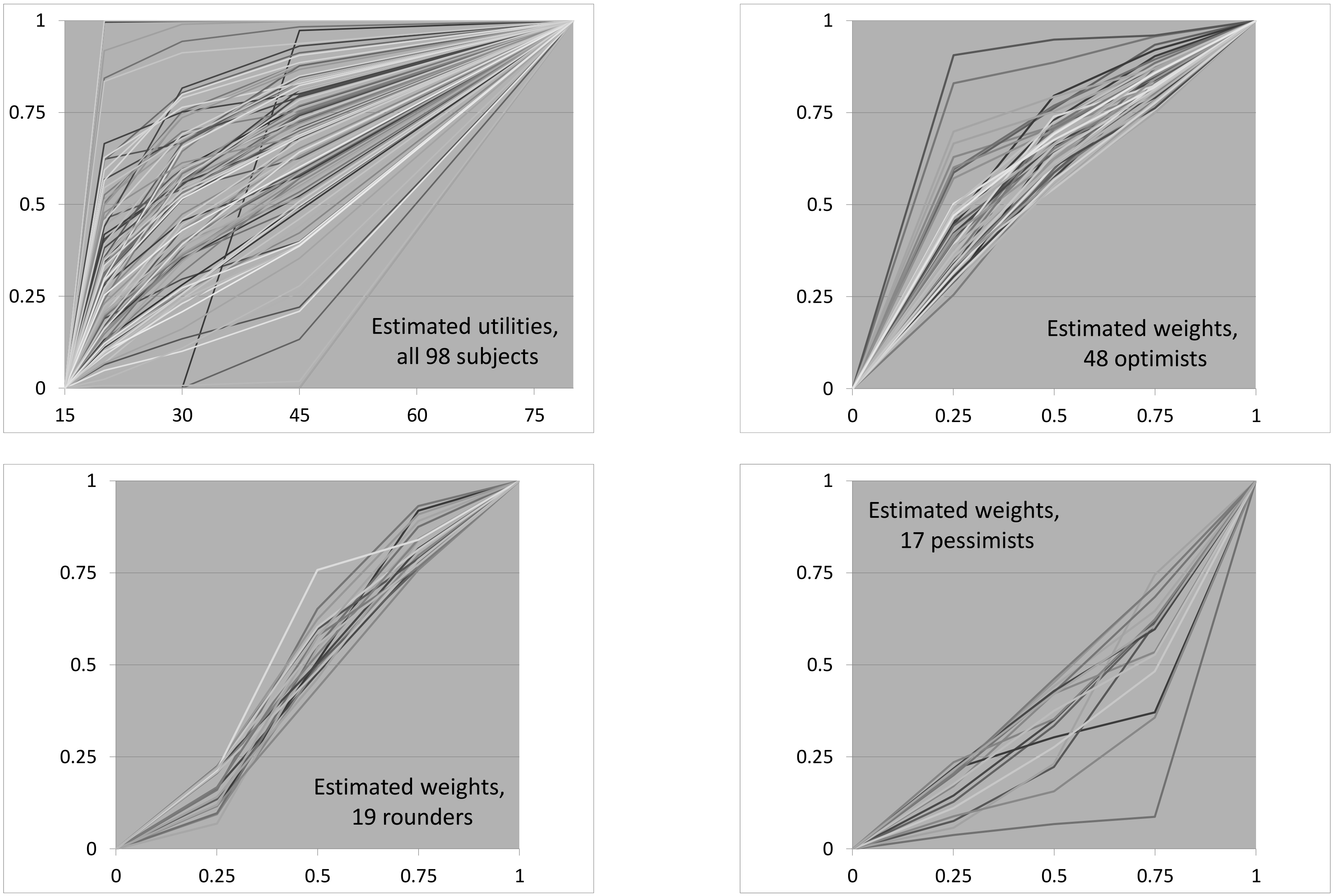
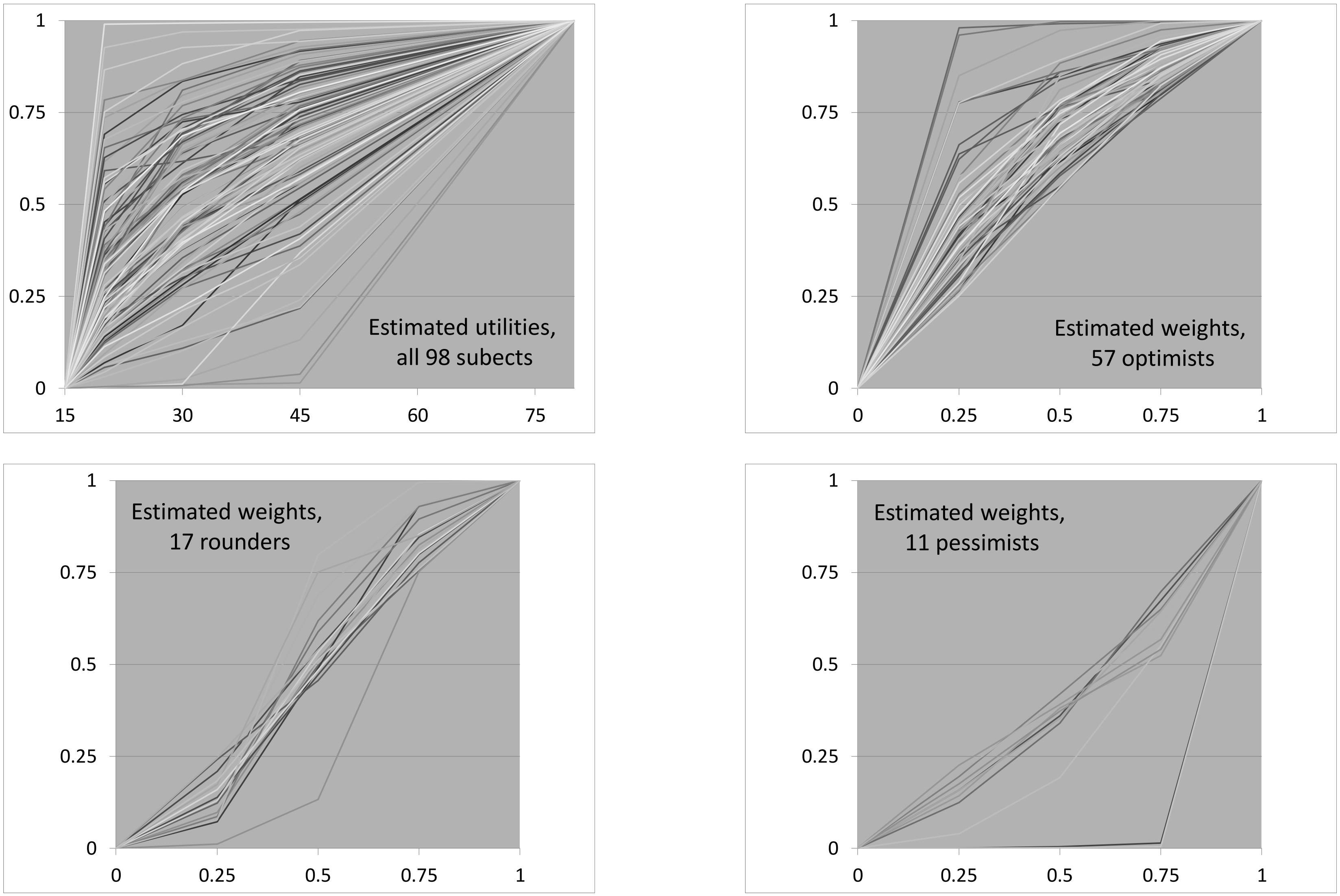


Figure 11: 98 function-free individual estimates, estimated with the stronger utility model using data from the second experiment. The first panel shows estimated utility functions together; the next three panels show estimated weighting functions for the most commonly estimated shapes (85 of the 98 subjects). The median estimated λ^s is about 2.31.



Estimated optimist shapes are an outright majority under contextual utility and stronger utility (and almost half of subjects under decision field theory) and more than twice as common as the second-most-common shape (rounders except with contextual utility, where it is pessimists). Keeping in mind that the set of outcomes, the probability device and the sampled population are all different in this second experiment, so the (tempting) conclusion that the coarser probability grid has resulted in less rounding and more optimism (relative to the first experiment) is not formally warranted. However, optimist shapes are again the most commonly observed shape. This has been replicated with a new sample from a different population, a new die and a new outcome set.

One can modify estimations to include the possibility raised by Andreoni and Sprenger (2012)—that subjects "exhibit a preference for certainty when it is available." To do this, I multiply all $safe$ option occurrences of $u^s(m_i)$ in equations 8, 9 and 10 by a factor $[1 + \beta^s 1(p_{mi} = 1)]$, shifting the estimated utility of the middle outcome m_i in $safe$ by a multiplicative effect β^s whenever $safe$ is a sure thing. Having done this and estimated these models, I find no evidence that β^s is systematically and significantly positive as suggested by the findings of Andreoni and Sprenger. Additionally, this does not change the qualitative findings in Figures 9 and 10 much: Optimist shapes are nearly half of the estimates using either decision field theory or contextual utility as the probabilistic model.[5]

9. Second discussion

Optimism is again prevalent in the second experiment—even more prevalent than in the first experiment. I have suggested that the received prospector shape, characterized by relatively flat weighting functions on interior probability ranges, may frequently occur because close probabilities tend to be regarded as similar and ignored. Here is one econometric path for addressing this possibility. Begin with a design resembling that of the second experiment: It identifies utilities $\mathbf{u}^s = \langle u^s(20), u^s(30), u^s(45) \rangle$ and weights $\mathbf{w}^s =$

[5] With estimates $\hat{\beta}^s$ in hand for the 98 subjects, the null hypothesis $\beta^s = 0$ fails to be rejected at the 10% significance level by either a sign, signed-rank or t-test when contextual utility is the probabilistic model, and 48 of the 98 estimated weighting functions have optimist shapes. When decision field theory is the probabilistic model, estimates of β^s have a weakly significantly <u>negative</u> location (p = 0.081) by a signed-rank test, but not by the sign or t-test, and 51 of the 98 estimated weighting functions have optimist shapes. Application of stronger utility to this case is not straightforward since it is unclear how stochastic dominance is to be defined when there is one utility function for certain outcomes and another for uncertain outcomes.

$\langle w^s\left(\frac{1}{4}\right), w^s\left(\frac{2}{4}\right), w^s\left(\frac{3}{4}\right)\rangle$. Now replace the 4-sided die with a 12-sided die: We can still use the same design to identify the same weights and utilities, but suppose we wish to add some choice pairs to identify $w^s\left(\frac{1}{3}\right)$ as well and, in particular, the marginal weight between $q = 1/4$ and $q = 1/3$. Let $dw^s = w^s\left(\frac{1}{3}\right) - w^s\left(\frac{1}{4}\right)$ denoted this marginal weight.

Two routes to this identification can be imagined. The first route depends on adding pairs such as $safe = (0,1,0)$ and $risky = (2/3,0,1/3)$. In pairs like this, only the most aggressive rounder would view the 1/3 probability (of h in $risky$) as zero, and almost no one would view the 1/3 probability (of h in $risky$) as similar to certainty (of m in $safe$). I call this a “dissimilar pair” for those reasons: It does not encourage computational shortcuts based on either similarity judgments or rounding behavior. Add enough pairs like this one to the pre-existing design and we should be able to estimate $w^s\left(\frac{1}{3}\right)$ directly and then estimate the marginal weight dw^s as the difference between the estimates $\hat{w}^s\left(\frac{1}{3}\right)$ and $\hat{w}^s\left(\frac{1}{4}\right)$. Call this estimate $\widehat{dw}^s_{dis}$ (the subscript dis meaning “dissimilar”).

The second route to identification depends on adding a different sort of choice pair such as $safe = (2/3,1/3,0)$ and $risky = (3/4,0,1/4)$. Add enough pairs like this one to the pre-existing design and we should also be able to estimate $w^s\left(\frac{1}{3}\right)$ and hence dw^s. But this is not a “dissimilar pair” as defined in the previous paragraph: I believe that many decision makers would regard the 1/3 probability (of m in $safe$) as similar to the 1/4 probability (of h in $risky$), and would therefore ignore that probability difference and choose according to most-preferred outcome (that is, choose $risky$ since $h > m$). For that reason, I will call these “similar pairs.” Although we can estimate dw^s by adding only such similar pairs, I expect that our resulting estimate—call this $\widehat{dw}^s_{sim}$—will be much smaller than we would estimate by adding only dissimilar pairs to the pre-existing design (that is, following the first identification strategy).

Under the hypothesis that rank-dependent weighting exists independently of similarity, the two identification strategies outlined above should result in equivalent estimates of dw^s. The final observation is that nothing prevents us from constructing a design which simultaneously follows both paths to identifying dw^s—that is, in which dw^s

is overidentified, once with similar pairs and once with dissimilar pairs. The third experiment does this.

10. Design of the third experiment

The option pairs in this third experiment begin with design considerations and choices very like those of the second experiment. As before, subjects choose between $safe$ and $risky$ in each pair presented to them. There are ten distinct 3-outcome contexts, all created from the five positive money outcomes $15, $20, $30, $45 and $80. Table 4-A shows 34 of the option pairs used in the experiment: Some of these are repeated up to four times as indicated in the "trials" column, for a total of 68 choice tasks. These choice tasks are the "trunk" of the design: The probabilities in this set of pairs are constrained to the set of fourths (0, 1/4, 1/2, 3/4 or 1).

The pairs in Tables 5-B and 5-C are the two different identification "branches" of the design: These pairs introduce options that contain the 1/3 probability of a highest outcome in various options. There are six dissimilar pairs in Table 4-B, and six similar pairs in Table 4-C, each repeated up to four times as indicated in the "trials" column, for a total of 16 choice tasks from each of these tables. With the 68 choice tasks from Table 4-A, this is a total of 100 choice tasks in the design. As with the design of the second experiment, the 68 choice tasks in Table 4-A were chosen by an iterated Monte Carlo simulation procedure aimed at maximizing the efficiency of estimation for the worst decile of the sampled population. Then, the same kind of iterated Monte Carlo procedure was used to select contexts and numbers of trials for the two branches aimed at efficient estimation of dw^s in both branches.

The subjects for the third experiment were 92 undergraduate students, again at Chapman University as with the second experiment. The experimental protocol was almost identical to that of the second experiment, except that a twelve-sided die was used as the random device—this being the lowest-sided die capable of producing both fourths and thirds as probabilities.

Estimation closely resembles that undertaken for data from the second experiment. The third experiment now involves four distinct probabilities $q \in \left\{\frac{1}{4}, \frac{1}{3}, \frac{2}{4}, \frac{3}{4}\right\}$, and hence a

Tables 5. Option pairs used in the third experiment.

Table 4-A. 34 pairs used for both dissimilar and similar estimations (68 total trials).

pair #	trials	context $\langle l, m, h \rangle$	*safe* option outcome probabilities			*risky* option outcome probabilities		
			p_l	p_m	p_h	q_l	q_m	q_h
1	4	⟨15,20,30⟩	0	1	0	0.75	0	0.25
2	3		0	1	0	0.5	0	0.5
3	1		0	1	0	0.25	0	0.75
4	4		0	0.75	0.25	0.5	0	0.5
5	1		0.25	0.5	0.25	0.5	0	0.5
6	1		0	0.5	0.5	0.25	0	0.75
7	1	⟨15,20,45⟩	0	0.5	0.5	0.25	0	0.75
8	1	⟨15,20,80⟩	0.25	0.75	0	0.75	0	0.25
9	1		0.25	0.75	0	0.5	0	0.5
10	1	⟨15,30,45⟩	0.25	0.75	0	0.75	0	0.25
11	1		0	0.75	0.25	0.5	0	0.5
12	1		0	0.5	0.5	0.25	0	0.75
13	4	⟨15,30,80⟩	0.25	0.75	0	0.75	0	0.25
14	1		0.5	0.5	0	0.75	0	0.25
15	4		0.25	0.5	0.25	0.5	0	0.5
16	4		0	0.75	0.25	0.25	0	0.75
17	1	⟨15,45,80⟩	0.25	0.75	0	0.75	0	0.25
18	1		0.5	0.5	0	0.75	0	0.25
19	1		0	1	0	0.25	0	0.75
20	2		0	0.75	0.25	0.5	0	0.5
21	2		0	0.5	0.5	0.25	0	0.75
22	3	⟨20,30,45⟩	0	1	0	0.75	0	0.25
23	1		0	1	0	0.25	0.5	0.25
24	4		0	1	0	0.5	0	0.5
25	2		0.25	0.75	0	0.5	0	0.5
26	3		0	1	0	0.25	0	0.75
27	1		0	0.75	0.25	0.5	0	0.5
28	1	⟨20,45,80⟩	0	0.5	0.5	0.25	0	0.75
29	3	⟨30,45,80⟩	0	1	0	0.75	0	0.25
30	3		0	1	0	0.5	0	0.5
31	3		0.25	0.75	0	0.5	0	0.5
32	1		0	1	0	0.25	0	0.75
33	2		0	0.75	0.25	0.5	0	0.5
34	1		0	0.75	0.25	0.25	0	0.75

Tables 5 (continued). Option pairs used in the third experiment.

Table 4-B. 6 “dissimilar pairs” used only for dissimilar estimations (16 trials in all).

pair #	trials	context $\langle l,m,h\rangle$	*safe* option outcome probabilities			*risky* option outcome probabilities		
			p_l	p_m	p_h	q_l	q_m	q_h
35	3	⟨15,20,30⟩	0	1	0	0.67	0	0.33
36	4	⟨15,20,80⟩	0	1	0	0.67	0	0.33
37	1	⟨15,45,80⟩	0	1	0	0.67	0	0.33
38	3	⟨20,30,45⟩	0	1	0	0.67	0	0.33
39	1	⟨20,30,80⟩	0	1	0	0.67	0	0.33
40	4	⟨30,45,80⟩	0	1	0	0.67	0	0.33

Table 4-C. 6 “similar pairs” used only for similar estimations (16 trials in all).

pair #	trials	context $\langle l,m,h\rangle$	*safe* option outcome probabilities			*risky* option outcome probabilities		
			p_l	p_m	p_h	q_l	q_m	q_h
41	2	⟨15,20,30⟩	0.67	0.33	0	0.75	0	0.25
42	3	⟨15,20,80⟩	0.67	0.33	0	0.75	0	0.25
43	4	⟨15,30,45⟩	0.67	0.33	0	0.75	0	0.25
44	4	⟨15,45,80⟩	0.67	0.33	0	0.75	0	0.25
45	2	⟨20,30,45⟩	0.67	0.33	0	0.75	0	0.25
46	1	⟨30,45,80⟩	0.67	0.33	0	0.75	0	0.25

vector of four weights to estimate. As suggested by the second discussion in the previous section, we can think of the two different branches of the design as creating two possibly different vectors of weights $\mathbf{w}^s_{dis}$ and $\mathbf{w}^s_{sim}$. The dissimilar estimation uses only the 84 choice tasks of tables 4-A and 4-B to produce an estimate $\hat{\mathbf{w}}^s_{dis}$, while the similar estimation uses only the 84 choice tasks of Tables 4-A and 4-C to produce an estimate $\hat{\mathbf{w}}^s_{sim}$. The same penalized maximum likelihood procedure was used for this estimation (see Appendix II). With these estimates in hand, two different estimates of the marginal weight may be computed as $\widehat{dw}^s_{dis} = \hat{w}^s_{dis}\left(\frac{1}{3}\right) - \hat{w}^s_{dis}\left(\frac{1}{4}\right)$ and $\widehat{dw}^s_{sim} = \hat{w}^s_{sim}\left(\frac{1}{3}\right) - \hat{w}^s_{sim}\left(\frac{1}{4}\right)$.

11. Results of experiment three

Figure 12 shows the results of the dissimilar estimation (the left panel) and the similar estimation (the right panel) side by side, using contextual utility as the probabilistic model.

Vertical lines at $q = 1/4$ and $q = 1/3$ focus attention on the change in estimated weights across this probability interval. The dissimilar estimations result in a handful of flat weighting function segments, that is $\widehat{dw}^s_{dis} = 0.0001$,[6] across the interval—11 out of 92 subjects in fact. The similar estimation shows well more than a handful of flat weighting function segments: In fact 57 of the 92 estimates result in $\widehat{dw}^s_{sim} = 0.0001$. This alone is strong evidence that the similar pairs quite commonly provoke the computational shortcut suggested throughout this study. Decision field theory produces the same kind of figures. Stronger utility, on the other hand, only rarely produced bottom-bounded estimates.

Table 5 shows the sample mean value of $\widehat{dw}^s_{dis} - \widehat{dw}^s_{sim}$, which I will call "the similarity effect," along with related statistics. In absolute terms as well as the estimated effect size, contextual utility estimations produce the strongest similarity effect: Across the 92 subjects, the sample mean of $\widehat{dw}^s_{dis} - \widehat{dw}^s_{sim}$ is 0.0791 with a standard error[7] of 0.011 (a p-value would be gratuitous). Perspective on the size of this estimated effect is provided by identity weights (expected utility for instance) for which $dw^s = 1/3 – 1/4 = 0.0833$. That

Table 5. The estimated similarity effect $\widehat{dw}^s_{dis} - \widehat{dw}^s_{sim}$ in the third experiment.

	probabilistic model used for estimation		
	contextual utility	decision field theory	stronger utility
sample mean	0.0791	0.0627	0.0462
standard error	0.011	0.0104	0.0092
standard deviation of $\widehat{dw}^s_{dis}$	0.0855	0.0868	0.0841
effect size	0.925	0.723	0.549

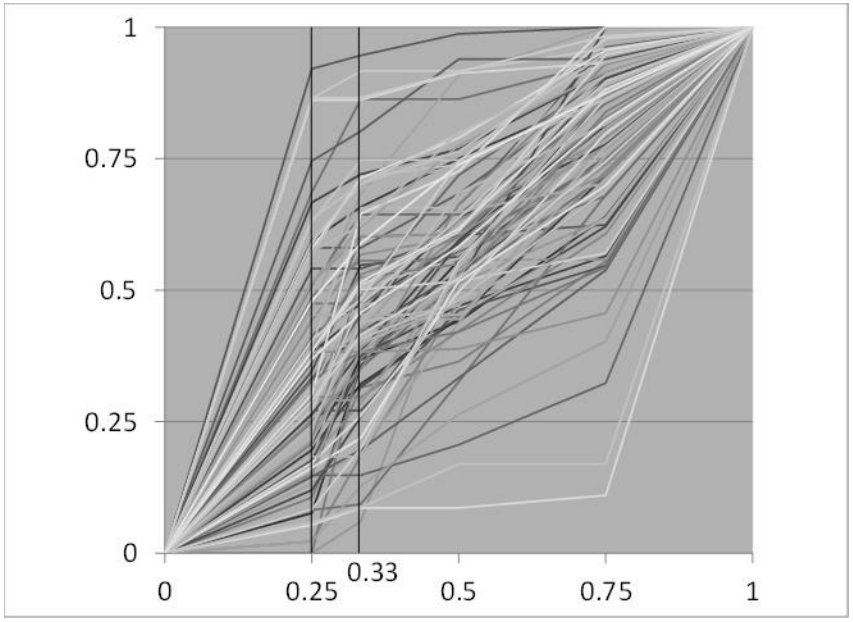
Notes: The effect size is calculated as the ratio of the sample mean of $\widehat{dw}^s_{dis} - \widehat{dw}^s_{sim}$ to the standard deviation of $\widehat{dw}^s_{dis}$. An effect size of 0.5 is considered moderate while an effect size of 0.8 is considered large. (The standard deviation of $\widehat{dw}^s_{sim}$ is always smaller than that of $\widehat{dw}^s_{dis}$, so the effect sizes would be larger if that information was used too.)

6 As mentioned in Appendix II, monotonicity is imposed on estimated utilities and weights in such a manner that the minimum estimated value of dw^s is constrained to be no smaller than 0.0001.

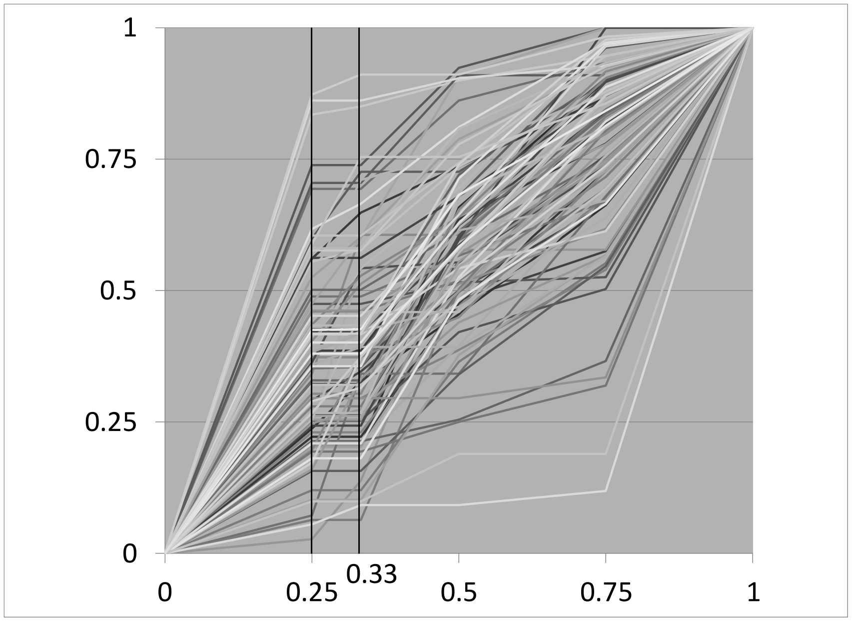
7 I treat each calculated value of $\widehat{dw}^s_{dis} - \widehat{dw}^s_{sim}$, for each subject, as an independent single observation (with one degree of freedom), and do simple statistics based on that. Any scalar quantity $y^s = f(\mathbf{r}^s) - g(\mathbf{r}^s)$, where $\mathbf{r}^s$ is an observation vector from subject s, can be represented as $y^s = E_s[f(\mathbf{r}^s) - g(\mathbf{r}^s)] + \varepsilon^s$, where $E_s[f(\mathbf{r}^s) - g(\mathbf{r}^s)]$ is the population mean of $f(\mathbf{r}^s) - g(\mathbf{r}^s)$ and the ε^s are independent across subjects. Therefore, where inferences about $E_s[f(\mathbf{r}^s) - g(\mathbf{r}^s)]$ are concerned, the very simplest statistics can be applied.

Figure 12. 92 function-free individual estimates of weighting functions, estimated with the contextual utility model using data from the third experiment. The left panel shows estimations using only the "dissimilar pairs" to identify the weight at $q = 0.33$; the right panel shows estimations using only the "similar pairs" to identify the weight at $q = 0.33$.

Estimated weights using only the pairs in Tables 4-A and 4-B (the "dissimilar pairs"). 11 of 92 estimated weighting functions are flat on the interval [0.25,0.33] in this case.



Estimated weights using only the pairs in Tables 4-A and 4-C (the "similar pairs"). 57 of 92 estimated weighting functions are flat on the interval [0.25,0.33] in this case.



is, the estimated size of the similarity effect would very nearly erase identity weighting. The sample mean of the estimated similarity effect is smaller when I perform the estimation with either decision field theory or stronger utility, but still significantly positive at any conventional significance level. This casts strong doubt on the null hypothesis that estimated probability weights are independent of similarity effects.

12. Conclusions

Optimism is the most prevalent form of rank-dependent weighting functions estimated here—not the widely believed inverse-s shape. I attribute this to several potential factors. The designs of the first and second experiments deliberately set about to minimize opportunities for reducing decision complexity by way of computational shortcuts based on similarity of probabilities and rounding of probabilities. This was done by confining probabilities to relatively coarse grids and avoiding choice pairs that juxtapose similar probabilities of high outcomes. The third experiment showed that such pairs do produce flatter probability weighting function estimates on low to moderate interior probabilities—a defining feature of the inverse-s prospector shape—in the predicted manner.

If $\underline{p}$ and $\overline{p}$ are the smallest and largest probabilities of the maximum lottery outcomes found across all lotteries in some binary choice experiment, it can always be the case that, for probabilities of a maximum outcome below $\underline{p}$, or above $\overline{p}$, that experiment will miss some bit of curvature or elevation different from that observed on $[\underline{p},\overline{p}]$ or miss a fixed point not found on $[\underline{p},\overline{p}]$. Of course that possibility is true of my three experiments. But if similarity-induced shortcuts and rounding short-circuit "normal" probability weighting of outcome utilities, pushing designed outcome probabilities ever closer to zero or one increasingly runs the risk that those probabilities get treated as zero and one, respectively. Indeed Kahneman and Tversky (1979) originally thought the probability weighting function was poorly behaved—maybe not a continuous function at all, or even a function—near its endpoints for those kinds of reasons.

Study of risk preferences requires a researcher to make several interrelated choices. Here I chose binary discrete choice as an elicitation method. This has its virtues, not least of which is the fact that binary preference relations are the primitive of most axiomatic theories. Yet each elicitation method comes with its own econometric conundrums—for instance, where and how functional form assumptions should be deployed. Here, I chose to minimize functional assumptions concerning the utilities and weights that are the structural entities of axiomatic rank-dependent representation theorems. This has costs. Assumptions concerning probabilistic models of binary discrete choices will be needed, and I have used three such models. By and large my results are not too sensitive to a choice of one of those models or another. Another approach might minimize assumptions about probabilistic models—say exploiting new semiparametric methods for discrete choice estimation. This has its own cost: A more parametric approach to the decision-theoretic entities. This would be good and useful future work.

Finally, others choose to elicit certainty equivalents, usually employing a choice list procedure to approximate certainty equivalents (e.g. Tversky and Kahneman 1992; Gonzalez and Wu 1999; Bruhin, Fehr-Duda and Epper 2010) and then use econometric methods appropriate to certainty equivalents. Such estimations generally find the conventional inverse-s shape. However, Wilcox (2017) describes a fundamental ambiguity concerning certainty equivalent evidence having to do with different sources of the probabilistic nature of decision and judgment. Elicitation of certainty equivalents just doesn't free one from the necessary business of making probabilistic modeling assumptions.

Throughout this chapter I have motivated discussions in terms of "bias" in probability weighting estimation due to similarity-induced computational shortcuts or rounding behavior that short-circuit "normal" rank-dependent weighting of utilities. I argued certain kinds of experimental designs would more likely produce that kind of "bias." Of course "poor" experimental design can result in all kinds of biased estimation. From that perspective, readers may think that the chapter simply points to issues that a careful experimenter can consider to detect and avoid bad designs. But there may be no gold standard experimental design that allows estimation of that one stable probability weighting function. So now I need to come clean about that previous "bias" rhetoric. What I fear (but do not know) is that with considerations of similarity, rounding, and other things in mind, an experimenter might—by judicious selection of choice pairs—be able to demonstrate almost any rank-dependent probability weighting function shape. Asked what the probability weighting function looks like, the reply of a worldly experimenter might resemble that of the famously broad-minded corporate accountant: What do you want it to look like?

References

Abdellaoui, M., 2000, Parameter-free elicitation of utility and probability weighting functions. Management Science 46, 1497-1512.

Abdellaoui, M., H. Bleichrodt and C. Paraschiv, 2007, Loss aversion under prospect theory: A parameter-free measurement. Management Science 53, 1659-1674.

Anderson, S., G. W. Harrison, M. I. Lau and E. E. Rutstrom, 2008, Eliciting risk and time preferences. Econometrica 76, 583-618.

Anderson, S., G. W. Harrison, M. I. Lau and E. E. Rutstrom, 2013, Discounting behavior and the magnitude effect: Evidence from a field experiment in Denmark. Economica 80, 670-697.

Andreoni, J. and C. Sprenger, 2012, Risk preferences are not time preferences. American Economic Review 102, 3357-3376.

Apesteguia, J. and M. Ballester, 2018, Monotone stochastic choice models: The case of risk and time preferences. Journal of Political Economy 126:74-106.

Ballinger, T. P., and N. Wilcox, 1997, Decisions, error and heterogeneity. Economic Journal 107, 1090-1105.

Blavatskyy, P. R., 2011, Probabilistic risk aversion with an arbitrary outcome set. Economics Letters 112:34-37.

Blavatskyy, P. R., 2013, Which decision theory? Economics Letters 120, 40-44.

Blavatskyy, P. R., 2014, Stronger utility. Theory and Decision 76, 265-286.

Brown, A., and P. J. Healy, 2018, Separated decisions. European Economic Review 101:20–34.

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

Busemeyer, J. and J. Townsend, 1992, Fundamental derivations from decision field theory. Mathematical Social Sciences 23, 255-282.

Busemeyer, J. and J. Townsend, 1993, Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. Psychological Review 100, 432-59.

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

Camerer, C. and T.-H. Ho, 1994, Violations of the betweeness axiom and nonlinearity in probability. Journal of Risk and Uncertainty 8, 167-196.

Cheung, S., 2013, Comment on 'Risk preferences are not time preferences': On the elicitation of time preference under conditions of risk. American Economic Review 105, 2242-2260.

Conlisk, J., 1989, Three variants on the Allais example. American Economic Review 79, 392-407.

Cox, J. C., V. Sadiraj, and U. Schmidt, 2015, Paradoxes and mechanisms for choice under risk. Experimental Economics 18:215–250.

Fishburn, P., 1978, A probabilistic expected utility theory of risky binary choices. International Economic Review 19, 633-646.

Gonzalez, R. and G. Wu, 1999, On the shape of the probability weighting function. Cognitive Psychology 38, 129-166.

Harrison, G. W. and J. T. Swarthout, 2014, Experimental payment protocols and the bipolar behaviorist. Theory and Decision 77, 423-438.

Hey, J. D., 2001, Does repetition improve consistency? Experimental Economics 4, 5-54.

Hey, J. D. and C. Orme, 1994, Investigating generalizations of expected utility theory using experimental data. Econometrica 62, 1291-1329.

Hey, J. D., G. Lotito, and A. Maffioletti, 2010, The descriptive and predictive adequacy of theories of decision making under uncertainty/ambiguity. Journal of Risk and Uncertainty 41:81-111.

Kahneman, D. and A. Tversky, 1979, Prospect theory: An analysis of decision under risk. Econometrica 47, 263-291.

Langer, E. J., 1982, The illusion of control. In D. Kahneman, P. Slovic and A. Tversky, eds., Judgment Under Uncertainty: Heuristics and Biases. New York: Cambridge University Press.

Leland, Jonathan W., 1994, Generalized similarity judgments: An alternative explanation for choice anomalies. Journal of Risk and Uncertainty 9, 151–172.

Loomes, G. and R. Sugden, 1995, Incorporating a stochastic element into decision theories. European Economic Review 39, 641-648.

Loomes, G. and R. Sugden, 1998, Testing different stochastic specifications of risky choice. Economica 65, 581-598.

Luce, R. D., 1959, Individual Choice Behavior. New York: Wiley.

Mosteller, F. and P. Nogee, 1951, An experimental measurement of utility. Journal of Political Economy 59, 371-404.

Nilsson, H., J. Rieskamp and E-J. Wagenmakers, 2011, Hierarchical Bayesian parameter estimation for cumulative prospect theory. Journal of Mathematical Psychology 55:84-93.

Pratt, J. W., 1964, Risk aversion in the small and in the large. Econometrica 32, 122-136.

Prelec, D., 1998, The probability weighting function. Econometrica 66, 497-527.

Quiggin, J, 1982, A theory of anticipated utility. Journal of Economic Behavior and Organization 3, 323-343.

Quiggin, J, 1993. Generalized Expected Utility Theory: The Rank-Dependent Model. Norwell, MS: Kluwer.

Rieskamp, J., 2008, The probabilistic nature of preferential choice. Journal of Experimental Psychology: Learning, Memory and Cognition 34, 1446-1465.

Rubinstein, A., 1988, Similarity and decision making under risk (Is there a utility theory resolution to the Allais paradox?). Journal of Economic Theory 46, 145-153.

Scheibehenne, B. and T. Pachur, 2015, Using Bayesian hierarchical parameter estimation to assess the generalizability of cognitive models of choice. Psychonomic Bulletin and Review 22:391–407.

Starmer, C. and R. Sugden, 1989, Probability and juxtaposition effects: An experimental investigation of the common ratio effect. Journal of Risk and Uncertainty 2, 159-78.

Starmer, C. and R. Sugden, 1991, Does the random-lottery incentive system elicit true preferences? An experimental investigation. American Economic Review 81, 971-978.

Tversky, A., 1969, Intransitivity of preferences. Psychological Review 76, 31-48.

Tversky, A. and D. Kahneman, 1992, Advances in prospect theory: Cumulative representation of uncertainty. Journal of Risk and Uncertainty 5, 297–323.

Wald, A., 1947. Sequential Analysis. New York: Wiley.

Wilcox, N., 1993, Lottery choice: Incentives, complexity and decision time. Economic Journal 103, 1397-1417.

Wilcox, N., 2008, Stochastic models for binary discrete choice under risk: A critical primer and econometric comparison. In J. C. Cox and G. W. Harrison, eds., Research in Experimental Economics Vol. 12: Risk Aversion in Experiments pp. 197-292. Bingley, UK: Emerald.

Wilcox, N., 2011, 'Stochastically more risk averse:' A contextual theory of stochastic discrete choice under risk. Journal of Econometrics 162, 89-104.

Wilcox, N., 2015, Error and generalization in discrete choice under risk. Working paper, Economic Science Institute, Chapman University.

Wilcox, N., 2017, Random expected utility and certainty equivalents: Mimicry of probability weighting functions. Journal of the Economic Science Association 3:161-173.

Wu, G. and R. Gonzalez, 1996, Curvature of the probability weighting function. Management Science 42, 1676-1690.

Appendix I: Background on the three probabilistic choice models

The contextual utility or CU model (Wilcox 2011) makes comparative risk aversion properties of the RDU representation and its stochastic implications consistent within and across contexts. For representations such as RDU, utility functions $u(z)$ are only unique up to a ratio of differences: Intuitively, contextual utility exploits this uniqueness to create a correspondence between functional and probabilistic definitions of comparative risk aversion. Consider the choice pairs in the first experiment: Under RDU and contextual utility, eq. 4 can be rewritten as

(A1) $P^{rdcu} = F(\lambda[-v(l,m,h) + w(q_h)])$, where

$$v(l,m,h) = [u(m) - u(l)]/[u(h) - u(l)].$$

This probability is decreasing in the ratio of differences $v(l,m,h)$. Consider two subjects Anne and Bob with identical weighting functions (this includes the case where both have EU preferences) and identical scale parameters λ, and assume that Bob is globally more risk averse than Anne in Pratt's sense (Bob's local absolute risk aversion $-u''(z)/u'(z)$ exceeds that of Anne for all z). These assumptions, Pratt's (1964) main theorem, and simple algebra shows that $v^{Bob}(l,m,h) > v^{Anne}(l,m,h)$ on all contexts. As a result (A1) implies that Bob will have a lower probability than Anne of choosing $risky$ on all contexts in the first experiment. Wilcox (2011) shows that the received homoscedastic latent index model cannot share this property, and this was my primary motivation for the contextual utility model. In the second and third experiments, the property is a somewhat weaker one appropriate to RDU when one or both options in a pair have nonzero probabilities of all three outcomes (see Wilcox 2011, p. 97, Proposition 2), reflecting the role probability weights play in observed risk aversion.

This volume focusses on Cumulative Prospect Theory (or CPT). Wilcox's (2011) results concerning the contextual normalization's good theoretical properties flow from RDU and EUT because the utility function $u(z)$ in those theories is an interval scale. Since CPT's value function is a ratio scale (not an interval scale), there's no reason to think the contextual normalization is appropriate for CPT.

Here I give a conjecture on an appropriate normalization of ΔCPT, the overall difference between option values according to CPT. For the purpose of this discussion I adopt one bit of notation Harrison and Swarthout [this volume, eqs. 16a and 16b] use to define CPT: Specifically $U(m)$ is CPT's value function. First consider what the contextual normalization does for interval scales. Pratt (1964) shows in his main theorem that suitably specified ratios of differences create unique quantities that correspond to risk attitudes in a one-to-one mapping. This is equivalent to the observation that, for interval scales, ratios of scale differences are unique. Formally the contextual normalization simply exploits that part of Pratt's main theorem.

Similarly for ratio scales, ratios of scale values (rather than scale value differences) are unique. That suggests that for pure gain pairs, a suitable normalization of ΔCPT might be $\Delta CPT/U(\overline{m})$ where $\overline{m}$ is the maximum outcome in such a pair. In pure loss pairs, a suitable normalization might be $\Delta CPT/U(\underline{m})$, where $\underline{m}$ is the minimum outcome in such pairs. For mixed pairs, I conjecture that a suitable normalization would be $\Delta CPT/[U(\overline{m}) - U(\underline{m})]$, resembling the contextual normalization. See Appendix C of Harrison and Swarthout [this volume] for a sensitivity analysis using this normalization.

Note that eq. 5 is the <u>decision field theory</u> model or DFT <u>only</u> for pairs like those found in the first experiment. For those choice pairs, DFT shares CU's main property: Holding constant λ and $w(q_h)$, globally greater risk aversion (in the sense of Pratt) will imply a lower probability of choosing $risky$ in all pairs on all contexts. The general formulation of $D(risky, safe)$ in DFT, which is needed for the estimations using data from the second and third experiments, depends on the underlying events that generate outcome probabilities as well as outcome utilities. Index events by $e = 1,2, \dots, E$, let w_e be the decision weight given to event e, and let u_e^{risky} and u_e^{safe} be the utilities resulting from the choice of options $risky$ and $safe$, respectively, when event e occurs. Then the general formulation of $D(risky, safe)$ in decision field theory is:

(A2) $D(risky, safe) = \sqrt{\sum_e w_e \left(u_e^{risky} - u_e^{safe} - \Delta RDU\right)^2}$.

Busemeyer and Townsend (1992, 1993) derive DFT from a computational argument: The theory is one of the early "diffusion" models of probabilistic choice. A simple intuition can be given for the model. Suppose that a decision maker's computational resources can effortlessly and quickly provide utilities of outcomes, and also suppose the decision maker wishes to choose according to relative RDU value; but suppose she does not have an algorithm for effortlessly and quickly multiplying utilities and weights together. The decision maker could proceed by sampling events in option pairs in proportion to their decision weights, keeping running sums of the sampled utility differences between the options, and choose when the summed differences exceed some threshold determined by the cost of sampling. In essence, the choice probability in eq. 5 results from this kind of sequential sampling decision procedure, which can be traced back to Wald (1947). Busemeyer and Townsend also show that, as the sampling rate gets large, the function F will be the logistic c.d.f.—the reason I employ the logistic c.d.f. throughout this work.

Because decision field theory's D function is defined in terms of events, with decision weights assigned to events rather than ranked outcomes, application of decision field theory to members of the rank-dependent family is only sensible if all choice options in an experiment are comonotonic. In this case, event weights and rank-dependent weights coincide, and all three experiments are structured in this way. For example, in the first experiment, lotteries $risky$ all have probabilities q_h of receiving their high outcome that are in sixths, generated by the roll of a six-sided die. All lotteries are constructed so that $q_h = k/6$ is always the roll "1 or 2 or…k". So $w(k/6)$, the rank-dependent weight on the high outcome h in *risky*, can always be thought of as the decision weight of the event "the die roll is 1 or 2 or…k", while $1 - w(k/6)$, the rank-dependent weight on the low outcome l in *risky*, can always be thought of as the decision weight of the event "the die roll is $k + 1$ or $k + 2$ or… 6." The events and outcome ranks are identically ordered across all option pairs in each experiment: This is comonotonicity (see Quiggin 1993).

Blavatskyy's (2014) stronger utility or SU model is a general approach to constructing probabilistic models of risky choice that will respect first order stochastic dominance or FOSD: That is, the model always attaches a zero probability to choice of first order stochastically dominated options. In its general form, the SU model begins with a definition of two important benchmark options. Let $(risky \vee safe)$ and $(risky \wedge safe)$ denote the least upper bound and greatest lower bound, respectively, on both $risky$ and $safe$ in terms of FOSD.[8] Let V denote the functional representation of option value for some decision theory. Then in the general SU model, $D(risky, safe) = V(risky \vee safe) - V(risky \wedge safe)$, and

(A3) $$P^{rdbf} = Prob(risky) = H_\lambda\left(\frac{V(risky)-V(safe)}{V(risky \vee safe)-V(risky \wedge safe)}\right).$$

For the choice pairs in the first experiment, $(risky \vee safe) = (0,1-q_h, q_h)$ and $(risky \wedge safe) = (1-q_h, q_h, 0)$. Applying the RDU representation to these lotteries,

(A4) $$RDU(risky \vee safe) - RDU(risky \wedge safe) =$$
$$w(q)u(h) + [1-w(q)]u(m) - w(q)u(m) - [1-w(q)]u(l) =$$
$$w(q)[u(h)-u(m)] + [1-w(q)][u(m)-u(l)],$$

which is the denominator appearing in eq. 6 defining the SU model for these choice pairs.

Given a suitable choice of the function H_λ, equivalence of eqs. A3 and 7 may be established as follows. Let $R = risky$ and $S = safe$. From eq. A3 and the definitions $U = (R \vee S) = (0,1-q_h, q_h)$ and $L = (R \wedge S) = (1-q_h, q_h, 0)$ for the option pairs in the first experiment, Blavatskyy's model is

(A5) $$P^{rdbf} = Prob(R) = H_\lambda\left(\frac{V(R)-V(S)}{V(U)-V(L)}\right).$$

Choose $H_\lambda(x) = \Lambda\left[\lambda \ln\left(\frac{1+x}{1-x}\right)\right]$. For $x \in (-1,1)$, this has the needed properties $H_\lambda(0) = 0.5$ and $H_\lambda(x) = 1 - H_\lambda(-x)$. With $x = \frac{V(R)-V(S)}{V(U)-V(L)}$, we have

---

[8] That is, $(risky \vee safe)$ stochastically dominates both $risky$ and $safe$, but is itself stochastically dominated by every other option that stochastically dominates both $risky$ and $safe$. Similarly, $risky$ and $safe$ both stochastically dominate $(risky \wedge safe)$, and every other option stochastically dominated by both $risky$ and $safe$ is itself stochastically dominated by $(risky \wedge safe)$.

(A6) $$\frac{1+x}{1-x} = \frac{1+\frac{V(R)-V(S)}{V(U)-V(L)}}{1-\frac{V(R)-V(S)}{V(U)-V(L)}} = \frac{V(U)-V(L)+V(R)-V(S)}{V(U)-V(L)+V(S)-V(R)} = \frac{[V(U)-V(S)]+[V(R)-V(L)]}{[V(U)-V(R)]+[V(S)-V(L)]}.$$

Applying the RDU representation theorem to the four key options,

(A7) $$V(R) = w(q_h)u(h) + [1-w(q_h)]u(l), \quad V(S) = u(m),$$
$$V(U) = w(q_h)u(h) + [1-w(q_h)]u(m), \text{ and}$$
$$V(L) = w(q_h)u(m) + [1-w(q_h)]u(l).$$

Substitute these into the four bracketed terms at the right end of (A6) to get

(A8) $$[V(U)-V(S)] = w(q_h)[u(h)-u(m)],$$
$$[V(R)-V(L)] = w(q_h)[u(h)-u(m)],$$
$$[V(U)-V(R)] = [1-w(q_h)][u(m)-u(l)], \text{ and}$$
$$[V(S)-V(L)] = [1-w(q_h)][u(m)-u(l)].$$

Clearly $\frac{1+x}{1-x} = \frac{w(q_h)[u(h)-u(m)]}{[1-w(q_h)][u(h)-u(m)]}$, so the equivalence to eq. 7, given RDU and a suitable choice of H_λ, has been established.

In the case of the second and third experiments, where $safe = (p_l, p_m, p_h)$ and $risky = (q_l, q_m, q_h)$, we have $(risky \vee safe) = (p_l, 1-p_l-q_h, q_h)$ and $(risky \wedge safe) = (q_l, 1-q_l-p_h, p_h)$. Algebraic steps resembling those from eqs. A3 to A8 lead to the following elaborated version of eq. 7 that is suitable for data from the second and third experiments:

(A9) $$P^{su} = Prob(risky) = F\left[\lambda \ln\left(\frac{[w(q_h)-w(p_h)][u(h)-u(m)]}{[w(1-p_l)-w(1-q_l)][u(m)-u(l)]}\right)\right].$$

This particular instance of Blavatskyy's (2014) stronger utility is also an instance of Fishburn's (1978) incremental expected utility advantage model of probabilistic choice, so one might usefully refer to eq. A9 as the Blavatskyy-Fishburn model.

Appendix II: Estimation notes

All estimations were carried out in SAS 9.2 using the nonlinear programming procedure ("Proc NLP" in the SAS language) using the quasi-Newton algorithm. For function-free estimations all parameters bounded in the interval [0,1], that is utilities and weights, were constrained to lie in [0.0001,0.9999]; additionally, monotonicity was imposed on estimated utilities and weights.

Monte Carlo simulations showed that both finite sample biases of parameter estimates and prediction log likelihoods could be noticeably improved by penalizing estimation that produced fitted probabilities very close to zero or one. By a grid search across Monte Carlo simulations, the following piecewise quadratic penalty function $p_i(\boldsymbol{\theta}^s)$ was arrived at as a good kludge for penalizing such fitted probabilities:

$p_i(\boldsymbol{\theta}^s) = 0$ if $P_i^{spec}(\boldsymbol{\theta}^s) \in [0.001,0.999]$;

$p_i(\boldsymbol{\theta}^s) = -10 \cdot \left(1 - 1000P_i^{spec}(\boldsymbol{\theta}^s)\right)^2$ if $P_i^{spec}(\boldsymbol{\theta}^s) < 0.001$; and

$p_i(\boldsymbol{\theta}^s) = -10 \cdot \left(1000P_i^{spec}(\boldsymbol{\theta}^s) - 999\right)^2$ if $P_i^{spec}(\boldsymbol{\theta}^s) > 0.999$.

This simply imposes a very steep but smoothly differentiable penalty on probabilities that wander within 0.001 of zero or one. The adjusted log likelihood function is

$$\mathcal{L}^{spec}\left(\mathbf{r}_{set(k)}^s \middle| \boldsymbol{\theta}^s\right) = \sum_{it \in set(k)} \ell^{spec}(r_{it}^s | \boldsymbol{\theta}^s) + \sum_i \tau_i p_i(\boldsymbol{\theta}^s),$$

where τ_i denotes the number of trials of pair i in any experiment. This penalty was imposed on all maximum likelihood estimations.

For each subject and specification, estimations were started from a grid of starting parameter vectors to a "finalist" estimated vector from each starting vector, and the finalist with the best adjusted log likelihood was selected as the maximum likelihood estimate.

Appendix III: First experiment protocol explanation and instructions to subjects

I want to estimate utilities and weights without aggregation assumptions. Decision theories are about individuals, not aggregates, and aggregation mutilates and destroys many observable properties of decision theories (Wilcox 2008). A large amount of choice data from each subject is needed to reliably estimate utilities and weights at the individual level. A subject will become bored, and will become careless, if she makes hundreds of decisions at one sitting. So the decisions are divided up across three days, and on each day into two parts separated by unrelated tasks providing a break from decisions.

The separation across three days, in particular, introduces a risk that some substantial event altering a subjects' wealth or background risks will occur between days, which could arguably undermine the assumption that utilities of outcomes and hence choice probabilities are stationary throughout the protocol. This is a risk I am willing to run to mitigate subject boredom with hundreds of choice tasks, and I can check whether distributions of risky choice proportions across subjects appear to be stationary across subjects' three days of decisions. No test finds any significant difference between these three daily distributions. Within-subject differences between risky choice proportions on the first and third day have zero mean by all one-sample tests. There is some evidence that decisions are less noisy on the second and third days versus the first day (see Wilcox 2015). Econometric allowance for this (estimating a separate precision parameter for each day) has no qualitative effect on my results in Section 5.

Random problem selection or RPS is meant to result in truthful, motivated and unbiased revelation of preferences in each pair: That is, subjects should make each of their 300 choices as if it was the only choice being made, for real, and there should be no portfolio or wealth effects making choices interdependent across the tasks. Both the independence axiom of EUT and the "isolation effect" of prospect theory would imply this. To see this for EUT, notice that the independence axiom in its "unreduced compounds" form (i.e. "compound independence") implies

$$safe \succcurlyeq risky \quad \text{if and only if} \quad (risky \text{ with Prob} = 1/300; Z \text{ with Prob} = 299/300) \succcurlyeq (safe \text{ with Prob} = 1/300; Z \text{ with Prob} = 299/300)$$

…where Z is any other outcome or risk, including the "grand lottery" created by the subject's other 299 choices over the course of this experiment. Therefore, if subjects' preferences satisfy independence in this unreduced compounds form, random problem selection should be incentive compatible. Some evidence suggests that preferences generally satisfy the independence axiom in its unreduced compounds form (Kahneman and Tversky 1979; Conlisk 1989), and older direct examinations of random problem selection in binary lottery choice experiments found no systematic choice differences between tasks selected with relatively low or high probabilities (Wilcox 1993) nor between tasks presented singly or under random problem selection (Starmer and Sugden 1991), at least for relatively simple tasks like the pairs used here. The literature examining the RPS is moderately large; recent pessimistic evidence includes Harrison and Swarthout (2014) and Cox, Sadiraj and Schmidt (2015) while Brown and Healy (2018) found more optimistic evidence but only when using the SED (separated decisions) feature I use in my three experiments (each option pair is presented alone on its own separate screen in random order).

The choice pairs in Table 1 are on twenty-five distinct contexts, all constructed from nine positive money outcomes ($40 to $120 in $10 increments). I want to estimate the utilities and weights in the function-free manner Hey and Orme (1994) pioneered for utilities, Hey et al. (2010) did for utilities and subjective probabilities, and Blavatskyy (2013) did for utilities and weights. Monte Carlo simulations showed that function-free identification of utilities, weights and scale parameters is greatly improved when the same events (the die rolls) are matched with many different outcomes on different contexts.

Finally, the choice of a six-sided die for the first experiment was deliberate. Sixths are well-suited for estimation given widely-held priors about the shape of weighting functions. Consider Prelec's (1998) single-parameter weighting function $w(q|\gamma) = \exp(-[-\ln(q)]^{\gamma})$ $\forall\ q \in (0,1)$, $w(0)=0$ and $w(1)=1$: Prelec proposed $\gamma = 0.65$ as a rough estimate consistent with other estimates using different weighting functions. At that value of γ, $q - w(q|0.65)$

attains its maximum very close to $q = 5/6$; and at $q = 1/6$, $q - w(q|0.65) \approx -0.065$, about 80% of the minimum value taken by $q - w(q|0.65)$ (this is about –0.081 at $q \approx 0.07$). So the differences between linear weighting (that is EU) and received priors concerning probability weighting are about as strong as they could be at $q = 5/6$ and $q = 1/6$.

Instructions [first experiment only]

You will participate in 3 different sessions—one session on each of 3 different days. On **each** of the three days, you will make **100 choices** from each of 100 pairs of monetary options. Some of the options will involve chance, in the form of a die roll. Option pairs will be presented to you as pie charts, on a computer screen: In each option pair you see, you will choose the option you would prefer to play.

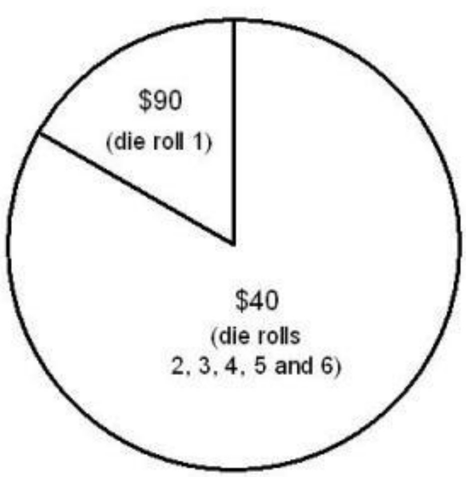
At the end of your third day with us, you will have made 300 choices over your three sessions. ONE of your 300 option choices will then be randomly selected using a bag of 300 tickets with the numbers 1, 2, 3,…, 299, 300 written on them. The numbers 1 to 100 correspond to the 100 choices you will make today, in the order you make them today. Likewise, the numbers 101 to 200 (and 201 to 300) correspond to the 100 choices you will make on your second day (and then on your third day) with us, in the order you make them on those days.

At the end of your third day with us, you will reach into the bag of tickets (without looking inside), pull one out and show us the number. We will then enter that number into the computer, and it will recall that option pair and show the option you chose. That option will determine your payment for participation in this project. If the option you chose requires a die roll, we will then roll a six-sided die to determine your payment.

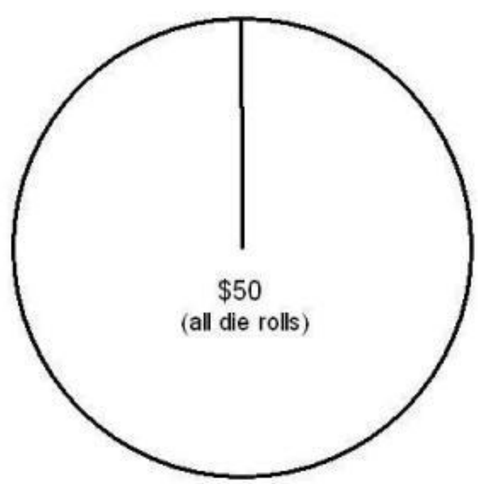
Notice that since **every** option pair choice you make has a 1 in 300 chance of determining your payment for participation, you have a real reason to consider each option pair with equal care. Also, notice that **only one** of your 300 option pair choices **will** determine your payment.

Please note that you won't be able to use a calculator, or pencil and paper, to make your choices. That would take too long for 100 choices…our lab schedule will not accommodate this.

Left Option



Right Option



(Instructions to subjects—continued)

An example of an option pair is shown above. The left option is a 1 in 6 chance of $90 and a 5 in 6 chance of $40: If you chose this option and it was selected to determine your payment, a die roll would be needed to determine the payment. The right option is a sure $50: If you chose this option and it was selected to determine your payment, no die roll would be needed.

The option pair you just saw is only one example. The money outcomes in the option pairs you see will range from $40 to $120, in ten dollar increments. Also, the connection between die rolls and money outcomes varies a lot over those options that involve a die roll, so remember to notice those die rolls when new option pairs appear on the screen for your consideration. Finally, note that the computer will present option pairs to you in a randomized order, and will also randomly select the left/right placement of the options in each pair. So you do not want to assume that option pairs appear in any kind of patterned sequence: They do not. The computer will remember the exact sequence, as well as what you chose, so that you can be paid properly on your last day with us.

Some questions for a break

It is difficult to maintain good attention over 100 choices. Even though the amounts of money in option pairs are not small, almost anyone will get a bit bored with making these kinds of choices after awhile.

Partly for that reason, the 100 option pair choices will be broken into two halves (50 pairs in each half) on each day. Between the halves, on each day, you will answer some survey questions and respond to some questionnaire items. This will go pretty quickly on all three days (a little longer on the second day), and will give you a break each day from the option pair choices.

You'll be able to do everything at your own pace. We believe that each session will last about one hour for most people on most days, but remember that we expect you to have 90 minutes available on each day, so that you are not rushed.

If there is anything you do not understand, please ask us. We will be happiest if you understand exactly how your decisions affect you: We want you to be able to do well for yourself, whatever you believe “doing well” is. We encourage you to do what you want.

Finally, the money for this study comes from grants. This money is earmarked for payment to student participants. We have no alternative use for this money: It must be paid out to participants like you. We must of course make payments only in accordance with the procedure we have described above. But do not worry about taking that money from us: It is specifically earmarked for this and we cannot use it for anything else. We say this, only because some students worry about taking such money from professors. You should not worry about it. The money is grant money, not Dr. Wilcox’s money, and it is earmarked specifically for paying out to student participants like yourself.