http://dx.doi.org/10.1007/s11166-012-9154-4

Butler, D.J., Isoni, A. and Loomes, G.C. (2012) Testing the 'standard' model of stochastic choice under risk. Journal of Risk and Uncertainty, 45 (3). pp. 191-213.

Testing the 'standard' model of stochastic choice under risk

David Butler[1], Andrea Isoni[2,3], and Graham Loomes[2]

1.Murdoch Business School, Perth, Australia

2.Behavioural Science Group, Warwick Business School, Coventry, UK

3.Department of Economics, University of Warwick, Coventry, UK

Abstract

Models of stochastic choice are intended to capture the substantial amount of noise observed in decisions under risk. We present an experimental test of one model, which many regard as the default—the *Basic Fechner model*. We consider one of the model's key assumptions—that the noise around the subjective value of a risky option is independent of other features of the decision problem. We find that this assumption is systematically violated. However the main patterns in our data can be accommodated by a more recent variant of the Fechner model, or within the random preference framework.

Keywords: Choice under risk; Stochastic choice; Fechner model; Random preference

JEL classification: D81 C91

Due to its simplicity, elegance and versatility, John von Neumann and Oskar Morgenstern's (1947) Expected Utility Theory (EUT) has been one of the most influential and widely used tools in economic analysis during the past 65 years. However, early attempts to elicit individuals' utility functions (e.g. Mosteller and Nogee 1951) highlighted the fact that many people's responses were characterised by a substantial degree of 'noise', which prompted the development of research into probabilistic choice (see Luce and Suppes 1965 for an early review). At the same time, experimental tests of the descriptive validity of EUT identified a number of seemingly systematic violations of its axioms (see Camerer 1995 for a review), which led to the development of a large number of alternative theories (see Starmer 2000).

These violations are 'seemingly systematic' because it has become clear that whether any particular empirical regularity can be regarded as a violation of EUT (or any other theory) turns out to depend crucially on the assumptions made about the stochastic specification of the decision-making process (see Loomes and Sugden 1995; Wilcox 2008). Many of what have for years been treated as systematic violations of EUT can be accommodated within the EU framework under *some* assumptions about the stochastic specification, although they contradict EUT under *other* specifications of the noise in people's responses (e.g. Loomes 2005; Blavatskyy 2006, 2009). Indeed, the assumptions made about the stochastic specification may be as important, if not even more important, than assumptions about which 'core' theory is invoked (e.g. Blavatskyy and Pogrebna 2010).

We cannot yet identify just which stochastic specification is most appropriate. But in some settings, one can rule out specifications that are known to be problematic. For instance, a model in which individuals are assumed to commit errors at a rate that does not depend on any of the characteristics of the decision problem (Harless and Camerer 1994) is inadequate in most circumstances (Loomes and Sugden 1998).

The literature has also documented various reasons for concern about another very popular stochastic model, the *Basic Fechner (BF)* model (Fechner 1860, 1966; Hey and Orme 1994), which many regard as the ‘standard’ stochastic model of choice under risk. In this model, a symmetric, zero-mean error[1] term is attached to the true subjective value of a risky option. On the empirical side, it has been shown that this model entails many more violations of first-order stochastic dominance than are usually observed (e.g. Loomes and Sugden 1998; Loomes et al. 2002). On the theoretical side, even if such a model is used to estimate parameters meant to reflect agents’ degrees of risk aversion in the Pratt (1964) sense, the estimated parameters do not necessarily satisfy a ‘statistically more risk averse’ relation (Wilcox 2011).

However, in spite of such serious limitations, the BF model is still popular in applied work (e.g. Harrison et al. 2007; Bruhin et al. 2010; von Gaudecker et al. 2011; Epper et al. 2011).[2] A possible reason for this is the relative ease of econometric implementation. But tractability, while useful, is not sufficient: using an inadequate model could well result in the serious mis-estimation of parameters and incorrect inferences about the structure of people’s preferences.

In this paper we report an experimental test of one of the central assumptions of the BF model, namely, that the error terms attached to the subjective values of the risky prospects in a binary choice are independent of each other. Our experiment was motivated by some data from a preference reversal experiment by Butler and Loomes (2007), who found that imprecision intervals around certainty and probability equivalents for standard P-bets and $-bets were systematically related to the characteristics of the prospect being valued: contrary to the BF independence assumption, the $-bet appeared to exhibit more noise than the P-bet in the certainty equivalent exercise, while the P-bet showed greater imprecision than the $-bet in the probability equivalence task. We say more about this in Section 2.

In order to explore this issue more directly, we constructed a series of binary choice problems in which P- and $-bets were matched with two distinct sets of progressively better common alternatives. As we explain in Section 1, the BF independence assumption entails that any inference concerning the relative noisiness of the errors associated with the P- and $-bets should not depend on which set of comparators is used. Our results show that this independence assumption is systematically violated. Prospects like the P-bets, which are more similar to certain amounts, are inferred to have lower variance error terms than the $-bet when both are matched with an increasing sequence of certainties. However, when both bets are matched with a series of lotteries that are riskier than the $-bet, the P-bets are inferred to have higher variance error terms.

Our new test adds to the evidence of the inadequacy of the BF model in many contexts. In Section 3, we turn to the important issue of which stochastic specification is able to come to grips with the systematic patterns that we find. We conduct some simple calibration exercises which show that these patterns can be accommodated by the extension of the BF model proposed by Blavatskyy (2009, 2011), but not by the contextual utility model proposed by Wilcox (2011). These models have the common feature that the variance of the error term is, in effect, conditioned on features of a particular binary choice. However, they implement this contextualisation in different ways, and in Section 3 we explain why this can organise the data in one case but not in the other.

An alternative way of explaining our findings is provided by the random preference framework (e.g. Becker et al. 1963; Loomes and Sugden 1995) in which agents act as if the parameters of their core preferences have random components. In Section 3 we show that because P-bets are less risky than $-bets, their evaluation against a common set of certainties on the basis of a given distribution of risk aversion coefficients is characterised by less variability than the corresponding evaluation of $-bets, whereas the opposite is true when both bets are matched with prospects which are even riskier than the $-bet.

As we point out in the concluding section of the paper, our findings emphasise the pressing need to go beyond the BF modelling approach, and we suggest avenues for further research into how best to model the stochastic component in decision-making under risk.

1 A test of the Basic Fechner model

In line with most of the literature, we consider decisions between pairs of lotteries. A *lottery* is a probability distribution over a set of monetary *consequences* $X = \{x_0, \ldots, x_n\}$, with $x_0 < x_1 < \ldots < x_n$. Typical lotteries are denoted by vectors of probabilities such as $\mathbf{p} \equiv (p_0, \ldots, p_n)$ and $\mathbf{q} \equiv (q_0, \ldots, q_n)$, with $p_0 + p_1 + \ldots + p_n = 1$ and $q_0 + q_1 + \ldots + q_n = 1$. A *binary choice* is denoted by $\{\mathbf{p}, \mathbf{q}\}$.$x_0$

Most deterministic decision theories assume the existence of some function V(.) that associates each lottery with an individual's *subjective value*. Different theories are defined by different ways of computing V(.). In any choice $\{\mathbf{p}, \mathbf{q}\}$, $\mathbf{p}$ is preferred to, equally preferred to or less preferred than $\mathbf{q}$ according to whether $V(\mathbf{p})$ is greater than, equal to or less than $V(\mathbf{q})$. More succinctly:

$$\mathbf{p} \begin{array}{c}\succ\\ \sim\\ \prec\end{array} \mathbf{q} \text{ iff } V(\mathbf{p}) - V(\mathbf{q}) \begin{array}{c}>\\ =\\ <\end{array} 0 \qquad (1)$$

The Fechner model (Fechner 1860, 1966) offers a simple mechanism for making the deterministic decision rule (1) stochastic. According to this model, each $\mathbf{p}$ is evaluated with some noise, ε_p, which is usually assumed to be symmetric and to have mean zero and variance σ^2_p. This error term is also assumed to be *lottery-specific*, i.e. not to depend on features of the decision problem *other than* the characteristics of the lottery itself. This assumption, which is of particular importance for the test that we derive in the remainder of this Section, is not shared by more recent variants of the Fechner model such as those proposed by Blavatskyy (2011) and Wilcox (2011). We will come back to these variants later in the paper, but for the moment our principal concern is with the *Basic Fechner* version just described. This version is less restrictive than the one often used in applied work, in which a symmetric, mean-zero error term is attached to the *difference* between the subjective values of the two lotteries, entailing that the lottery-specific error term has the same variance for all lotteries.[3] Such an assumption is unnecessarily restrictive.

Once the error terms are introduced, choice becomes probabilistic:

$$\Pr(\mathbf{p}\,|\{\mathbf{p},\mathbf{q}\}) = \Pr[\mathrm{V}(\mathbf{p}) - \mathrm{V}(\mathbf{q}) + (\varepsilon_{\mathrm{p}} - \varepsilon_{\mathrm{q}}) > 0] \tag{2}$$

where Pr(**p**|{**p**, **q**}) is the probability that **p** is chosen in decision problem {**p**, **q**}. To keep with the standard practice followed in most applications, ε_p and ε_q can be thought of as normally distributed error terms. The BF *independence* assumption then entails that cov(ε_p, ε_q) = 0, so that the joint variance of the error terms is $\sigma^2_p + \sigma^2_q$.

Now consider what happens to Pr(**p**|{**p**, **q**}) if lottery **q** is progressively improved by moving probability mass from low-value to higher-value consequences. For simplicity, let us assume for the moment that the error term associated with each **q** has the same variance σ^2_q. Figure 1illustrates. The solid line represents Pr(**p**|{**p**, **q**}) as a function of V(**q**). When V(**q**) is very low, **p** is always chosen. But as **q** is improved and V(**q**) increases, a point is reached where the difference between V(**p**) and V(**q**) may occasionally be outweighed by particular combinations of ε_q and ε_p, so that the probability of choosing **p** falls below 1; and Pr(**p**|{**p**, **q**}) falls further as V(**q**) continues to rise. When V(**q**) = V(**p**), errors in the two directions balance each other on average, so that Pr(**p**|{**p**, **q**}) = 0.5. We will call this the *stochastic indifference point*. Further increases in V(**q**) progressively reduce Pr(**p**|{**p**, **q**}), until it reaches a point where **p** is rarely, and eventually never, chosen.

The slope of the Pr(**p**|{**p**, **q**}) function is determined by the combined variance of the error terms—that is, by $\sigma^2_p + \sigma^2_q$. When ε_p and ε_q are normally distributed, the shape is the mirror image of a normal c.d.f.

Now take another lottery, **r**, with error term ε_r, and such that V(**r**) < V(**p**). To start with, suppose that $\sigma^2_r = \sigma^2_p$, so that $\sigma^2_r + \sigma^2_q = \sigma^2_p + \sigma^2_q$ for all **q**. The resulting Pr(**r**|{**r**, **q**}) is shown as the dashed line in Fig. 1. As the combined variances are the same, the dashed line is everywhere to the left of the solid line. Because of the strict monotonicity of the c.d.f. function, there is no point at which the two lines in Fig. 1 cross. The same would happen with any other symmetric, strictly increasing c.d.f. such as those used in applications.

However, the simplification that $\sigma^2_r = \sigma^2_p$ is by no means an essential characteristic of the BF model. Figure 2 considers the case in which $\sigma^2_r > \sigma^2_p$, i.e. there is more noise around V(**r**) than around V(**p**). This alters the Pr(**r**|{**r**, **q**}) curve, increasing the range over which Pr(**r**|{**r**, **q**}) lies between 1 and 0: for any given **q** there is a greater likelihood that noise overturns the difference between the subjective values of the two lotteries, making the whole curve flatter around the stochastic indifference point. Because V(r) < V(p), the dashed line must lie to the left of the solid line for all Pr(**p**|{**p**, **q**}) ≥ 0.5. However, the flatter trajectory of the dashed line allows the possibility that it *may* cut the solid line from below[4]; in which case, for all V(**q**) to the left of that crossover point, Pr(**p**|{**p**, **q**}) > Pr(**r**|{**r**, **q**}), whereas for all V(**q**) to the right of that point, Pr(**p**|{**p**, **q**}) < Pr(**r**|{**r**, **q**}).

So, if we were to observe repeated choices between **p** and the $\mathbf{q}_{low}$ marked in Fig. 2, and also between **r** and $\mathbf{q}_{low}$, we should see **p** chosen from {**p**, $\mathbf{q}_{low}$} more often than **r** is chosen from {**r**, $\mathbf{q}_{low}$}. This difference is represented by the vertical distance (vd_{low}) shown by the down arrow. By contrast, with a comparator lottery sufficiently attractive that it lies to the right of the crossover point—e.g. $\mathbf{q}_{high}$—we should see **p** chosen from {**p**, $\mathbf{q}_{high}$} *less* often than **r** is chosen from {**r**, $\mathbf{q}_{high}$}, with this difference being represented by the vertical distance (vd_{high}) shown by the up arrow.

Under the BF model, where by assumption ε_p, ε_q and ε_r are all independent of one another, this is the case whatever the particular nature of **q**. So it is true if **q** is a certainty, with V(**q**) increasing as the size of the sure payoff increases. And it is also true if **q** is some probability mix of a positive sum x_m and zero, with V(**q**) increasing as the probability of x_m increases (and the probability of zero correspondingly decreases). This provides a simple way of testing the BF model. If one is able to map the Pr(**p**|{**p**, **q**}) and Pr(**r**|{**r**, **q**}) curves by matching both **p** and **r** with two distinct sets of progressively better **q** lotteries, then the mappings deriving from the two distinct sets should have the same implications as to whether $\sigma^2_r = \sigma^2_p$, or $\sigma^2_r > \sigma^2_p$, or $\sigma^2_r < \sigma^2_p$.

Notice that our working assumption that all **q**'s have the same variance is inconsequential if **p** and **r** are matched with exactly the same set of **q**'s. What matters for the test is whether, for any given **q**, the joint variance $\sigma^2_r + \sigma^2_q$ is less than, equal to, or greater than the joint variance $\sigma^2_p + \sigma^2_q$, and this clearly depends only on the relationship between σ^2_r and σ^2_p.

2 Experimental design and results

2.1 Motivation

Our experimental design was motivated by the findings of the Preference Reversal (PR) experiment by Butler and Loomes (2007). In a typical PR experiment, subjects make decisions involving two lotteries, a *P-bet* offering a modest prize with a fairly high probability (and nothing otherwise), and a *$-bet* offering a larger prize with a smaller probability (and nothing otherwise).[5] Subjects choose between the two lotteries and report a money equivalent (often called a certainty equivalent) for each. It is usually found that the pattern of inconsistent responses is highly asymmetrical, with the subjects who choose the P-bet but value the $-bet more (a *standard* reversal) greatly outnumbering those who choose the $-bet but put a higher value on the P-bet (a *counter* reversal); see Seidl (2002) for a review. Butler and Loomes reported that this pattern was reversed if, instead of reporting a money equivalent for the two lotteries, subjects were asked to report a probability equivalent in the form of a probability of winning a sum of money fixed at a higher level than the payoff offered by the $-bet. In this setting, counter reversals outnumbered standard ones.

Butler and Loomes also elicited the *intervals* of money and probability equivalents over which subjects were less than completely sure about their preferences. They found that money equivalent intervals were wider for the $-bet, while probability equivalent intervals were wider for the P-bet. The patterns of money equivalent intervals were consistent with the possibility that the variance of the error term was greater for the $-bet than for the P-bet. On the other hand, the patterns of probability equivalent intervals were in line with the opposite possibility, namely that the variance of the error term was greater for the P-bet than for the $-bet. Clearly, according to the BF model, only one of these possibilities can be true. So the Butler and Loomes imprecision interval data cast doubt upon the BF model.

However, those intervals were elicited by a procedure that could not be made incentive compatible,[6] and might therefore be regarded with a degree of scepticism by those who believe that participants in experiments will only provide valid responses if these are linked to material rewards. On the other hand, Butler and Loomes's experiment did also produce some incentive-compatible data that provide further suggestive evidence. The participants made a number of straight choices in which the $-bet (offering A$80 with 0.25 probability) and the P-bet (A$24 with 0.7 probability) were matched with five predetermined sure amounts of Australian dollars (A$8, A$12, A$16, A$20 and A$24 for the $-bet and A$4, A$8, A$12, A$16 and A$20 for the P-bet) and five lotteries offering different predetermined chances of winning an amount (A$160) twice as large as that offered by the $-bet (namely, 0.1, 0.12, 0.15, 0.18 and 0.2 for the $-bet, and 0.1, 0.15, 0.2, 0.25 and 0.3 for the P-bet). The proportions of participants choosing the $-bet and P-bet in these tasks are plotted in Fig. 3. When the two PR lotteries are matched with increasing certain amounts (see panel A), the proportion of participants choosing the $-bet decreases more slowly than that of participants choosing the P-bet. When the two lotteries are matched with increasing chances of winning A$160 (see panel B), the proportion of those choosing the $-bet is always smaller, but it decreases more rapidly than the proportion of participants choosing the P-bet. The two curves in panel A cross in a manner consistent with the error of the $-bet having greater variance. Although the two curves in panel B do not cross, the flatter slope of the P-bet curve is consistent with a greater error variance for the P-bet than for the $-bet. These patterns are analogous to those found in the data on money and probability equivalent intervals and are in conflict with the BF model. However, they are somewhat limited in scope and cannot be regarded as much more than suggestive of the need for further experimental investigation.

2.2 Design

Our experiment uses the same kinds of comparisons, but attempts to do so for parameters for which both sets of curves cross, and for sets of lotteries which are exactly the same for all PR bets. Participants make a series of decisions in which a P-bet and a $-bet are each paired with two sets of common alternatives. The $-bet offers a 25% chance of winning £40 and is the same for all subjects. There are two P-bets, one offering a 90% chance of winning £10 and the other offering a 65% chance of winning £10, randomly allocated between two subsamples of participants. We use two different P-bets in order to increase the likelihood of identifying the regions where the two curves cross for both sets of common alternatives.

The first of these sets consists of four sure sums £4, £6, £8 and £10. The second set consists of lotteries offering, respectively, a 10%, 15%, 20% and 25% chance of winning £60 (and nothing otherwise). In what follows, it will sometimes be convenient to refer to these lotteries as L10, L15, L20 and L25 respectively, or *L lotteries* as a group. Each participant faces the following choices: four in which the P-bet is matched with the four certainties; four in which the P-bet is matched with the four L lotteries; four in which the $-bet is matched with the four certainties and four in which the $-bet is matched with the four L lotteries. In addition, participants make straight choices between the P-bet and the $-bet. An important aspect of our design, which was absent in Butler and Loomes's, is that every one of these choices is repeated a minimum of three times during the course of the experiment (the {$, P} choice is repeated four times). We will come back to this aspect later in this Section. So, for each individual, we have a total of 52 decisions involving these lotteries.[7]

In each choice, the alternatives are labelled A and B. An example of the way choices were displayed is given in Fig. 4, which reproduces one of the pairs used in the experiment. The amounts to be won are shown in boxes for which the width is proportional to the probability of winning, which is also shown as a percentage underneath the money amounts. Lotteries are played out by drawing one of 100 numbered discs from a bag; the numbers associated with each amount are shown at the top of each box.

At the beginning of each task, the cursor is located in the middle of the bar. Subjects choose A by moving it to the left towards the large letter A; they choose B by moving it to the right towards the large letter B. They are asked to move the cursor along the bar so as to indicate their strength of preference for whichever lottery they choose.[8] Subjects are not allowed to report indifference. At the end of the experiment, one decision problem is randomly selected for each subject: the chosen lottery is played out and the subject is paid the resulting amount in cash, with no show-up fee.[9] With this procedure, choices are incentive-compatible.[10]

2.3 Results

The experiment was conducted at the University of East Anglia in May 2009. The 138 subjects who took part were recruited via email shots from the general student population. The instructions were read aloud at the beginning of each session. Clarifying questions were answered publicly. The instructions also included some questions designed to check subjects' understanding of how lotteries were displayed and played out (see Appendix for details). On average, experimental sessions took less than an hour to complete.

The experimental results are presented in Table 1, which reports the proportions of subjects choosing each PR bet against each of the four certainties, and against each of the four L lotteries that offer various chances of winning £60. The first column of the table indicates which observations the data refer to. Each choice was repeated three times. The proportions of subjects choosing each PR bet in each repetition are reported separately. We also report the proportions aggregating over the three repetitions (denoted by 'Aggr.') and the proportions computed using each individual's median choices ('Median'). Half of the sample (69 subjects) made choices involving the relatively attractive P-bet, P1 = (£10, 0.9), while the other 69 subjects faced the somewhat less attractive P-bet, P2 = (£10, 0.65). The $-bet, $ = (£40, 0.25) was the same for both subsamples.[11] The observations for the $-bet are pooled across the two subsamples.[12]

Table 1 shows that the choice proportions vary somewhat between repetitions. This is exactly the kind of variability that models of stochastic choice are intended to capture. Another feature of the data is that some choice proportions—those of decision problems {P1, £10}, {P2, £10} and {$, L25}, where L25 = (£60, 0.25)—are extremely low. In those problems, the PR lotteries are first-order stochastically dominated by the alternative. According to most deterministic decision theories, dominated lotteries should never be chosen, except in error.[13]

The two sets of decisions in which the $-bet and the P-bets are matched with progressively better alternatives can be used to map aggregate versions of the kinds of curves shown in Figs. 1 and 2. Regardless of which set of data is used, much the same picture emerges. Figure 5 illustrates the curves aggregating across the three repetitions.

Panel A shows how the proportions vary when the PR lotteries are matched with increasing certainties. At the aggregate counterpart of the stochastic indifference point—i.e. at the 0.5 point on the vertical axis—we can see that P1 ≻ $ ≻ P2. The two curves relating to the P-bets are steeper than that for the $-bet. If we were to take aggregate data as reflecting the preferences of 'representative agents' and if we were to apply the BF model accordingly, this would lead us to infer that the error term associated with $ has greater variance than that associated with each P-bet.

In Panel B, the curves for both P-bets are flatter than the curve for the $-bet, with P1 ≻ P2 ≻ $ at the 0.5 point on the vertical axis. From this picture, we would infer that the error term associated with $ now has a *smaller* variance than those associated with P1 and P2.

These inferences appear to contradict the BF model, but are subject to an important caveat. The test described in relation to Fig. 2 refers to a single individual, while our analysis has focused on aggregate choice proportions *as if* they were typical of a representative agent. In order to obtain a really detailed description of each separate individual, one would need to ask each respondent a very large number of questions, which poses serious practical problems. In our experiment, each choice was repeated three times. This is clearly not enough for us to be able to map the probability curves separately for each individual in the kind of detail we would ideally have liked. However, we can use these repetitions to see whether individuals show the same broad patterns that we observe at the aggregate level. Consider the choices between the PR bets and the certain amounts. The key tendencies that are driving our results are the following. The $-bet is chosen over high certainties relatively more frequently than the P-bets, but as the certainties decrease the pattern tends to be reversed. The opposite happens when the comparator lotteries involve different chances of winning £60: the $-bet is chosen relatively more frequently than the P-bet over low probabilities of £60, but relatively less frequently for higher probabilities.

To see how far individual participants show these patterns, it is useful to refer back to the vertical distances between the curves in Fig. 2. In order to draw a parallel with that Figure, we look at the first and third lowest certainties and probabilities, which seem to best approximate the $\mathbf{q}_{\text{low}}$ and $\mathbf{q}_{\text{high}}$ shown there. For both sets of choices, we define chP_{low} ($\text{ch\$}_{\text{low}}$) as the number of times out of three that the individual chooses the P-bet ($-bet) over the *low*certainty (£4) or probability (0.1), and we define chP_{high} ($\text{ch\$}_{\text{high}}$) as the number of times the individual chooses the P-bet ($-bet) over the *high* certainty (£8) or probability (0.2). On that basis, we construct the following variable:

$$\Delta\text{vd} = (\text{chP}_{\text{low}} - \text{ch\$}_{\text{low}}) - (\text{chP}_{\text{high}} - \text{ch\$}_{\text{high}}) \tag{3}$$

where (chP_{low}–$\text{ch\$}_{\text{low}}$) is the vertical distance between the two curves at the low certainty (probability) for the individual—i.e. the counterpart of vd_{low} in Fig. 2—while (chP_{high}–$\text{ch\$}_{\text{high}}$) is the vertical distance at the high certainty (probability) for the same individual—i.e. vd_{high} in Fig. 2. We denote this variable by Δvd to indicate that it stands for the *change in vertical distance*. For individuals who show the pattern we observe at the aggregate level, Δvd should tend to be positive for certainties and negative for L lotteries, and hence greater for certainties than for L lotteries. Table 2 reports the distributions of this variable as one goes from £4 to £8 and from L10 to L20, separately for the two subsamples.

Table 2 leaves little room to doubt that the patterns we observe in the aggregate are a reflection of most individuals behaving in a similar way. In the top eight rows, we can see that there is a far greater tendency for changes to be negative in comparisons involving the L lotteries than in comparisons involving certainties, as reflected by the mean and medians shown in the ninth and tenth row. In both subsamples, the distribution of Δvd shifts significantly in the predicted direction. For 50 out of 69 (72%) members of the P1 subsample, and for 51 out of 69 (74%) of those in the P2 subsample, Δvd is greater in tasks involving certainties than in tasks involving different chances of winning £60. The opposite slope reversal occurs much less frequently: in 9 out of 69 cases (13%) in the P1 subsample, and in 10 out of 69 cases (15%) in the P2 subsample. We can easily reject the null hypothesis that both types of reversal are equally likely to occur ($p < 0.001$ in a binomial test).

3 Going beyond the Basic Fechner model

Our experiment demonstrates that a key assumption of the BF model is systematically violated. In this Section, using calibration exercises we consider whether any extant alternative approaches can accommodate the main patterns we observe in our data. We consider two recent extensions of the Fechnerian approach: the contextual utility model proposed by Wilcox (2011); and the model proposed by Blavatskyy (2009, 2011). We also consider the random preference approach originally proposed by Becker et al. (1963) and revived by Loomes and Sugden (1995).

Before turning to the details of our calibration exercises, it is important to clarify that we are not seeking to fit a particular dataset. Our main objective is to identify different approaches that might reproduce the key features of our data that contradict the BF model, namely, the reversal of the relative slopes of the curves for the $-bet and the P-bets, which is clearly visible in Fig. 5. For this reason, we try as far as possible to keep our assumptions about parameters and functional forms constant across the various approaches.

The two extensions of the basic Fechner model that we consider are similar in spirit, as they both normalise the differences between the subjective values of the two lotteries. In the contextual utility model, the subjective value difference is normalised with respect to the difference between the subjective value of the highest and lowest monetary consequences in the set over which the two alternatives are defined. In this model, the probability that lottery **p**is chosen over lottery **q** can be written as:

$$\Pr(\mathbf{p} \mid \{\mathbf{p}, \mathbf{q}\}) = \mathrm{F}\left[\lambda (\mathrm{V}(\mathbf{p}) - \mathrm{V}(\mathbf{q})) / (\mathrm{V}(\mathrm{x}_n) - \mathrm{V}(\mathrm{x}_0))\right] \tag{4}$$

where F is a cumulative distribution function, λ is a precision parameter, $V(x_n)$ and $V(x_0)$ are, respectively, the utilities of the degenerate lotteries offering the certainty of winning the highest (x_n) and lowest (x_0) monetary consequences amongst those offered by **p** and **q**. This set of monetary consequences represents the *context* of decision problem {**p**, **q**}.

Since x_n and x_0 may vary with the decision problem, it is intuitive that the model has the potential to explain the patterns of our data in the pairs in which the PR lotteries are matched with various certain amounts. For all four pairs, the context ranges from £0 to £40 for the $-bet, and from £0 to £10 for the two P-bets. Since the differences in subjective values are divided by a smaller amount for the two P-bets, for these bets choice probabilities should be more responsive to an increase in the matched certainty than for the $-bet, for which the subjective value difference is divided by a larger amount. A simple calibration exercise, which assumes EUT at the core[14] with a power utility function $U(x) = x^{0.8}$, a precision parameter $\lambda = 5$ and normally distributed errors, shows that this is indeed the case. The resulting choice proportions are shown in panel A of Fig. 6.

For the reasons just discussed, however, the model cannot mimic equally well the patterns in the pairs in which the PR lotteries are matched with increasing chances of winning £60. In this case, the extreme consequences in the context of the decision problem are the same for the $-bet and the two P-bets (£0 and £60), and therefore, the differences in subjective values are divided by the same amount for all bets, which rules out the inversion of the relative slopes of the curves observed in our data. This is shown in panel B of Fig. 6.

The model proposed by Blavatskyy is similar in some respects, but uses a different normalisation. For any {**p**, **q**} we can identify a greatest lower bound (GLB) **p**^**q**, which is the best lottery that is stochastically dominated by both **p** and **q**. The model then considers the subjective value differences between each lottery and that GLB—i.e. (V(**p**)–V(**p**^**q**)) and (V(**q**)–V(**p**^**q**))—and puts functions of these differences into a Luce choice formulation (Luce 1959). Thus, the probability that lottery **p** is chosen over lottery **q** can be written as:

$$\Pr(\mathbf{p}|\{\mathbf{p},\mathbf{q}\}) = \varphi(V(\mathbf{p}) - V(\mathbf{p}^\wedge q))/[\varphi(V(\mathbf{p}) - V(\mathbf{p}^\wedge q)) + \varphi(V(\mathbf{q}) - V(\mathbf{p}^\wedge q))] \qquad (5)$$

where φ is a non-decreasing function such that φ(0) = 0, and V(.) is the EU of the relevant lottery.[15]

In effect, this modelling strategy cuts across the BF assumption of independence between the noise associated with each V(.) because the same **q** interacts differently with different **p**'s when generating the GLB **p**^**q** to be applied in any particular choice. To see this, consider the two bets P2 = (£10, 0.65) and $ = (£40, 0.25) presented to the second subsample, each in conjunction with a **q** offering some sure amount X and then each in conjunction with a **q** offering some probability π of £60.

In the comparisons with the sure sums, the GLBs are P2^X = (X, 0.65) and $^X = (X, 0.25). Since the EU of the GLB for each {P2, X} choice is 2.6 times the EU of the GLB for each corresponding {$, X} choice, Expression (5) gives different values even when the EUs of the P-bet and the $-bet are equal. The general result is that Pr($ |{$, X}) falls more slowly as X rises than does Pr(P2 |{P2, X}), in much the same way that the dashed line in Fig. 2 is flatter than the solid line.

But now consider the comparisons with the L lotteries. The GLBs are P2^L = (£10, π) and $^L = (£40, π). Now the GLB in the $-bet cases is a multiple of the GLB in the P-bet cases, so that Expression (5) tends to change more slowly in the P-bet comparisons, with the result that Pr(P2 |{P2, L}) will tend to fall more slowly than Pr($ |{$, L}) as π rises, reversing the relationship. This is illustrated in panels C and D of Fig. 6. Here, we have conducted a calibration exercise in which we assume EUT at the core with a power utility function $U(x) = x^{0.8}$, and for simplicity we take φ to be the identity function. Such a parameterisation of this model can reproduce the qualitative patterns in our data.

Finally, we turn to the random preference (RP) approach. Instead of assuming that individuals have one true deterministic preference relation, to which an error term is attached, the RP approach models individuals as if their preferences consist of a number of deterministic preference functions, each corresponding with (slightly) different moods or states of mind. That is, RP allows that an individual's attitudes, perceptions and ways of processing decisions—and hence their responses—may vary somewhat from one occasion to another. To capture this, RP supposes that when facing a decision problem, the individual acts as if randomly drawing one of these functions and applying it to the decision problem as a whole. In this setting, the probability that lottery **p** is chosen over lottery **q** is given by the proportion of preference functions for which V(**p**) > V(**q**).

This approach is rather flexible, in principle, if one allows any possible distribution over any possible kind of preference function. However, we can put substantial restrictions on the allowable preference functions, and can still mimic the key features of our data quite closely. In our calibration, we assume preferences to be represented by a family of twenty power utility functions of the form that we have assumed for the other two models, $U(x) = x^r$, with r—the coefficient of constant relative risk aversion (CRRA)—ranging from 0.1 to 1.5 in equally spaced steps (median 0.8). As can be seen in panels E and F of Fig. 6, this is sufficient to reproduce the patterns found in our data that the BF model cannot explain.[16]

The intuition for this result is best seen if the problem is framed in terms of money and probability equivalents. At the most risk averse end of the individual's spectrum of utility functions, the riskier $-bet will be much less attractive than the safer P-bet, implying that the money equivalents for the $-bet will be much lower than the money equivalents for the P-bet. By contrast, at the most risk seeking end of the spectrum, the $-bet will be much more attractive than the P-bet and it will have much higher money equivalents. So, the distribution of money equivalents for the $-bet is wider than the corresponding distribution for the P-bet. The implication is that a given range of sure amounts will span a bigger proportion of the P-bet distribution than the $-bet distribution, leading to bigger changes in the frequencies with which the P-bet is chosen. This is reflected by the curves for the P-bets in Panel E of Fig. 6 being steeper than the curve for the $-bet.

But for probability equivalents, the relative riskiness of the lotteries produces the opposite results. At the most risk averse end of the spectrum, the more attractive P-bet requires higher probability equivalents than the $-bet, while at the most risk seeking end of the spectrum, the P-bet is less attractive and the probability equivalents are lower than for the $-bet. Hence for probability equivalents, it is the P-bet that has a wider distribution and this is reflected by the curves for the two P-bets being flatter than that for the $-bet in Panel F of Fig. 6.

4 Concluding remarks

Researchers in the field of decision-making under risk are becoming increasingly aware that the assumptions they make about the stochastic specification of the decision process can crucially affect the conclusions they are able to draw from their studies (e.g. Loomes 2005; Wilcox 2008; Hey et al. 2010). There have already been attempts at comparing the adequacy of alternative stochastic specifications (e.g. Loomes and Sugden 1998; Loomes et al. 2002; Wilcox 2008), which have highlighted some of the limitations of the BF model—in particular the over-prediction of violations of first-order stochastic dominance, which are actually quite rarely observed when transparent. In spite of these concerns, this model is still popular in applied work.

In this paper we have taken a different approach. Motivated by the findings of Butler and Loomes (2007), we have designed an experiment in which a key assumption of the BF model—that the error attached to the subjective value of a risky option is independent of the characteristics of all other options in the choice set—makes distinct and refutable predictions. Our results are incompatible with those predictions, adding to the evidence of the inadequacy of that model.

In the face of the evidence against the BF model, it is natural to start looking for alternatives. Using simple calibration exercises, we have identified just two extant approaches that are able to come to grips with the systematic patterns in our data. The approach proposed by Blavatskyy (2009, 2011) uses both lotteries in the construction of the greatest lower bound (or, equivalently, least upper bound), which is then used to compute their relative attractiveness and their respective probabilities of being chosen in a way which is consistent with our data. The random preference approach (Becker et al. 1963; Loomes and Sugden 1995) takes a rather different perspective, but by imposing some regularity upon the distribution of preferences—in our example, supposing a uniformly distributed family of CRRA utility functions ordered by the coefficient of relative risk aversion—we can also reproduce the key patterns in our data.

Our calibration exercises are meant to be simple illustrations of possible ways to overcome the limitations of the BF model. We do not have enough evidence to conclude in favour of either of the two stochastic specifications that are consistent with our findings. Indeed, there may be other yet-to-be-developed approaches which might turn out to be superior to both of them. Further work will be required to explore new approaches or discriminate between existing ones. Meanwhile, our results provide further evidence cautioning against the uncritical use of some form of BF specification, and suggesting the need to thoroughly investigate the properties of various extensions and/or alternatives to the 'standard' BF model.

Footnotes

[1]Although we shall sometimes refer to 'error', we are not suggesting that the variability is due to mistakes, in the sense of miscalculations or lapses of attention. Many experimental datasets may indeed contain cases of such mistakes (sometimes referred to as 'trembles') but our usage is in the econometric spirit of a disturbance term.

[2]In some of these applications, a BF-like error term is attached to certainty equivalents instead of subjective values.

[3]In cases in which estimation is done separately for each individual, the variance of the error term *is* allowed to vary across individuals (see, for example, Hey and Orme 1994).

[4]If the difference V(**p**)–V(**r**) is sufficiently large relative to the difference between σ^2_r and σ^2_p it is possible that the dashed line will lie to the left of the solid line for all V(**q**). The discussion that follows applies to those cases where the differences involved are such that the two lines cross.

[5]In early PR experiments, lotteries often involved small losses (e.g. Lichtenstein and Slovic 1971).

[6]For instance, one way of eliciting the money equivalent of a bet involved starting with a choice between the bet and a very low certainty (A$1) and asking the respondent to report whether she 'definitely' preferred the bet, 'probably' preferred the bet, probably preferred the sure sum, or definitely preferred the sure sum. The respondents almost always signified a definite preference for the bet in this choice. The sure sum was then changed by increments of A$1 and the question was repeated for each new value. As that sum was made progressively more attractive, most respondents moved from a definite preference for the bet to a probable preference for the bet, then to a probable preference for the sure amount, and eventually to a definite preference

for the sure sum. The interval of money equivalents over which the respondent is less than completely sure about her preference is given by the difference in switching values between a definite and a probable preference for the bet, and a probable and a definite preference for the sure amount. These switching points were not linked to incentives. Analogous intervals were also elicited for probability equivalents.

[7]Subjects face all pairings before any pair is repeated. The sequence of pairs is predetermined, separately for the two subsamples, and differs in each of the repetitions.

[8]As they move the cursor, some text appears underneath the bar. In the first quarter of each side, the text says that the chosen lottery is 'slightly better' than the alternative, in the second that it is 'better', then 'much better' and finally 'very much better'. In fact, each side of the bar maps into a 100-point scale (not visible to the subjects); when subjects confirm their decision, the value corresponding to the position of the cursor is recorded. The full text of the instructions is reported in the Appendix.

[9]There were a total of 100 tasks in the experiment. In addition to the 52 discussed in this paper, there were another 48 decision problems. These are described in Butler et al. (2012a).

[10]The strength of preference judgments, which are not linked to incentives, are reported elsewhere (Butler et al. 2012a, b).

[11]Given that all the lotteries we work with have just one non-zero consequence, from now on, we denote a lottery K which offers £x with probability p and nothing otherwise as K = (£x, p).

[12]For each choice involving the $-bet, we test the hypothesis that the proportion of subjects choosing it over the alternative is the same in both subsamples using the median choice. We only reject this hypothesis in one out of eight comparisons.

[13]As we have noted in the introduction, violations of first-order stochastic dominance are not ruled out by the BF model. However, irrespective of whether or not the BF model is the correct one, the frequencies of these violations are normally so low that they can be regarded as 'trembles' (see Loomes et al. 2002).

[14]Notice that extending the analysis to other transitive core models that allow for non-linear probability weighting, such as rank-dependent expected utility theory, cannot *in itself* do much to accommodate the patterns in our data. This is because any probability distortion would be the same for the P-bets and the $-bet in the two series of comparisons and could have no effect on the relative slopes of the curves we consider, which reflect the variability of the error terms.

[15]Blavatskyy takes EUT as the basis for V(.) and shows that, for binary choice, an equivalent expression for Pr(**p** |{**p**, **q**}) can be derived in terms of a least upper bound (LUB). Everything we say in terms of the GLB has a counterpart in terms of the LUB and no conclusions are altered by using the LUB rather than the GLB. The model can, in principle, be used with non-EU V(.) functions.

[16]Further details on these calibration exercises are available from the authors upon request.

Acknowledgements

Andrea Isoni and Graham Loomes acknowledge the financial support of the UK Economic and Social Research Council (grant no. RES-051-27-0248) and David Butler acknowledges the support of the Australian Research Council (grant: DP1095681) in this collaboration. We thank the Centre for Behavioural and Experimental Social Science at the University of East

Anglia for the resources and facilities used to carry out the experimental work reported here. Finally, we have benefitted from helpful comments and suggestions by an anonymous referee. The usual disclaimer applies.

Appendix

Experimental instructions

Introduction

In this session you will be offered a number of choices displayed in the form shown lower down this page. These are choices between different chances of different sums of money.

After you have made all your choices, just ONE of them will be picked at random to be run for real. Your ENTIRE PAYMENT will depend on how your choice in that one question works out, so please think carefully about every choice—any one of them could turn out to be the one on which your entire payment depends, and you will NOT be able to change your mind once each choice has been made.

The choice shown below is only an example and will not be run for real. Alternative A gives you a 55% chance of receiving £29 and a 45% chance of getting £3. Alternative B gives you an 85% chance of receiving £17 and a 15% chance of nothing.

If this were the choice that is being run for real, here's what would happen. We would see which alternative you had chosen. Then we would ask you to dip into a bag containing a hundred little discs, each with a different number on it, and you would pick one out at random.

If you had chosen A, you would be paid £29 if the number on the disc was between 1 and 55 inclusive; but if you chose A and the number turned out to be between 56 and 100 inclusive, you would get £3. On the other hand, if you had chosen B, you would receive £17 if the number on the disc was between 1 and 85 inclusive; but if it was between 86 and 100, you would get nothing.

	1-55	56-100
A	£29	£3
	55%	45%

	1-85	86-100
B	£17	0
	85%	15%

If you have a question, please ask.

If you are ready to continue, click here. [CONTINUE]

Checking for understanding

We have to make sure that everyone who takes part understands how things work. So please answer the check questions below by clicking on the answer you believe to be correct and then clicking on OK.

	1-55	56-100
A	£29	£3
	55%	45%

	1-85	86-100
B	£17	0
	85%	15%

Question 1. If the number on the randomly chosen disc is 73, how much will someone receive if they have chosen A?

[ ] £29 [ ] £17 [ ] £3 [ ] 0

[OK]

(check if answer is correct. If wrong, explain right answer and give chance to ask questions)

Question 2. If the number on the randomly chosen disc is 23, how much will someone receive if they have chosen B?

[ ] £29 [ ] £17 [ ] £3 [ ] 0

[OK]

(check if answer is correct. If wrong, explain right answer and give chance to ask questions)

Practice

There are no right or wrong answers—in each decision we ask you to make, we just want you to tell us what YOU personally prefer.

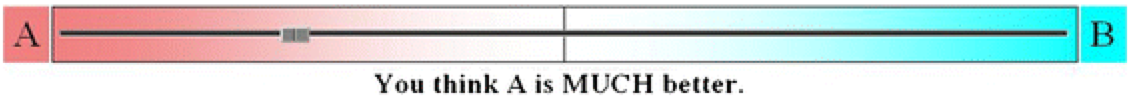
We also want you to tell us HOW MUCH BETTER you think one alternative is than the other. So each decision will be shown in the form below.

To tell us which alternative you choose AND how much better you think it is, put the cursor on the button in the middle of the bar below and move it either left (if you want to choose A) or right (if you want to choose B).

If you feel that both alternatives are almost equally good so that you think the one you are choosing is only SLIGHTLY better than the other one, just move the button a little way in the direction of your choice. However, if you think the one you are choosing is VERY MUCH BETTER than the other one, move the button a long way along the bar in the direction of your choice, possibly as far as the end if you feel very strongly indeed.

Once you have moved the button to the position that shows which alternative you choose and how much better you think it is, press OK. Then you will be asked to confirm your choice (or change it, if you change your mind) before moving to the next decision.

Try it now on this PRACTICE question.



[OK]

A	1-65	66-100
	£13	£7
	65%	35%

B	1-40	41-100
	£37	0
	40%	60%

(After OK was pressed, the following message appeared beneath the two prospects)

You have chosen **A**

To confirm this choice, click on Yes.

To change your decision, click on No and make your choice again.

[YES] [NO]

Checking for understanding

Just to be sure that everyone is understanding how things work, please answer the check questions underneath the display.

In this choice, you chose **A**

	1-65	66-100
A	£13	£7
	65%	35%

	1-40	41-100
B	£37	0
	40%	60%

Question 1: If the number on the randomly-picked disc were 19, how much would you receive?

[ ] £37 [ ] £13 [ ] £7 [ ] 0

[OK]

(check if answer is correct. If wrong, explain right answer and give chance to ask questions)

Question 2: If the number on the randomly-picked disc were 81, how much would you receive?

[ ] £37 [ ] £13 [ ] £7 [ ] 0

[OK]

(check if answer is correct. If wrong, explain right answer and give chance to ask questions)

End of practice

That is the end of the Practice. From now on, every decision you make could be the one that YOUR ENTIRE PAYMENT depends on.

Remember, after you have made your final decision, one of your choices will be selected at random (they are all equally likely to be picked, as you will see when the time comes). We will then see what you chose and your choice will be played out FOR REAL.

Whatever amount you receive will be paid straight away, in cash. You will be asked to sign a receipt for it.

If you have any more questions, please raise your hand and someone will come to your desk.

If you are happy to proceed to the real choices, click on OK below.

[OK]

References

Becker, G. M., DeGroot, M. H., & Marschak, J. (1963). Stochastic models of choice behaviour. *Behavioral Science, 8*, 41–55.

Blavatskyy, P. R. (2006). Violations of betweenness or random errors? *Economics Letters, 91*, 34–38.

Blavatskyy, P. R. (2009). Preference reversals and probabilistic decisions. *Journal of Risk and Uncertainty, 39*, 237–250.

Blavatskyy, P. R. (2011). A model of probabilistic choice satisfying first-order stochastic dominance. *Management Science*, *57*, 542–548.

Blavatskyy, P. R., & Pogrebna, G. (2010). Models of stochastic choice and decision theories: Why both are important for analyzing decisions. *Journal of Applied Econometrics, 25*, 963–986.

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

Butler, D., & Loomes, G. (2007). Imprecision as an account of the preference reversal phenomenon. *American Economic Review*, *92*, 277–298.

Butler, D., Isoni, A., Loomes, G., & Tsutsui, K. (2012a). *Looking beyond choices: Investigating the sensitivity and validity of strength of preference measures*. Working paper.

Butler, D., Isoni, A., & Loomes, G. (2012b). *What lies beneath? systematic intransitive tendencies revealed by strength of preference measures*. Working paper.

Camerer, C. F. (1995). Individual decision-making. In J. Kagel & A. Roth (Eds.), *The handbook of experimental economics*. Princeton: Princeton University Press.

Epper, T., Fehr-Duda, H., & Bruhin, A. (2011). Viewing the future through a warped lens: Why uncertainty generates hyperbolic discounting. *Journal of Risk and Uncertainty, 43*, 169–203.

Fechner, G. T. (1860, 1966). *Elements of psychophysics*. New York: Holt, Rinehart and Winston.

Harless, D. W., & Camerer, C. F. (1994). The predictive utility of generalized expected utility theories. *Econometrica*, *62*, 1251–1289.

Harrison, G. W., List, J. A., & Towe, C. A. (2007). Naturally occurring preferences and exogenous laboratory experiments: A case study of risk aversion. *Econometrica, 75*, 433–458.

Hey, J. D., & Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica*, *62*, 1291–1326.

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

Lichtenstein, S., & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology, 89*, 46–55.

Loomes, G. (2005). Modelling the stochastic component of behaviour in experiments: Some issues for the interpretation of data. *Experimental Economics, 8,* 301–323.

Loomes, G., & Sugden, R. (1995). Incorporating a stochastic element into decision theories. *European Economic Review, 39*, 641–648.

Loomes, G., & Sugden, R. (1998). Testing different stochastic specifications of risky choice. *Economica, 65*, 581–598.

Loomes, G., Moffatt, P. G., & Sugden, R. (2002). A microeconometric test of alternative theories of risky choice. *Journal of Risk and Uncertainty, 24*, 103–130.

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. New York: John Wiley.

Luce, R. D., & Suppes, P. (1965). Preferences, utility, and subjective probability. In R. D. Luce et al. (Eds.), *Handbook of mathematical psychology, 3*. New York: John Wiley.

Mosteller, F., & Nogee, P. (1951). An experimental measurement of utility. *Journal of Political Economy, 59*, 371–404.

Pratt, J. W. (1964). Risk aversion in the small and in the large. *Econometrica, 32*, 122–136.

Seidl, C. (2002). Preference reversal. *Journal of Economic Surveys, 16*, 621–655.

Starmer, C. (2000). Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature, 38*, 332–382.

von Gaudecker, H., van Soest, A., & Wengström, E. (2011). Heterogeneity in risky choice behavior in a broad population. *American Economic Review, 101*, 664–694.

von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior*. Princeton: Princeton University Press.

Wilcox, N. T. (2008). Stochastic models for discrete choice under risk: A critical primer and econometric comparison. In J. C. Cox & G. W. Harrison (Eds.), *Research in experimental economics, 12*. Bingley: Emerald Group Publishing Limited.

Wilcox, N. T. (2011). 'Stochastically more risk averse': A contextual theory of stochastic discrete choice under risk. *Journal of Econometrics, 162*, 89–104.

Fig. 1 Pr(**p**|{**p**, **q**}) **and** Pr(**r**|{**r**, **q**}) **as a function of** V(**q**) **when** $\sigma^2_r = \sigma^2_p$, **and** V(**r**) < V(**p**). The *solid line* depicts how the probability that lottery **p** is chosen from the pair {**p**, **q**} varies as a function of the subjective value assigned to **q** according to the individual's core preferences, V(**q**). When V(**q**) = V(**p**), the individual is stochastically indifferent and chooses each lottery with equal probability. The *dashed line* shows how the probability that lottery **r** is chosen from the pair {**r**, **q**} varies as a function of V(**q**). The two curves are parallel because it is assumed that the variance around V(**r**), σ^2_r, is the same as the variance around V(**p**), σ^2_p

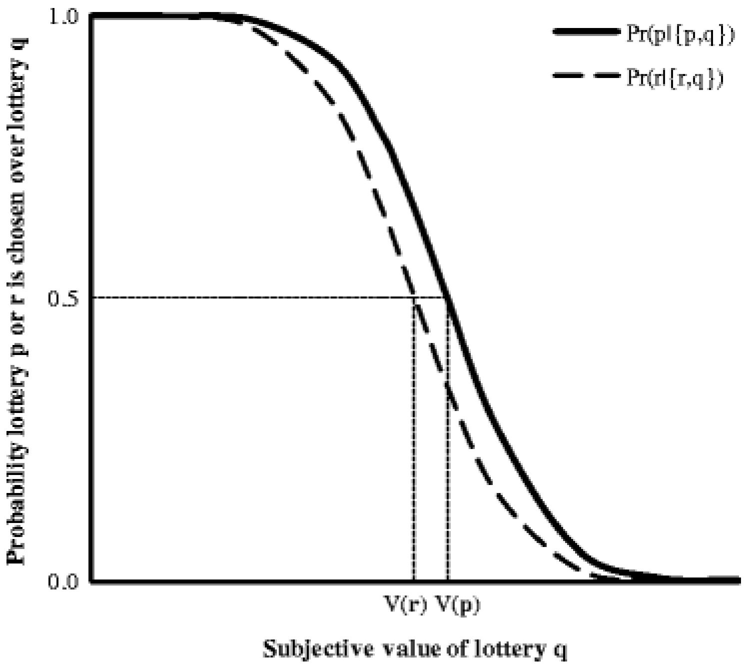


Fig. 2 Pr(**p**|{**p**, **q**}) **and** Pr(**r**|{**r**, **q**}) **as a function of** V(**q**) **when** $\sigma^2_r > \sigma^2_p$, **and** V(**r**) < V(**p**). The *solid line* is just as in Fig. 1. The *dashed line* shows how the probability that lottery **r** is chosen from the pair {**r**, **q**} varies as a function of V(**q**). The two curves may now cross because $\sigma^2_r > \sigma^2_p$ entails the c.d.f. being flatter for **r** than for **p**. The *down arrow* denoted vd_{low} shows that, for a $\mathbf{q}_{low}$ to the left of the crossing point, **p** is chosen more frequently over **q** than is **r**. The *up arrow* denoted by vd_{high} shows that, for a $\mathbf{q}_{high}$ to the right of the intersection point, **p** is chosen less frequently over **q** than is **r**

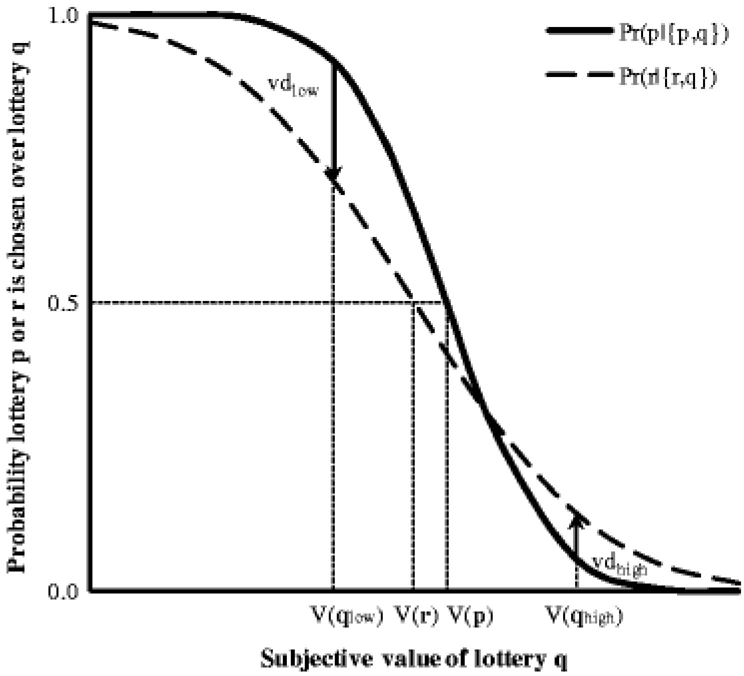
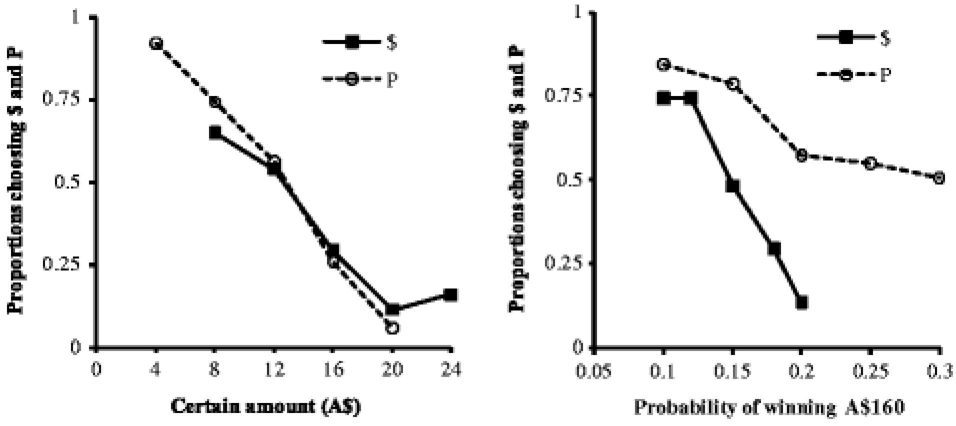


Fig. 3 Choice proportions from Butler and Loomes (2007). The *solid lines* depict the probability with which $ = (A$80, 0.25) is chosen over a series of certain amounts (Panel **A**) or a series of different chances of winning A$160 (Panel **B**). The *dashed lines* depict the probability with which P = (A$24, 0.7) is chosen over a series of certain amounts (Panel **A**) or a series of different chances of winning A$160 (Panel **B**)



(A) $-bet and P-bet against certainties

(B) $-bet and P-bet against chances of A$160

Fig. 4 An example of a binary choice. The two lotteries in each pair are labelled **A** and **B**, and presented as strips in which the amounts to be won are shown in *boxes* for which the width is proportional to the chances of winning, which are shown as percentages underneath the amounts. Lotteries are played out by drawing one of 100 numbers from an opaque bag. The correspondence between numbers and prizes can be read above the amounts. Choices are made by dragging the slider to the left (to choose **A**), or to the right (to choose **B**). The slider starts at the middle of the bar and allows participants to express the strength of their preference for their chosen lottery

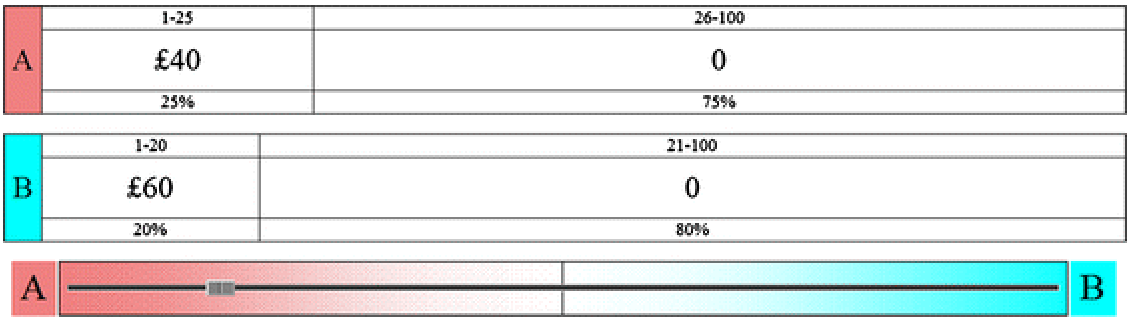
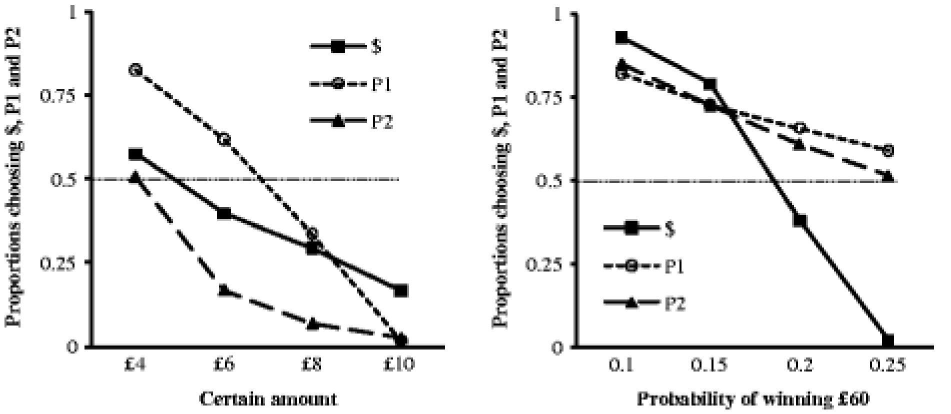


Fig. 5 Choice proportions using aggregate data. Panel **A** presents the choice proportions for $ = (£40, 0.25), P1 = (£10, 0.9) and P2 = (£10, 0.65) when each is matched against four certain amounts (£4, £6, £8 and £10). Panel **B** presents the choice proportions for $, P1 and P2 when each is matched against the four L lotteries offering different chances (0.1, 0.15, 0.2 and 0.25) of winning £60. Contrary to the independence assumption of the BF model, the curve for the $-bet is flatter than those for the two P-bets in one case, and steeper in the other



(A) $-bet and P-bets against certainties

(B) $-bet and P-bets against chances of £60

Fig. 6 Calibration of alternative stochastic EU models. For Wilcox's (2011) model, $U(x) = x^{0.8}$ and $\lambda = 5$. For Blavatskyy's (2011) model, $U(x) = x^{0.8}$ and $\varphi(z) = z$. For Random Preference model, $U(x) = x^r$, with r drawn from 20 equally spaced risk aversion coefficients ranging from 0.1 to 1.5 (median 0.8). Predicted choice proportions on the vertical axis

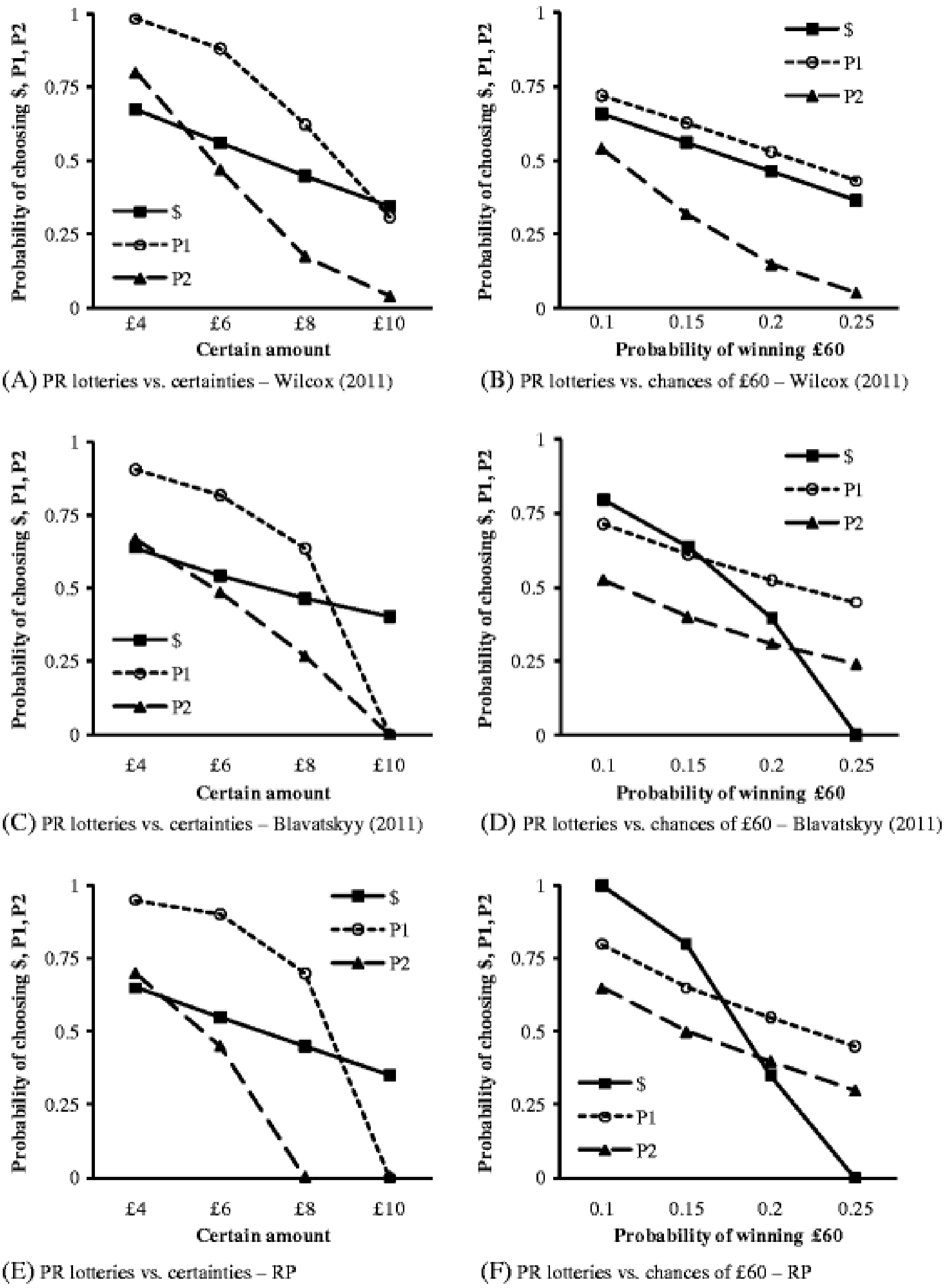


Table 1 Experimental results

PR Lottery–Observation	Common certainties				Common chances of £60			
	£4	**£6**	**£8**	**£10**	**0.1**	**0.15**	**0.2**	**0.25**
$–1st	0.529	0.268	0.283	0.138	0.957	0.768	0.239	0.029
$–2nd	0.609	0.457	0.304	0.167	0.913	0.783	0.449	0.014
$–3rd	0.587	0.464	0.297	0.203	0.920	0.819	0.449	0.014
$–Aggr.	0.575	0.396	0.295	0.169	0.930	0.790	0.379	0.019
$–Median	0.580	0.449	0.297	0.159	0.935	0.804	0.370	0.007
P1–1st	0.913	0.638	0.261	0.043	0.855	0.754	0.681	0.638
P1–2nd	0.754	0.623	0.391	0.014	0.797	0.725	0.580	0.551
P1–3rd	0.812	0.594	0.362	0.000	0.812	0.696	0.710	0.580
P1–Aggr.	0.826	0.618	0.338	0.019	0.821	0.725	0.657	0.589
P1–Median	0.841	0.609	0.348	0.014	0.826	0.783	0.681	0.609
P2–1st	0.507	0.145	0.072	0.043	0.826	0.696	0.580	0.536
P2–2nd	0.522	0.174	0.058	0.014	0.841	0.754	0.609	0.478
P2–3rd	0.493	0.188	0.072	0.014	0.884	0.725	0.638	0.522
P2–Aggr.	0.507	0.169	0.068	0.024	0.850	0.725	0.609	0.512
P2–Median	0.507	0.116	0.058	0.014	0.855	0.768	0.580	0.536

$ = (£40, 0.25); P1 = (£10, 0.90); P2 = (£10, 0.65). Observations for $-bet pooled.

Each choice was presented three times. The choice proportions for each presentation are denoted by 1st, 2nd and 3rd.

The proportions obtained by aggregating across the three repetitions are denoted by 'Aggr'. We also report proportions using the median choice for each participant.

Table 2 Individual-level tendencies

$\Delta dv = (chP_{low} – ch\$_{low}) – (chP_{high} – ch\$_{high})$	P1 = (£10, 0.9)		P2 = (£10, 0.65)	
	£4 to £8	**L10 to L20**	**£4 to £8**	**L10 to L20**
−3	1	7	0	10
−2	4	15	6	14
−1	8	20	11	27
0	23	22	20	14
1	11	5	14	1
2	14	0	14	3
3	7	0	4	0
4	1	0	0	0
Mean	0.65	−0.9565217	0.45	−1.130435
Median	0	−1	0	−1
Cert. > Prob.		50		51
Cert. = Prob.		10		8
Cert. < Prob.		9		10

chP_{low} = number of times P is chosen against the low sure amount (£4) or the low probability of winning £60 (L10). $ch\$_{low}$ = number of times $ is chosen against the low sure amount (£4) or the low probability of winning £60 (L10). chP_{high} = number of times P is chosen against the high sure amount (£8) or the high probability of winning £60 (L20). $ch\$_{high}$ = number of times $ is chosen against the high sure amount (£8) or the high probability of winning £60 (L20).

The top eight rows report the whole distributions. The means and medians are in the ninth and tenth row.

The bottom three rows report the number of individuals for whom the change in vertical distance when the P and $ bets are matched with certainties (Cert.) is greater than, equal to, or less than that for L lotteries (Prob.)