Modelling Noise and Imprecision in Individual Decisions

Graham Loomes, University of Warwick, Coventry, CV4 7AL, UK

Jose Luis Pinto-Prades, Universidad Pablo de Olivade, Sevilla, Spain

Jose Maria Abellan-Perpinan, Universidad de Murcia, Spain

Eva Rodriguez-Miguez, Universidade de Vigo, Spain

Abstract

When individuals take part in decision experiments, their answers are typically subject to some degree of noise / error / imprecision. There are different ways of modelling this stochastic element in the data, and the interpretation of the data can be altered radically, depending on the assumptions made about the stochastic specification. This paper presents the results of an experiment which gathered data of a kind that has until now been in short supply. These data strongly suggest that the 'usual' (Fechnerian) assumptions about errors are inappropriate for individual decision experiments. Moreover, they provide striking evidence that core preferences display systematic departures from transitivity which cannot be attributed to any 'error' story.

February 2010

Keywords: Error Imprecision Preferences Transitivity

Introduction

Most theories of decision making under risk are expressed in deterministic form, as if an individual's preferences are precise, stable and consistent with some 'core' of axioms or well-defined components (such as a utility function and/or a probability weighting function). Interpreted literally, the implication is that if a particular individual is asked to reveal his preference by choosing between two alternatives, G and H, he will (except in the very special case of indifference) always make the same choice every time the two are offered under the same conditions.

However, this is in sharp contrast with extensive experimental evidence going back more than 50 years which suggests that, when asked to choose between pairs of lotteries on two or more occasions within a short period of time, many individuals do not always choose the same alternative. Indeed, it is not uncommon to find 15%-30% 'switching rates' in experimental repeated binary choice data (see, for example, Mosteller and Nogee, 1951; Luce (1962); Starmer and Sugden, 1989; Camerer, 1989; Hey and Orme, 1994; Ballinger and Wilcox, 1997; Loomes and Sugden, 1998).

This raises two questions. How should we understand and model the observed variability in the data? And what are the implications for the way(s) in which we might specify and test different 'core' theories?

Most researchers in this area respond to these questions using some variant of a 'Fechner' model. Fechner models assume that each individual gives a subjective value (SV) to an object, but that her perception of this value is subject to some degree of 'noise'. That is, if we ask the jth individual to value some object G a number of times on separated occasions, she will give a set of responses distributed around some central tendency. Thus on any particular occasion, the value of G perceived and reported by that individual can be denoted by $SV_{Gj} + \varepsilon_{Gj}$ where SV_{Gj} represents the 'core' subjective value of G to that individual – this being determined by whatever theory best accounts for the way that individual combines payoffs and probabilities – while ε_{Gj} signifies some independent random deviation from the core value on that particular occasion.

Applying this approach to the situation where the individual is asked to choose between two lotteries, G and H, the model entails the choice being made according to which lottery is perceived to have the higher value at the moment when the choice is made. On those occasions when $SV_{Gj} + \varepsilon_{Gj}$ is greater than $SV_{Hj} + \varepsilon_{Hj}$, G is chosen; but, depending on the difference between SV_{Gj} and SV_{Hj} and on the values that ε_{Gj} and ε_{Hj}

happen to take, there may be occasions when $SV_{Gj} + \varepsilon_{Gj}$ is less than $SV_{Hj} + \varepsilon_{Hj}$, in which case H is chosen. Thus choice becomes probabilistic, with the probability of G being chosen over H, $\Pr(G \succ H)$, given by:

$$\Pr(G \succ H) = \Pr(SV_{Gj} + \varepsilon_{Gj} > SV_{Hj} + \varepsilon_{Hj}) \qquad (1)$$

For many economists and econometricians, such a formulation is in line with the tradition of taking deterministic core functional forms and simply adding some 'error' term which has well-established, analytically tractable properties. Not surprisingly, then, Fechner models have featured in some form or other in many of the econometric analyses of experimental binary choice data. For example, Hey and Orme (1994) examined the performance of a number of alternative core theories on the assumption that ε is symmetrical around zero and has constant variance. Buschena and Zilberman (2000) allowed the variance of ε to be correlated with some measure of the complexity of the lotteries being evaluated. And Blavatskyy (2007) considered the implications of truncating the distribution of ε in particular ways. If the broad Fechner framework is regarded as the appropriate way of modelling the stochastic component in risky decisions, all of these variants – and others, perhaps – are potentially admissible: the best way of specifying the distribution of ε is then principally a matter of empirical investigation[1].

However, the Fechner approach is not the only way in which a stochastic component can be incorporated into decision modelling. Becker, DeGroot and Marschak (1963) proposed an alternative *random preference* (RP) approach. Rather than supposing that each person is characterised by just one core preference function, RP allows that individuals' perceptions, moods, attitudes and judgments may fluctuate to some extent from one moment to another. So on one occasion an individual may tend to feel more optimistic, impulsive, risk seeking, etc., while on a different occasion he might focus more on the downside, exhibiting greater caution and risk aversion. Thus it is as if an individual's judgmental apparatus comprises of some continuum of states of mind, with each state of mind represented by a (slightly) different preference function. For example, someone who is essentially a von Neumann-Morgenstern expected utility

[1] Different assumptions about the distribution of ε can have very different implications: see Loomes (2005) or Bardsley et al (2009, Chapter 7) for examples and discussion.

(EU) maximiser may, within each state of mind, always weight the utilities of different payoffs by their respective probabilities; but different states of mind may be characterised by different utility functions, sometimes more concave, sometimes less concave or even on occasions convex, reflecting a degree of variability in risk attitude from one occasion to another. Thus a particular individual's preferences might be modelled as a distribution over some *set* of functions, one for each state of mind, with any one of these having some probability of being the current state of mind at the time a particular judgment is made.

In a choice between two lotteries, G and H, there may be some functions in the set which would favour G over H and others which would favour H over G. The probability that an individual chooses G can thus be modelled as the probability of drawing at random from the set a preference function which evaluates G more highly than H.

To illustrate how fundamentally the RP approach can differ from the Fechner approach, consider the case where the individual is, at core, an EU maximiser and where he is presented with a choice between H, which offers a 50-50 chance of 20€ or 0, and G, which offers a 50-50 chance of 20€ or 5c – that is, G first-order stochastically dominates H, although the difference in their expected values is relatively small.

Under the RP approach, the individual's state of mind – that is, the particular vN-M utility function he applies to the choice – may vary from moment to moment; but *every one* of these functions respects first-order stochastic dominance (FOSD), so that whatever his state of mind at the moment of choice, he *always* prefers G to H: hence $\Pr(G \succ H) = 1$.

Under the Fechner approach, SV_{Gj} is greater than SV_{Hj}. But the difference is small, and may be dwarfed by the variances of ε_{Gj} and ε_{Hj}, with the result that on a substantial minority of occasions, ε_{Hj} may exceed ε_{Gj} to a degree which more than offsets the difference between SV_{Gj} and SV_{Hj}. On these occasions, $SV_{Gj} + \varepsilon_{Gj} < SV_{Hj} + \varepsilon_{Hj}$ and the dominated lottery H is chosen. In such cases, then, the Fechner model entails a substantial probability (less than 0.5, but conceivably not *much* less) of observing a violation of FOSD.

In this respect, such evidence as there is comes much closer to RP than to Fechner. For example, Loomes and Sugden (1998) asked 92 respondents to make 45 binary choices, with each choice presented twice. In 40 of these pairs, neither lottery dominated the other, and out of 3,680 (92 x 40) instances, there were 676 cases (18.4%) where the choice on the second occasion was different from that on the first. In the other 5 pairs, one lottery dominated the other, usually by offering a 0.05 higher chance of the best payoff and a 0.05 lower chance of the worst payoff. In these pairs, out of a total of 920 observations (92 respondents each making 5 choices on two occasions), dominance was violated in just 13 cases – a rate of less than 1.5%. Under the Fechner approach, there is no reason to expect the rate to be so much lower in cases involving dominance: indeed, since the differences in expected values were mostly smaller in the pairs involving dominance, the rate might, if anything, have been expected to be higher in those pairs. Such a low rate can be more readily reconciled with RP (which entails a 0% rate) supplemented by the occasional lapse of attention or 'trembling hand' kind of error[2].

Of course, it might be argued that if the only shortcoming of the Fechner approach were its overprediction of violations of FOSD, that could be finessed by assuming (as Kahneman and Tversky's (1979) Prospect Theory does) some prior editing phase which identifies and eliminates any transparently dominated options.

However, we wished to investigate the robustness and appropriateness of the Fechner and RP approaches in other scenarios which did not involve dominance but which required trade-offs between the countervailing attractions of different alternatives. To that end, we conducted an experiment that would allow us to explore other possible differences between those two approaches.

In the next section we set out the key features of the experimental design and the main issues we sought to examine. In section 3, we report the results. These results raise serious doubts about the appropriateness of Fechner models in this area and suggest that RP may provide a more suitable framework for modelling stochastic decision processes. The final section discusses the potentially far-reaching and radical implications of our findings.

2. Basic Principles of the Design and the Issues to be Investigated

At the centre of the experimental design were six lotteries, as listed in Table 1. In that table, each lottery is shown in the form: higher payoff, probability of higher

[2] Loomes, Moffatt and Sugden (2002) argue that some allowance for such 'trembles' may be a useful adjunct to both Fechner and RP models, but they suggest that the prevalence of such 'pure' errors is low.

payoff; lower payoff, probability of lower payoff. In all cases, the payoffs were in Euros.

Table 1

Label	Description	EV	Label	Description
A	84, 0.25; 0, 0.75	21.00	D	60, 0.25; 8, 0.75
B	36, 0.55; 0, 0.45	19.80	E	36, 0.40; 9, 0.60
C	22, 0.8; 0, 0.2	17.60	F	20, 0.8; 8, 0.2

These six lotteries constitute three triples, {A, B, C} and {D, E, F}. Within each triple, safer lotteries offer lower expected values (EVs). D, E and F respectively offer the same EVs as A, B and C, but each involves a smaller spread than its {A, B, C} counterpart. For all six lotteries, the individuals in our sample were asked to undertake three types of task, as follows:

BC. Nine different binary choices (BC) were constructed and each choice was presented to every respondent on six different occasions (separated from one another by being interspersed with the other types of task described below). Those choices were: {A, B}, {B, C}, {A, C}, {D, E}, {E, F}, {D, F}, {A, D}, {B, E} and {C, F}.

ME. For each lottery, each respondent was asked on six different (dispersed) occasions to state the sure sum of money that would make them indifferent between that sum and the lottery in question. These were the 'money equivalent' (ME) questions.

PE. Every lottery was worth less than 120€, so for each lottery, each respondent was asked on six different (dispersed) occasions to state the probability p of receiving 120€ and the 1-p chance of receiving 0 that would make them indifferent between playing the lottery in question and playing that 'probability equivalent' (PE) lottery.

The data from these tasks allow us to examine two respects in which the Fechner and RP approaches are liable to differ substantially. These relate to: 1) the relationship between the distributions of MEs and PEs; and 2) the relationship between equivalences and binary choices. In the next two subsections we expand upon each of these in turn.

2.1: The relationship between the distributions of MEs and PEs

We start with the Fechner approach. Within that framework, standard deviations for the various lotteries should follow the same pattern for MEs and PEs.

To see why, consider first MEs. For each individual, any sure amount of money M can be regarded as a degenerate lottery with its own distribution of ε. Without being specific about how the variance of ε behaves for such degenerate lotteries[3], we can expect that for sufficiently low values of M, the overlap between an individual's $SV_M + \varepsilon_M$ and his $SV_G + \varepsilon_G$ is negligible, so that he would judge G to be better than those sure sums with a probability so close to 1 that he would be extremely unlikely to identify any of those sums as money equivalents for the lottery G. But as M is progressively increased over some intermediate range where $SV_M + \varepsilon_M$ and $SV_G + \varepsilon_G$ overlap, it becomes increasingly likely that M will be judged at least as good as G. Eventually M can be expected to become sufficiently high for these be no consequential overlap at the other end of the $SV_G + \varepsilon_G$ distribution, so that the individual would be extremely unlikely to judge G to be as good as these high values.

If the ε_Gs and ε_Ms are distributed symmetrically around zero means, the probability of M being judged at least as good as G reaches 50% at the point where $SV_M = SV_G$. Thus, so long as the preference functions are not highly nonlinear over the relevant range, it might be a reasonable approximation to suppose that the distribution of MEs is roughly symmetrical around a mean/median value located where $SV_M = SV_G$, with the variance of this distribution reflecting the joint distribution of ε_M and ε_G. If nonlinearities result in substantial asymmetries in the distribution of MEs – something we can examine – there might be an argument for taking the median as the better measure of central tendency[4].

Second, if different lotteries are associated with markedly different distributions of ε, we might expect to see this reflected in the distributions of their MEs. For example, suppose that the variance of ε_H is greater than the variance of ε_G. Under Fechnerian assumptions, for every M the joint distribution of ε_M and ε_H will have greater variance than the joint distribution of ε_M and ε_G, so that we might expect the variance of MEs for H to be greater than the variance of MEs for G. So if this were the appropriate error model, it might allow us to gain insights into the features of lotteries that are associated with different variances of ε.

Now consider the PEs. Let us denote the 'yardstick' lottery (offering 120€ with probability p and 0 with probability 1-p) by Y. The distributions of the ε_Y's associated

---

[3] One possibility, consistent with much psychophysical work, is that the variance increases somewhat as the magnitude of M increases. Another possibility, advocated by Blavatskyy (2007) – although without any cited empirical foundation – is that the variance for any sure sum is zero.

[4] Under EUT, for example, it could be that u(.) is markedly concave so that a symmetrical distribution of perceived expected utilities may map to a distribution of MEs with a longer right tail: in which case, the median ME will correspond more closely with the midpoint of the distribution of perceived EUs.

with the values of p that span the relevant range may be different from the distributions of the ε_M's over the corresponding range. So the PEs for G *might* be distributed rather differently than the MEs for G, reflecting the possibility that the joint distributions of ε_G and ε_Y could be rather different from the corresponding joint distributions of ε_G and ε_M.

Nevertheless, we should still expect to observe PEs exhibiting the same regularities as just outlined for the MEs: whatever similarities or differences in the standard deviations of the ME distributions are observed between G and H, we should expect (broadly) the same similarities or differences to be manifested in the standard deviations of the PE distributions for those same lotteries. For example, if the standard deviation of an individual's MEs were to decline progressively from A to B to C, we should, under the Fechner model, suppose this to reflect a tendency for the variance of ε_C to be less than the variance of ε_B which, in turn, is less than the variance of ε_A. But if that is the case, the joint distributions of ε_Y with each of the lottery error terms should vary correspondingly, so that we should also expect some progressive decline in the standard deviations of that individual's PEs as we move from A to B to C.

How does this compare with the implication of RP? Under RP, it is quite straightforward to model equivalence judgments. Suppose an individual is asked on a particular occasion to state a sure payoff M such that he is indifferent between the certainty of that payoff and playing out lottery G. RP models this as if one of that individual's preference functions is picked at random from his set of such functions and the individual then states the M corresponding with that function. Different functions are liable to entail different values of M, so that the distribution of those functions generates a distribution of ME_Gs for that individual.

Exactly the same reasoning applies to PE. For each preference function in the set that characterises that individual, there will be some probability of the yardstick payoff which will make the individual indifferent between G and Y. Denoting that probability by p_G, the probability distribution over the set of preference functions maps to some distribution of p_Gs. This applies to any core theory which entails the existence of PEs (and MEs) for any and all lotteries: the likelihood that a particular PE is stated is simply the likelihood that, at the moment when the decision is made, an individual is in a state of mind corresponding to a preference function that entails a mapping between the lottery and that PE; and likewise for MEs.

However, in order to go further and generate some testable hypotheses that can be contrasted with those emanating from the Fechner framework, we need to place some restriction on the distribution of any individual's preference functions. A fairly permissive restriction in keeping with conventional wisdom would be to suppose that each individual's core theory respects transitivity and that all of the preference functions in his set can be ordered according to some measure of risk attitude.

Under these assumptions, suppose that the decision maker is asked to undertake choice and equivalence tasks involving two binary lotteries, G and H, where both the variance and expected value of G is greater than for H. For some preference functions at the risk-seeking/risk-neutral/less risk-averse end of the distribution, the higher EV of G is sufficient for $G \succ H$, whereas some functions at the more risk-averse end of the spectrum entail $H \succ G$. Let the proportion of functions that entail $G \succ H$ be denoted by α: then α is the probability of observing $G \succ H$ on any occasion when the individual is asked to make a BC, while we can expect to observe $H \succ G$ with probability $1-\alpha$.

Now suppose we draw a representative sample of an individual's MEs for each lottery[5]. Were we to happen to draw the function where G ~ H, it would give the same ME for G as for H: call this value M*. All less risk-averse functions would give MEs for both lotteries that are above M*, but for each of those functions, $G \succ H$, so that the ME_G would be higher than the corresponding ME_H. On the other hand, for all functions involving greater risk aversion than the one where G ~ H, the MEs would all be less than M*; and for each of *those* functions, $ME_G < ME_H$. Under these conditions, a representative sample of that individual's MEs would exhibit a greater standard deviation for ME_G than for ME_H.

If $\alpha > 0.5$, the median function would entail $G \succ H$, and we should expect the median ME_G to be greater than the median ME_H (and if the underlying distribution were not very far from symmetrical and if the sample sizes were adequate, we might also expect the means to reflect the same inequality). If $\alpha < 0.5$, the median function would entail $H \succ G$ so that the opposite inequality would hold between medians (and probably means). But of course, the earlier conclusion about the direction of inequality of the standard deviations of the MEs would be unaffected.

[5] Notice that the assumption being made here is that the same probability distribution over an individual's utility functions applies to any type of task. This is not the only assumption that might be made, but it is the working assumption we shall operate with at present. In footnote 12 we shall briefly discuss a different possibility.

What about the PEs? Consider again the function where G ~ H. This entails some probability p* such that $PE_G = PE_H$. For all functions involving less risk aversion, the probability of the yardstick payoff would be lower than p* for both lotteries; but since G $\succ$ H for each such function, the corresponding PE_G would be higher than its PE_H counterpart. In other words, for the subsample of functions from the less risk-averse portion of the distribution, the PE_Hs would tend to take even lower values than the PE_Gs. On the other hand, for any function from the more risk averse part of the distribution, the probabilities of the yardstick payoffs would be greater than p*; and since these functions entail H $\succ$ G, the PE_Hs would here tend to take even higher values than the PE_Gs. Taking the distribution as a whole, then, we could expect the standard deviation of a sample of an individual's PE_Hs to be greater than the standard deviation of a comparable sample of that individual's PE_Gs.

This relationship between the standard deviations for the PEs is in the *opposite* direction to our expectation for the MEs and provides a sharp contrast with the implications of Fechner models, which suppose that the variances of ε_G and ε_H are primarily determined by the characteristics of the lotteries and that any differences between them will tend to manifest themselves in much *the same* way via MEs as via PEs. This is a contrast our experimental data will allow us to examine.

2.2: The relationship between equivalences and binary choices.

Within the terms of the Fechner framework, the degrees of overlap between any two sets of ME (alternatively, PE) responses should broadly correlate with the frequencies of choice in the repeated BC tasks.

If the relationship between the $SV_G + \varepsilon_G$ distribution and the $SV_H + \varepsilon_H$ distribution is such that (say) G is chosen over H significantly more than 50% of the time, we should, at the very least, expect the median (and probably the mean) ME for G to be higher than the median (mean) ME for H. We should expect the same with PE.

However, we may be able to go further than simply expecting the median/mean MEs and PEs to be ordered in the same way as each other and in line with the majority of choices between any two lotteries. If the distributions of MEs and PEs can be thought of as proxies for the SV + ε distributions, the relationships between those distributions for any two lotteries might allow us to proxy choice probabilities. For example, if the distribution of ME_Gs and ME_Hs were such that there is a 60% chance that an ME_G

drawn at random from its distribution would be greater than an ME_H drawn at random from its distribution, one might expect this to be indicative of $SV_G + \varepsilon_G$ being greater than $SV_H + \varepsilon_H$ about 60% of the time, so that under Fechnerian assumptions G would be chosen in roughly 60% of choice repetitions. Similarly, we should expect the choice proportions to be broadly in line with the probability that a randomly-drawn PE_G will be greater than an independently sampled PE_H. Because of the involvement of the ε_M's and ε_Y's, this correspondence may not be exact; but under Fechner assumptions, one would expect to find the probabilities inferred from equivalences and those observed in repeated choice being not too greatly out of alignment[6].

By contrast, the RP approach allows the possibility of very considerable disparities between the extent to which equivalences overlap and the pattern of binary choice. This was illustrated earlier in the case where G first-order stochastically dominated H, but only by a small amount, so that $ME_G > ME_H$ and $PE_G > PE_H$ just over 50% of the time but G $\succ$ H in direct binary choice on 100% of occasions (except, perhaps, for 'trembles'). However, in response to the suggestion that FOSD is a rather special and unusual situation which might be dealt with by some prior editing procedure, it is quite easy to construct cases which do not involve FOSD but where RP would allow considerable disparities between the choices we observe and those we might infer from the overlaps of equivalences.

For example, consider the choice between our lottery C = (22, 0.8; 0, 0.2) and our lottery F = (20, 0.8; 8, 0.2). Both have the same expected value of 17.60, so we could expect the two distributions of MEs to have a considerable degree of overlap; and likewise for the two distributions of PEs. But suppose (as is commonly done) that most individuals are predominantly risk averse, which in RP terms means that their preferences are characterised by sets of utility functions where the (great) majority are concave. Every concave function will entail F $\succ$ C, so that we might expect to find F chosen very much more often than the overlapping of the equivalence distributions would suggest. A similar argument applies to the {B, E} and {A, D} pairs. Substantial disparities of this kind would be compatible with RP, but would be contrary to Fechnerian models.

[6] If the utilities of sure sums of money are perceived with no noise/error, as assumed in Blavatskyy (2007), the correspondence between distributions of MEs and distributions of perceived SVs is exact; if the variance of ε is positive and liable to change with the magnitudes of the utilities (and perhaps with other characteristics of the risky lotteries), the correspondence is more approximate.

3. The Experiment – Implementation and Results

Every participant was required to attend two sessions, several days apart, during a three-week period in March/April 2008. Each session followed the same format. Having signed in and read the instructions (see Appendix 1), each participant answered 63 questions per session, organised in three successive 'phases', each consisting of the same 21 questions, as follows: 6 MEs, one for each of the six lotteries; 6 PEs, one for each of the six lotteries; and 9 binary choices. All MEs and PEs were elicited using an iterative choice format (see Appendix 1 for details and examples of displays) in order to make them as procedurally similar to BCs as possible. We gave no feedback until all tasks had been completed, at which point we paid each respondent on the basis of playing out one of those decisions picked at random at the very end of the experiment. Standard incentive mechanisms were used (again, see Appendix 1 for details).

A total of 274 individuals completed the full set of decisions[7]. In the way equivalences were elicited, we deliberately did not 'force' either ME or PE responses to respect stochastic dominance because we wanted to see how people behaved if unconstrained. In the course of the two sessions, there were 54 occasions when it was possible for each respondent to violate FOSD, either by stating an ME equal to or higher than the high payoff of the lottery being valued or equal to or less than the lower payoff (in the cases of D, E and F), or else by stating a PE at least as high as the probability of the high payoff in the {A, B, C} lotteries. The tables in the rest of this paper are based on the unedited responses of all 274 participants, including some responses that violate FOSD. However, in order to anticipate any concerns that such responses may be 'driving' the patterns in our data, we have also computed all tables using only the responses from individuals who *never* violated FOSD. These tables are presented in Appendix 2. They show that none of the patterns we report, nor the conclusions drawn from them, are materially altered when we apply even the fiercest exclusion criterion: indeed, if anything, the conclusions come through even more powerfully, since by excluding a number of outliers we reduce the standard errors used in various of the statistical tests and increase the corresponding significance levels.

[7] 45 others came to the first session but did not attend the second session within the time limit we imposed.

3.1: Results relating to the relationship between the distributions of MEs and PEs

In order to examine the relationship between the distributions of MEs and PEs and the contrasts between Fechner and RP outlined in subsection 2.1 above, for each respondent and for each lottery we computed the mean, median and standard deviation of the six ME responses – labelled, respectively, 'meanME', 'medME' and 'sdevME' – and the corresponding 'meanPE', 'medPE' and 'sdevPE' for each set of six PE responses. Table 2 reports the sample averages for these variables, plus the sample medians of the standard deviations, with those standard deviations in the middle rows for easier comparison between MEs and PEs.

Table 2: Key Statistics for Money Equivalents and Probability Equivalents

Lottery	A	B	C	D	E	F
Average meanME	20.93	17.48	15.09	20.93	18.82	15.50
Average medME	20.60	17.28	15.10	20.73	18.76	15.43
Average sdevME	4.43	3.17	2.51	4.08	2.97	2.24
Median sdevME	3.40	2.64	2.28	3.38	2.39	1.73
Average sdevPE	3.29	5.48	7.32	6.42	6.20	7.33
Median sdevPE	2.16	4.33	5.70	4.08	4.91	6.30
Average meanPE	18.33	25.27	28.45	23.79	25.66	29.37
Average medPE	18.05	24.88	27.94	23.24	25.23	29.05

This table enables us to see whether the trends in standard deviations follow the same pattern for MEs as for PEs, as the Fechner framework would suggest, or whether they move in opposite directions, as we might expect under RP.

For MEs, we find that as we move from A to B to C, and also as we move from D to E to F, standard deviations reduce to an extent that is highly significant ($p < 0.001$) in every pairwise comparison within each triple. By contrast, the standard deviations of PE responses show a strong tendency to change in the *opposite* direction: of the six binary comparisons of sdevPEs within the two triples, only the difference between D and E is in the same direction as for MEs (although insignificantly so), while the increase from D to F is significant at the 1% level and the other four binary differences are significant at the 0.1% level[8]. It is hard to see how such strong opposite trends in the standard deviations can be reconciled with *any* variant of the Fechner approach – and certainly not any in the existing literature of which we are aware.

3.1: Results relating to the relationship between equivalences and binary choices

In order to examine the relationship between equivalences and binary choices and the contrasts between Fechner and RP outlined in subsection 2.2 above, we collated the BC data as follows. For each respondent, we observed the number of times out of 6 repetitions that he/she chose the riskier lottery – that is, the one with the greater spread (always labelled alphabetically earlier than the safer lottery) – within each pair.

Table 3: Binary Choice Distributions

{R, S}	Frequency of Choice of Riskier Lottery							
Pair	6	5	4	3	2	1	0	Total R:S
{A, B}	25	19	20	31	33	41	105	525: 1119
{B, C}	20	17	25	29	37	50	96	516: 1128
{A, C}	36	20	16	28	31	41	102	567:1077
{D, E}	78	42	55	30	25	24	20	1062: 582
{E, F}	55	43	44	33	39	25	35	923: 721
{D, F}	80	31	44	36	23	28	32	993: 651
{A, D}	17	16	15	14	21	55	136	381: 1263
{B, E}	9	7	10	10	16	42	180	233: 1411
{C, F}	2	1	--	3	4	21	243	55: 1589

For each pair, Table 3 shows the 274 individuals categorised accordingly: that is, the column headed '6' shows how many respondents chose the riskier lottery on all 6 occasions; the column headed '5' shows how many chose the riskier lottery on five occasions and chose the safer lottery just once; and so on through to those in the column headed '0' who never chose the riskier lottery but chose the safer alternative every time

[8] As can be seen in the corresponding table in Appendix 2, removing cases where FOSD is violated has the effect of reducing all standard deviations, while bringing the direction of change of the average sdevPE from D to E in line with the other five pairwise comparisons – although this difference remains insignificant. It also increases the significance level of the D to F difference from 1% to 0.1%.

they faced that binary choice. The total number of times the riskier (R) and safer (S) lotteries were chosen is shown in the far right hand column.

It is immediately apparent that there is considerable variability at the individual level, while at the same time there are some very definite and quite intuitive trends. For each of the three {A, B, C} pairs, the overall majority of choices clearly favour the safer alternative, as shown in the far-right column. Even so, for each pair only about half of the respondents make the same choice consistently on all six occasions.

Turning to the {D, E, F} pairs, the effect of raising the minimum payoffs to 8€ or 9€ and reducing the spreads of these lotteries relative to their {A, B, C} counterparts is to cause the majority of choices now to favour the higher-EV alternatives in each case. Here, though, there is somewhat less within-person consistency, with only about 40% of respondents making the same choice on all six occasions.

Finally, when the EVs are equalised, as in the last three pairs, there are very substantial majorities favouring the safer alternatives, most strikingly in the {C, F} choice. So despite the clear evidence of variability in people's choices, there are also many signs of systematic tendencies underlying their behaviour.

How do the patterns of choice compare with the overlaps between equivalences? We focus on the three pairs {A, D}, {B, E} and {C, F} where the alternatives within each pair shared the same EV and differed only in terms of their spreads.

For each individual and for each pair of lotteries, we compared each individual's six MEs for one lottery with each of her six MEs for the other lottery. We recorded the number of comparisons in which the ME of the riskier lottery (ME_R) was strictly higher than the ME of the safer lottery (ME_S), the number of occasions when $ME_R = ME_S$, and the number of times when $ME_R < ME_S$. Since there were 36 comparisons for each of 274 respondents, the total number of comparisons per pair of lotteries is 9864. The distributions for each pair of lotteries is shown in Table 4.

Table 4: Comparing ME_Rs with ME_Ss

Lottery Pair	$ME_R > ME_S$	$ME_R = ME_S$	$ME_R < ME_S$
A vs D	4231 (42.9%)	928 (9.4%)	4705 (47.7%)
B vs E	3149 (31.9%)	976 (9.9%)	5739 (58.2%)
C vs F	4081 (41.4%)	1147 (11.6%)	4636 (47.0%)

It is immediately apparent that the BC distributions are very different from those implied by the overlaps of equivalences. On the basis of the ME responses, A should have been preferred to D *at least* 42.9% of the time in direct choice (that minimum figure being based on the rather extreme assumption that *all* cases where $ME_A = ME_D$ are interpreted as favouring D in direct choice). But in fact A was chosen in just over 23% of direct BCs. Similarly, the ME data suggest that B should be preferred to E on at least 31.9% of occasions, whereas the actual proportion was less than half of that (14.2%). The {C, F} case provides an even more striking contrast: whereas $ME_C > ME_F$ in more than 41% of comparisons, C is chosen only 55 times out of 1644 – a rate of less than 3.5%. These disparities are entirely in keeping with RP, but are incompatible with any Fechner formulation of which we are aware.

3.3: Results relating to the assumption of transitivity

The propositions and implications set out in subsections 2.1 and 2.2 were derived on the basis of fairly general assumptions about core preferences that might apply to many non-EU models as well as to EU: in particular, that core preference functions are transitive and can, for any individual, be ordered according to some measure of risk attitude. On that basis, we considered some contrasting implications of Fechner and RP approaches for the distributions of ME, PE and BC responses and the relationships between them. And on this basis, the evidence from the experiment strongly and consistently appeared to favour RP rather than any form of Fechner error term.

However, the data in Tables 2 and 3 give grounds for questioning the assumption of transitivity. In Table 2, the mean and median MEs order the lotteries in the first triple A $\succ$ B $\succ$ C and give the ordering D $\succ$ E $\succ$ F in the second triple, while the mean and median PEs produce exactly the opposite orderings within each triple. Meanwhile, Table 3 shows aggregate patterns of binary choices that do not fit either with MEs or with PEs: for each pair in the {A, B, C} triple, the majority choices favour the safer options, which is in line with the PEs but is contrary to the MEs; while for every pair in the {D, E, F} triple, the majority choices favour the riskier options, which tallies with MEs but runs counter to PEs.

Of course, those tables report aggregate data, whereas an examination of transitivity really requires individual-level analysis. In particular, as argued in Section 2, if individuals' underlying preference functions can be ordered according to some measure of risk attitude, we should expect individuals' median responses to provide insights into the nature of the functions at the centre of those distributions. If such functions entailed transitivity, we should expect this to be reflected at the level of the individual by the correspondence between median MEs, median PEs and majority choices.

The relevant individual-level analysis is reported in Table 5, which categorises all 274 respondents according to their median ME, median PE and majority BC responses to each pair, with R and S referring, respectively, to the riskier and safer lotteries in each pairing.

Table 5: Conjunctions of Median Responses

Direction of Median			**Lottery Pairs**					
ME	**PE**	**PC**	**{A, B}**	**{B, C}**	**{A, C}**	**{D, E}**	**{E, F}**	**{D, F}**
R ≥ S	R ≥ S	R ≥ S	36	48	57	80	73	84
R ≥ S	R ≥ S	R < S	20	46	12	12	29	13
R > S	R < S	R ≥ S	46	25	38	73	87	92
R > S	R < S	R < S	102	70	105	28	54	49
R < S	R > S	R > S	0	4	1	11	5	6
R < S	R > S	R ≤ S	0	16	1	8	5	5
R ≤ S	R ≤ S	R > S	6	5	1	33	5	3
R ≤ S	R ≤ S	R ≤ S	64	60	59	29	16	22

So, for example, the top cell in the {A, B} column shows that 36 of the 274 individuals had a median ME_A at least as high as their median ME_B *and* had a median

PE_A at least as high as their median PE_B *and* chose A over B on at least 3 of the six occasions they were presented with that choice. Such behaviour displays the kind of consistency that a transitive core theory would entail. The 20 in the next cell down favoured A both in terms of MEs and PEs but chose B on at least four of the six BC repetitions; while the 46 in the third cell chose A on at least 3 occasions and had median ME_A's strictly higher than their median ME_B's, but their median PE_A's were strictly lower than their median PE_B's. And so on.

By summing the numbers in the top and bottom rows, we can see how many individuals were weakly consistent with some transitive core preference for each pair. The emphasis shifts from the bottom row to the top row as we move from {A, B, C} to {D, E, F} but the total is fairly stable, always lying between 89 and 116 i.e. between 32.5% and 42.5% of the sample.

Thus for most pairs, more than 60% of the sample violated transitivity in one way or another. However, the ways in which they did so do not appear to be randomly distributed. On the contrary, they exhibit certain systematic patterns, as follows.

First, we observe the analogue to the classic preference reversal phenomenon where people choose one option but place a higher money equivalent on the other. In the literature – see Seidl (2002) – there is a clear asymmetry whereby it is relatively common to observe people placing a higher money value on the riskier option but choosing the safer option in the BC task (in our terms, median ME_R > median ME_S together with majority $S \succ R$) but it is relatively rare to observe people valuing the safer option more highly while choosing the riskier option (i.e. median ME_S > median ME_R together with $R \succ S$). In fact, taking the pairs in the left-to-right order of the columns in Table 5, the ratios we observe are 122:6, 116:9, 117:2, 40:44, 83:10 and 62:9: that is, with one exception, very strongly exhibiting the classic preference reversal asymmetry, especially among the {A, B, C} pairs where the safer options were more often chosen[9].

Although probability equivalents have been much less often studied, Butler and Loomes (2007) reported the opposite asymmetry when PEs and choices were compared. For the pair of lotteries they investigated, they found that instances where individuals chose the safer option but placed a higher PE on the riskier option were outnumbered by

[9] The corresponding table in Appendix 2, which excludes cases where a lottery is overvalued to the extent that the stated ME is greater than the high payoff, shows fewer reversals of both kinds: the corresponding ratios are 69:4, 70:6, 68:2, 24:26, 50:3 and 40:5. Thus even when all violations of FOSD are excluded, the asymmetry remains just as pronounced.

the opposite combination of choosing the riskier option while placing a higher PE on the safer one. For the six pairs in Table 5, the analogous ratios using majority choices and median PEs are: 13:35, 54:20, 11:25, 15:87, 29:72 and 12:74; thus, with one exception, these ratios show the same direction of asymmetry reported in Butler and Loomes (2007), with those asymmetries being much more pronounced for the {D, E, F} pairs where the riskier options were more often chosen[10].

Finally, the most striking and comprehensive asymmetry of all emerges from the conjunction of MEs and PEs. If we compare the numbers of individuals for whom median $ME_R > ME_S$ but median $PE_R < PE_S$ with those for whom median $ME_R < ME_S$ but median $PE_R > PE_S$, we obtain the following ratios: 148:0, 95:20, 143:2, 101:19, 141:10 and 141:11[11].

Remembering that these data are based on medians and majority responses – that is, they do not depend on single and possibly aberrant responses – we cannot see any way within the modelling framework outlined above that these patterns can be reconciled with a model which assumes that the great majority of individuals' decisions in such tasks reflect an essentially transitive core.

4. Discussion

The data presented in subsection 3.3 constitute powerful evidence against transitivity. This presents us with a Duhem-Quine problem: our basis for distinguishing between Fechner and RP involved the auxiliary assumption that any core theory was transitive (and in the case of RP, the additional assumption that preference functions could be ordered by risk attitude). If we have reason to doubt the assumption about the transitivity of core preferences, can we continue to be so confident that the Fechner framework is inferior to RP?

We think we can. Indeed, if core preferences are so often intransitive, that may constitute a further argument for doubting the appropriateness of applying the Fechner framework to experimental data about equivalences and choices between lotteries. To see why, consider what is involved in any core theory that allows systematic intransitivities. Intrinsic to such a theory is the idea that the evaluation of any lottery is liable to vary systematically from one context / choice set / decision task to another, so

[10] After excluding cases where FOSD was violated, the corresponding ratios are 7:24, 38:12, 6:17, 7:64, 20:32 and 8:36.

[11] From the Appendix 2 data, the corresponding ratios are: 91:0, 53:14, 86:1, 70:8, 74:6 and 80:6.

that G may be evaluated more favourably than H against sure amounts of money while H may be evaluated more favourably than G in a direct choice between the two and/or against some yardstick lottery. In other words, the assumption underpinning the Fechner approach – that the perceived SV of an object to an individual is purely a matter of how the characteristics of that object interact with the evaluation apparatus of the individual – does not hold. Trying to graft some form of independent Fechner error onto core preferences which allow systematic intransitivities as a result of contextual interactions would appear to involve a fundamental conceptual mismatch.

By contrast, the holistic processing of a decision task is entirely in keeping with the spirit of the RP model. The notion of a 'state of mind' entails both (all) alternatives being processed together and on the same basis on any given occasion. The effect of imposing transitivity is not to rule out such processing but rather to require that the results of evaluating two or more alternatives in conjunction with one another is indistinguishable from evaluating them separately and/or in conjunction with any/all other sets of options. Core theories which dispense with transitivity typically involve specifying the nature of interactions between alternatives which lead to systematic variations in the way the evaluation of a prospect is affected by the parameters of the other prospects in a particular set. Clearly, such interactions entail at least some degree of joint processing of the kind intrinsic to the RP approach. An RP specification of a non-transitive core theory is therefore very natural, and simply involves the extent of certain interactions varying from one occasion to another[12].

It is not our intention to say much more here about the detailed nature of an intransitive core theory that might fit the data[13]. Rather, the main focus of the present paper is upon the appropriate specification of the variability in most of our participants'

---

[12] However, it has occurred to us that there is a somewhat different approach that *might* reconcile the data with an RP formulation of some transitive core theory, as follows. Suppose that an individual's preferences are represented by some distribution of (say) von Neumann-Morgenstern u(.) functions, but that instead of sampling randomly from the same distribution for all types of task, the nature of the task biases the sampling in some way(s). In order to produce the patterns we have observed, it would need to be the case that the ME task prompts respondents to sample more heavily from the more risk-seeking/less risk-averse end of the distribution, while the PE task results in oversampling from the more risk averse end of the distribution, with binary choices perhaps being based on a sample somewhere between those other two. This kind of explanation would move us away from the more formal decision theoretic framework that underpins our analysis and towards something more in the 'heuristics and biases' (Kahneman, Slovic and Tversky, 1982) tradition. We put such a possibility 'on the table' as something that may merit future investigation, although we do not pursue it further in this paper.

[13] We *can* say that regret theory does not appear to fit the bill – in Butler and Loomes (2007) it was shown that regret theory is at odds with the form of PE-BC reversal found there and replicated in our data. One of the authors has proposed a model which does appear capable of accommodating that form of reversal alongside the classic ME-BC phenomenon – see Loomes (2010) – but it will require a much broader set of experiments to test more adequately the credentials of that model.

ME, PE and BC responses. Taken as a whole, our evidence strongly suggests that Fechner specifications are simply inappropriate.

The implications of such a conclusion are radical and potentially far-reaching. First, it raises serious doubts about much of the work to date that has used Fechner models to try to fit preference functions and to judge the relative merits of EU against other 'core' theories. If the whole Fechner approach is fundamentally inappropriate for these data, any estimates generated on the basis of such mis-specified error models and any inferences drawn from them must be regarded as questionable.

Second, the use of Fechner models has not been restricted to the analysis of data from individual decision experiments. As discussed in Loomes (2005) and Bardsley et al. (2009, Chapter 7), the 'quantal response equilibrium' (QRE) concept, developed by McKelvey and Palfrey (1995, 1998) and applied to numerous datasets generated by experimental games, is also an essentially Fechnerian model. If the Fechner approach is the wrong way of modelling the stochastic component of individual behaviour in the face of 'games against nature', it may also be the wrong way of modelling the stochastic component in individuals' behaviour when they are playing games against other individuals; and this may cast doubt on the robustness of QRE-based ways of fitting the data from experimental games and the inferences drawn from doing so.

Third, essentially the same assumptions underpin a much wider body of empirical and theoretical 'discrete choice' research (see Manski, 2001): if the model is unsound in the context of individual decisions about simple lotteries, how confident can we be about its suitability in many other areas where 'stated preference' methods have been used to guide private and public decision making?

Of course, it would be premature to discard a large body of existing literature on the basis of a single experimental study, no matter how striking the results of this study appear to be. Further work is clearly required in order to establish the robustness of our findings and explore the extent of their applicability. However, if such further work confirms our key findings and shows that they carry over into strategic behaviour and into other areas of preference elicitation, the implications are fundamental: techniques and results predicated upon Fechnerian assumptions may no longer be viable in these fields and we shall need in future to formulate hypotheses, conduct statistical tests, fit core functional forms and derive estimates of parameters in ways consistent with RP specifications of the stochastic nature of people's judgments and decisions.

References

Ballinger, T. and Wilcox, N., (1997). Decisions, Error and Heterogeneity, *Economic Journal*, **107**, 1090-1105.

Bardsley, N., Cubitt, R., Loomes, G., Moffatt, P., Starmer, C. And Sugden, R., (2009). *Experimental Economics: Rethinking the Rules*. Princeton: Princeton University Press.

Becker, G., DeGroot, M. and Marschak, J., (1963). Stochastic Models of Choice Behavior, *Behavioral Science*, **8**, 41-55.

Blavatskyy, P., (2007). Stochastic Expected Utility Theory, *Journal of Risk and Uncertainty*, **34**, 259-86.

Buschena, D. and Zilberman, D., (2000). Generalized Expected Utility, Heteroscedastic Error, and Path Dependence in Risky Choice, *Journal of Risk and Uncertainty*, **20**, 67-88.

Butler, D. and Loomes, G., (2007). Imprecision as an Account of the Preference Reversal Phenomenon, *American Economic Review*, **97**, 277-97.

Camerer, C., (1989). An Experimental Test of Several Generalized Utility Theories, *Journal of Risk and Uncertainty*, **2**, 61-104.

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

Hey, J. and Orme, C., (1994). Investigating Generalizations of Expected Utility Theory Using Experimental Data, *Econometrica*, **62**, 1291-1326.

Kahneman, D. and Tversky, A., (1979). Prospect Theory: An Analysis of Decisions Under Risk, *Econometrica*, **47**, 313-327.

Kahneman, D., Slovic, P. and Tversky, A., (1982). *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

Loomes, G., (2005). Modelling the Stochastic Component of Behaviour in Experiments: Some Issues for the Interpretation of Data, *Experimental Economics*, **8**, 301-23.

Loomes, G., (2010). Modelling Choice and Valuation in Decision Experiments, forthcoming in *Psychological Review*.

Loomes, G., Moffatt, P. and Sugden, R., (2002). A Microeconometric Test of Alternative Stochastic Theories of Risky Choice, *Journal of Risk and Uncertainty*, **24**, 103-130.

Loomes, G. and Sugden, R., (1998). Testing Different Stochastic Specifications of Risky Choice, *Economica*, **65**, 581-98.

Luce, R.D., (1962). Psychological Studies of Risky Decision Making. In *Social Sciences Approaches to Business Behavior*, Strother, G. B. (ed.). Homewood: Dorsey Press.

Manski, C., (2001). Daniel McFadden and the Econometric Analysis of Discrete Choice, *Scandinavian Journal of Economics*, **103**, 217-29.

McKelvey, R. and Palfrey, T., (1995). Quantal Response Equilibria for Normal Form Games, *Games and Economic Behavior*, **10**, 6-38.

McKelvey, R. and Palfrey, T., 1998). Quantal Response Equilibria for Extensive Form Games, *Experimental Economics*, **1**, 9-41.

Mosteller, F. And Nogee, P., (1951). An Experimental Measurement of Utility, *Journal of Political Economy*, **59**, 371-404.

Seidl, C., (2000). Preference Reversal, *Journal of Economic Surveys*, **6**, 621-655.

Starmer, C. and Sugden, R., (1989). Violations of the Independence Axiom in Common Ratio Problems: An Experimental Test of Some Competing Hypotheses, *Annals of Operations Research*, **19**, 79-102.

Appendix 1: Overview of the Experiment

Subjects and design

Participants in the experiment were students at three different Spanish universities: Vigo, Pablo de Olavide (Seville) and Murcia. The total number of people recruited was 319 (103 in Vigo, 144 in Murcia and 72 in Seville). Recruitment of participants took place during the 2 weeks before the experiment started: signs were posted in researchers' faculties and they also went to the classrooms to explain the aim of the experiment briefly and encourage participation.

The experiment was computer-based. Each participant had to attend two experimental sessions separated by at least 1 week. 45 subjects did not show up for the second session leaving the total sample as 274. There were two interviewers present during the group sessions to help subjects with any problems.

The questionnaire

The questionnaire was divided into three stages. In the first stage subjects were asked to enter their names, age, and gender. This request was for ensuring that responses of the same individual in the two sessions were correctly linked. They were then told that the experiment aimed to investigate how people make a series of choices between two options. It was explained that, in addition to a €5 'show-up' fee, there would also be a payment based on their decisions: at the end of the second session, one of their decisions would be retrieved and played out for real money. Because the payment depended on just one decision, participants were advised that it was in their interests to make each choice in a way that most accurately reflected their true preferences.

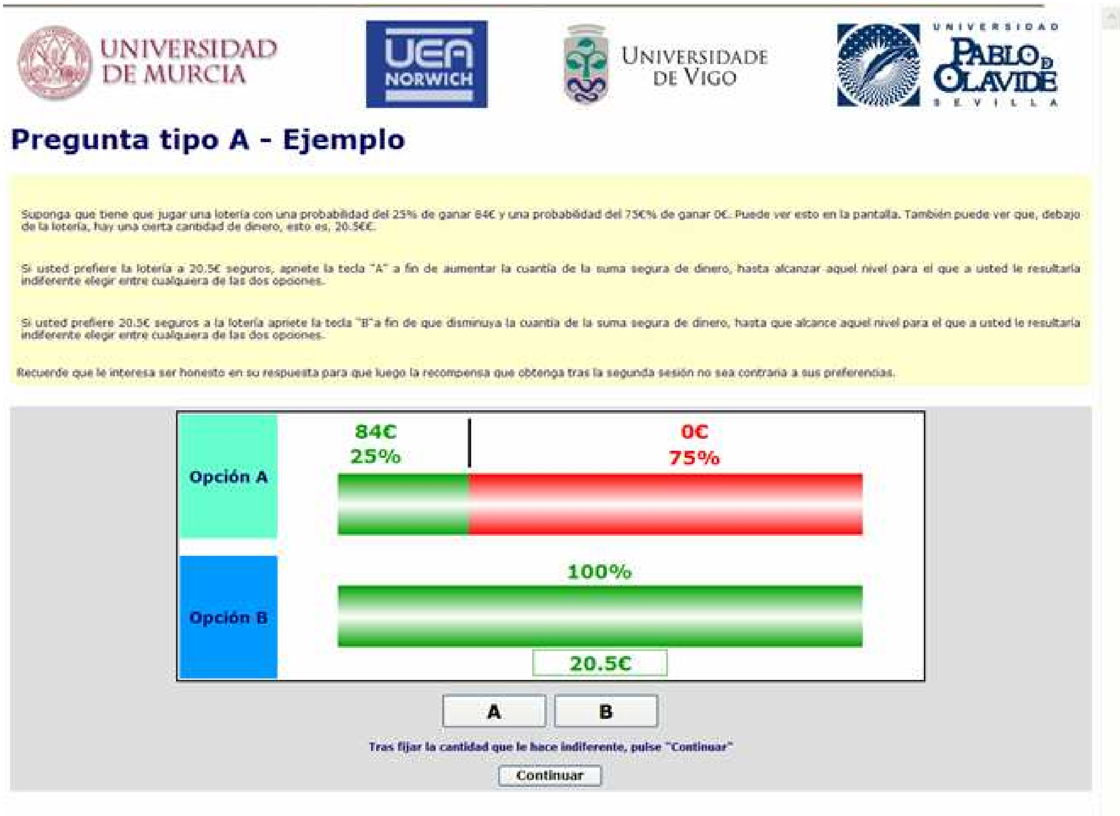
In the second stage, subjects were presented with three practice questions. Each of these questions illustrated the different types of question that participants would see in the third stage. The instructions for each type of question were displayed on the computer screen and they were also read aloud by the researchers. After being given an opportunity to clarify anything they were uncertain about, subjects were invited to proceed to the third stage.

The final stage consisted of nine sequences of questions, grouped in three blocks, 21 questions each. These questions were the same for the three blocks. The order in which the questions were administered within each block was as follows: first, there were 6 Money Equivalence (ME) questions; next, 6 Probability Equivalence (PE) questions; and finally, 9 Binary Choices (BC). Therefore each participant repeated this set of 21 tasks three times within each session – so, six times over the two sessions.

The Three Types of Question

Each ME question elicited the amount of money $€X_{ME}$ that made a subject indifferent between $€X_{ME}$ for certain and a lottery giving $€X_1$ with probability p and $€X_2$ with probability $(1-p)$. An example of the kind of display used is shown in Figure A1.

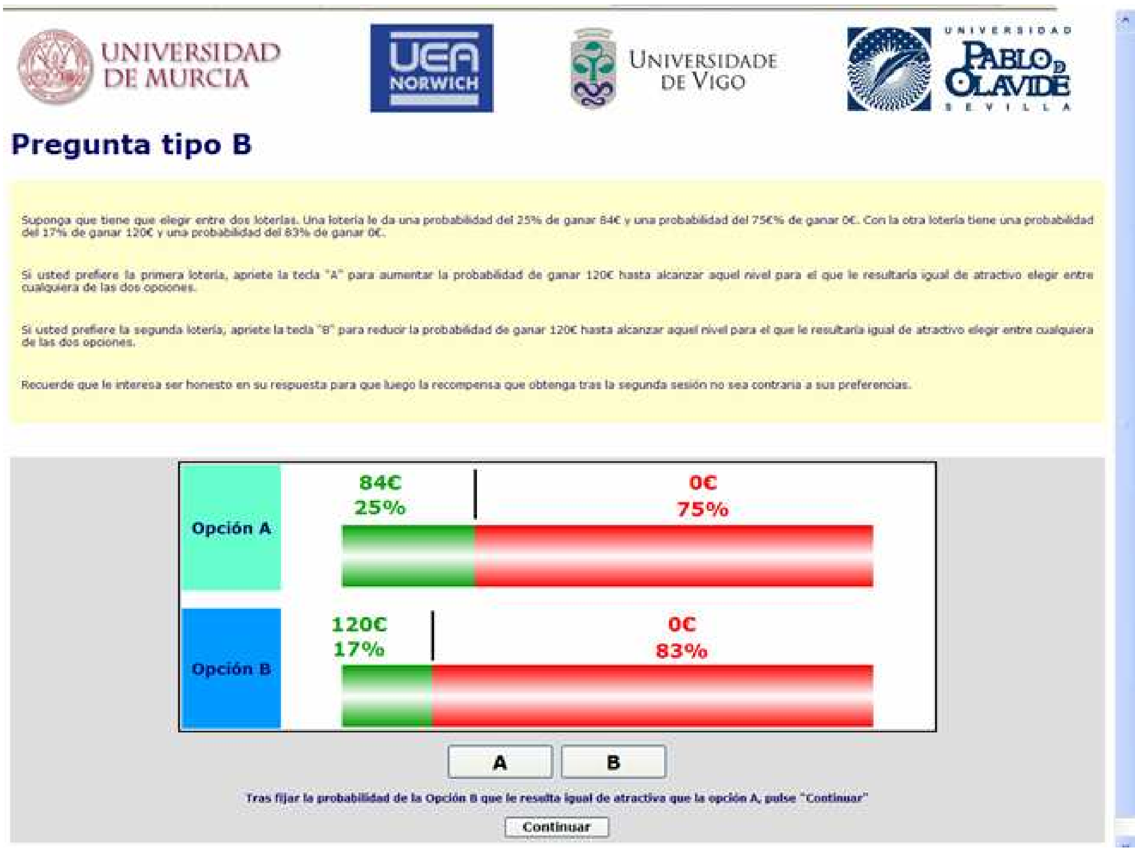
Figure A1: Screenshot of a Money Equivalence question



For each alternative, the bar was divided in proportion to the probability attached to the relevant payoff. For Option A (the lottery) the chance of winning the higher outcome (€84 in the example) was always coloured in green. The lower outcome (€0 in the example) was in red. As Option B only offered one sure amount, the entire bar was in green. Participants were asked to press "A" or "B" buttons until they considered both options equally attractive in terms of their preferences. Whenever the "A" button was pressed, the sure sum of money was increased, making Option B more desirable. The reverse occurred when subjects pressed the "B" button. Once subjects felt that they were indifferent between the two options, they registered this and moved on to the next question by pressing the "Continue" button.

Each PE question elicited the probability q that made the subject indifferent between a particular lottery and an alternative lottery giving *€120* with probability q and *€0* with probability *(1-q)*. Figure A2 shows an example of this type of question.

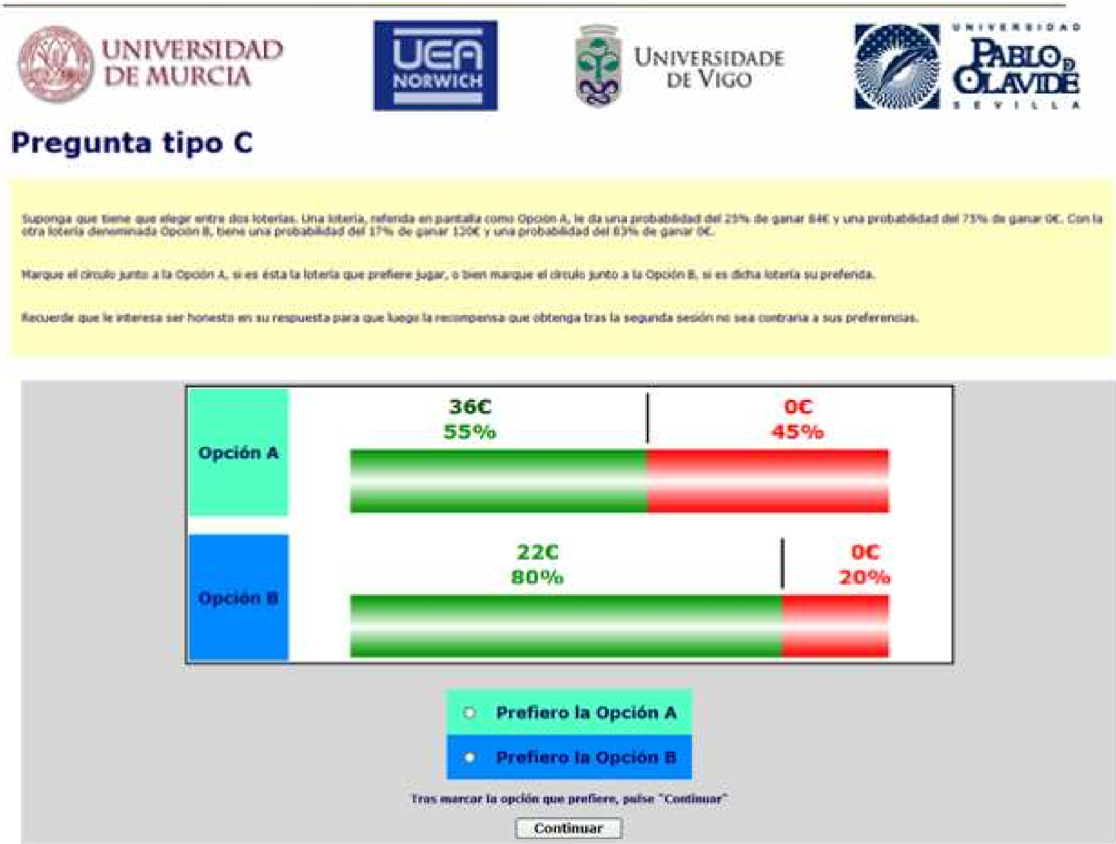
Figure A2: Screenshot of a Probability Equivalence question



The procedure to reach the indifference point was essentially the same as for the ME questions. So, when the subject indicated a preference for the fixed lottery by pressing the "A" button, the probability attached to €120 in Option B increased; and the opposite happened when the "B" button was pressed. When indifference was reached, the subject pressed "Continue" to register that value and move to the next question.

The BC questions presented subjects with two fixed lotteries and asked them to make a straight choice between them. An example of the display is shown in Figure A3.

Figure A3: Screenshot of a Binary Choice question



The Incentive System

After a subject had completed all questions in both sessions, one of his/her decisions was picked at random: it was equally likely to be any question from either of the two sessions. If it was a BC question, he/she simply played out whichever lottery s/he had chosen. If it was an equivalence question, an ‘offer’ – some sure sum of money in the case of an ME question, or a lottery offering some probability of €120 in the case of a PE question – was drawn at random: if this was as good as, or better than, the stated indifference sum/probability, the individual either received the full amount of the sure money offer or else played out the €120 lottery offered. If the offer was worse than the stated indifference value, s/he played out Option A instead.

Appendix 2: Results Tables After Exclusions (n = 165)

Table 2: Key Statistics for Money Equivalents and Probability Equivalents

Lottery	A	B	C	D	E	F
Average meanME	20.30	17.15	14.91	20.99	18.88	15.48
Average medME	20.20	17.00	15.02	20.87	18.84	15.52
Average sdevME	3.55	2.49	1.98	3.21	2.28	1.49
Median sdevME	2.88	2.15	1.80	2.89	2.01	1.41
Average sdevPE	1.81	3.98	5.93	4.91	4.98	6.20
Median sdevPE	1.64	3.22	4.83	3.33	3.78	5.24
Average meanPE	16.96	23.27	26.21	22.61	24.46	27.29
Average medPE	17.02	23.12	25.83	22.18	24.20	26.85

Table 3: Binary Choice Distributions

{R, S}	Frequency of Choice of Riskier Lottery							
Pair	6	5	4	3	2	1	0	Total R:S
{A, B}	17	12	15	19	14	26	62	333 : 657
{B, C}	16	11	15	12	22	22	67	313 : 677
{A, C}	27	13	7	17	16	27	58	365 : 625
{D, E}	54	27	34	16	13	11	10	680 : 310
{E, F}	37	25	27	20	15	19	22	564 : 426
{D, F}	57	20	20	21	15	15	17	630 : 360
{A, D}	10	13	8	3	12	32	87	222 : 768
{B, E}	5	3	9	5	10	17	116	133 : 857
{C, F}	1	1	-	-	1	10	152	23 : 967

Table 4: Comparing ME_Rs with ME_Ss

Lottery Pair	$ME_R > ME_S$	$ME_R = ME_S$	$ME_R < ME_S$
A vs D	2295 (38.6%)	611 (10.3%)	3034 (51.1%)
B vs E	1631 (27.5%)	603 (10.1%)	3706 (62.4%)
C vs F	2292 (38.6%)	802 (13.5%)	2846 (47.9%)

Table 5: Conjunctions of Median Responses

Direction of Median			Lottery Pairs					
ME	**PE**	**BC**	**{A, B}**	**{B, C}**	**{A, C}**	**{D, E}**	**{E, F}**	**{D, F}**
$R \geq S$	$R \geq S$	$R \geq S$	26	32	39	52	63	62
$R \geq S$	$R \geq S$	$R < S$	10	31	7	7	18	10
$R > S$	$R < S$	$R \geq S$	32	14	25	53	42	50
$R > S$	$R < S$	$R < S$	59	39	61	17	32	30
$R < S$	$R > S$	$R > S$	0	3	1	7	2	4
$R < S$	$R > S$	$R \leq S$	0	11	0	1	4	2
$R \leq S$	$R \leq S$	$R > S$	4	3	1	19	1	1
$R \leq S$	$R \leq S$	$R \leq S$	37	35	34	12	6	9