Modelling Noise and Imprecision in Individual Decisions

Graham Loomes, University of East Anglia, Norwich, 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.

October 2008

Draft: Please do not cite without permission from the corresponding author, Graham Loomes – g.loomes@uea.ac.uk

Introduction

This paper focuses on imprecision and 'noise' in decision making under risk. Although most decision theories are presented in a deterministic format, the fact is that actual behaviour exhibits a degree of imprecision and variability, and a fully specified model requires some way of allowing for that. This paper presents a dataset of a kind that does not seem to have been available until now and uses it to examine different approaches to modelling the stochastic component in various basic decision tasks. Section 1 sets the scene and shows how different assumptions may entail very different implications for the interpretation of data. Section 2 describes in more detail the experiment we conducted and reports the results. These results strongly suggest that the approach which, until now, has seemed to be the 'natural default' for economists and econometricians may not, after all, be the appropriate one to use in this area. The final section discusses some broad implications for this and other areas of economic and econometric modelling.

1. The Questions To Be Addressed

In order to give the non-specialist reader some feel for the questions we seek to address, we begin by outlining our experimental design in very broad terms. At the centre of our design were six lotteries, which we shall identify by the labels A to F. For all six lotteries, the individuals in our sample were asked to undertake three types of task, as follows:

ME. For each lottery, each respondent was asked on six different 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 Euros (€), so for each lottery, each respondent was asked on six different 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.

PC. Nine different pairwise choices were constructed and each choice was presented to every respondent on six different occasions.

The reader might like to reflect on the following questions:

(1): Should we expect each individual to give the same ME response for a particular item on each of the six occasions he is asked about that item? And if not, why might someone give different values for the same item on different occasions within a fairly short space of time? In the event of some variability, what is the best way to analyse the responses?

(2): Same as (1), but for the PE responses.

(3): Same as (1), but for the PC responses.

(4): If we do observe some variability in responses, should we expect the degree of variability to be much the same for every lottery and/or for PE as for ME? Or might it be quite understandable for the individual's spread of ME values for some lottery G to be quite different from the spread of ME values for some other lottery H? And if so, should we expect a similar difference in spread to apply to that individual's PE values for G as compared with his PE values for H?

(5): How should we expect any of the above to relate to choice? For example, suppose that we observe an individual's six ME values for lottery G and his six ME values for lottery H: could we expect to be able to predict his responses in the six PCs between G and H? Should we expect the same prediction to emerge from the PE responses?

In the current state of knowledge, what answers might we expect to those questions?

Perhaps the first point to note is that there is now no doubt that most individuals' responses *are* prone to some degree of variability. At present, there is little usable evidence concerning repetitions of ME and PE tasks[1], but there are now a number of databases where individuals make the same PCs on two or more occasions within the same experiment. The bulk of this evidence relates to lotteries with straightforward probabilities and usually involving no more than three money payoffs (one of which is often zero). So these are not complex alternatives requiring difficult calculations, although they are usually set up by experimenters so as to be considered reasonably evenly matched by typical respondents. Under these circumstances, as many as 30% of respondents may choose differently when presented with the same pair of lotteries on two occasions (see, for example, Camerer, 1989; Starmer and Sugden, 1989; Hey and Orme, 1994; Ballinger and Wilcox, 1997; Loomes and

[1] The kind of evidence that is needed involves exactly the same decisions being presented on two or more occasions within a fairly short timespan and without any intervening events that might be expected to influence later responses in any systematic way. We know of no databases comprising repetitions of PE tasks (which are anyway quite rare in the literature). There have been quite a number of experiments asking respondents to state sure sum values for the same lotteries in repeated rounds of buying or selling auctions or other markets; but these have invariably involved feedback to respondents at the end of each round about the market price (and sometimes other information) which is liable to influence subsequent stated values in systematic ways which would make it unwise to use those data to investigate any models which entail independence of the stochastic component in people's responses.

Sugden, 1998). And when we look at the record of any one individual making a number of twice-repeated choices, it is often only a minority who always make the same choices on both occasions for all pairs presented to them.

So variability of response is an established fact of experimental life, and it is therefore necessary for those building and testing decision models to try to take appropriate account of it. However, the majority of decision models developed during the past 30 years have tended to set the issue of variability / noise / error / imprecision to one side: they have generally been formulated deterministically rather than probabilistically. The challenge, then, is to reformulate such models to accommodate the somewhat stochastic nature of people's preferences while trying at the same time to remain faithful to what is distinctive about any particular deterministic 'core'. The answers to the questions posed in (1) to (5) above are liable to vary according to the ways in which certain core principles interact with different stochastic specifications.

A large number of different deterministic theories currently exist; and there are potentially many variants of assumptions about the stochastic nature of preferences. However, it turns out that we can make some progress by dividing 'core' theories into two classes – those that entail transitivity and those that do not – and by considering three categories of stochastic specification – namely, 'tremble' models, 'Fechner' models and 'random preference' models. We shall return to the issue of transitivity in due course. But first we focus on the three different approaches to modelling the stochastic component in behaviour, all of which can initially be illustrated by taking von Neumann-Morgenstern expected utility (EU) theory as our 'core'[2].

The notion of a 'tremble' in this context is as follows. It is supposed that an individual has some true preference according to EU (or any other core theory) but that on any particular occasion when he is asked to state his preference, his response is subject to the possibility of lapse of attention (or, in a computerised experiment, hitting the wrong key) such that he mis-states his real preference. Thus in a pairwise choice between lotteries G and H, and individual who truly prefers G will actually report a preference for H with probability ω (and hence only report his true preference with probability 1- ω) – and conversely, someone who truly prefers H will report that

[2] We take EU as the core because it is the theory most likely to be most familiar to non-specialist readers and because it is adequate for our current purpose. There is no presupposition that it is the 'best' core theory.

preference correctly with probability 1-ω but will (wrongly) state a preference for G with probability ω.

The simplest form of this model was used by Harless and Camerer (1994), who assumed that ω was the same for all individuals and all choices. However, there is plentiful evidence to show that both of these assumptions are false (see Loomes, 2005, for a discussion). Moreover, it is not obvious how to apply it to ME or PE tasks. For example, suppose that an individual's true ME is 10€: how many (and which) other values might be reported as a result of inattention or mis-typing?

However, given the evidence that this model performs poorly even in relation to PC tasks, there seems little point worrying about whether it can be generalised to a broader set of tasks. This is not to say that slips and momentary inattention play *no part at all* in decision data: when respondents are being asked to make a large number of choices of a somewhat unfamiliar kind in a relatively short period of time, it would be surprising if there were no such trembles; and indeed, Loomes, Moffatt and Sugden (2002) argued that a 'tremble term' is a useful adjunct to other ways of specifying the stochastic element in choice. But it is not viable as the principal model.

By contrast, Fechner models[3] have the potential to be much more general and flexible, while appearing to fit very well with a broader econometric tradition that adds a 'noise' or 'error' term to an otherwise deterministic relationship between dependent and independent variables. In the context of decision making under risk, Fechner models suppose that each individual has some true subjective value for any given lottery but that on any particular occasion, her judgment of that subjective value – a judgment which provides the basis of action on that occasion – may be subject to some variability.

To illustrate, consider a pairwise choice between two lotteries involving, between them, three payoffs $x_3 > x_2 > x_1 \geq 0$. Lottery G offers payoff x_3 with probability p and x_1 with probability 1-p, while lottery H offers x_2 with probability q and x_1 with probability 1-q, where q > p. If the decision maker's core preferences are EU, the true subjective value to her of lottery G, EU_G, is given as $p.u(x_3) + (1-p).u(x_1)$,

[3] The origins and essence of such models can be traced to the work of psychophysicist Gustav Fechner (1860/1966) who studied the relationship between the perceived values of physical stimuli (weight, sound, light, etc.) and their 'objective' values and found most subjects' perceptions to be distributed (more or less normally) around the true value but involving some probability of some deviation on any particular occasion of judgment.

while for H, EU_H is $q.u(x_2) + (1\text{-}q).u(x_1)$; and under the deterministic model, the individual's preference between G and H depends on whether EU_G is greater than, equal to, or less than EU_H.

However, under a Fechnerian specification, the judged subjective value of G on any occasion is given by $EU_G + \varepsilon_G$, where ε_G is a random deviation from that true value which on any particular occasion can take a positive or a negative value. The general supposition is that the degree of variability (i.e. the variance of ε_G) is a matter of how the characteristics of the lottery interact with the characteristics of the individual and his preferences. Meanwhile, the judged subjective value of H is given by $EU_H + \varepsilon_H$, where ε_H is a random deviation from EU_H which depends on how the individual interacts with *that* lottery's characteristics, and which is quite independent of ε_G. The assumption is that on any particular occasion when an individual is being asked to make a choice between those two lotteries, it is as if she compares a particular realisation of $EU_G + \varepsilon_G$ with an independent realisation of $EU_H + \varepsilon_H$ and chooses whichever of the two is perceived on that occasion to be greater.

With ε_G assumed to be independent of ε_H, this allows the possibility that even if G is truly preferred to H (i.e. $EU_G > EU_H$), H may be chosen on any occasion when the difference between the two noise terms, $\varepsilon_H - \varepsilon_G$, is greater than the difference between the true values $EU_G - EU_H$. How frequently this is likely to happen depends on the magnitude of the true difference $EU_G - EU_H$ in conjunction with the joint distribution of ε_G and ε_H: the stronger the true preference for G, the less likely it is that H will be perceived to be better on any particular occasion, all other things being equal; likewise, the smaller the variance of the joint distribution of ε_G and ε_H, the less likely it is that noise will 'overturn' true preference.

When this approach is applied to modelling decisions involving lotteries, we might ask various questions. For example, for any individual, is the variance of ε liable to be different from one lottery to another – and if so, for what reason(s)? Is it some function of the magnitude of the true subjective value? Does it depend on the complexity of the lottery? Are any other characteristics relevant? At present, relatively little empirical work has been done to examine these questions[4]: an

[4] But see Buschena and Zilberman (2000) for some investigation of some possible sources of heteroskedasticity in ε.

important part of the motivation for the present paper was to gather more information that would be useful in addressing at least some of those questions.

Irrespective of the answers to those questions, what is common to all of the different variants of Fechnerian error models is that the distribution of noise/error is seen as a reflection of how the characteristics of the lottery interact with the characteristics of the individual and his/her preferences. Under this approach, it is as if each lottery carries its own distribution of ε into every decision that particular individual is asked to make involving that lottery. So a pairwise choice between G and H is modelled in terms of the overlap between the distributions of the two lotteries' respective ε's around their respective true values. And although such models have until now been applied primarily to PC tasks, we shall see that they can be quite straightforwardly extended to ME and PE tasks.

However, before considering those implications, let us consider the third approach to modelling the variability in individuals' stated preferences. The random preference (RP) way of modelling EU supposes that on any occasion when asked to make some decision, the individual acts so as to maximise EU as perceived on that occasion, but that his perceived utility function is liable to vary from one occasion to another.

The conceptual foundation of RP is a model of human cognition that does not suppose that each individual is essentially characterised by a single true preference function, but rather that each individual's perceptual and judgmental apparatus comprises of some continuum of states of mind. Thus it is as if an individual's preferences are represented by a *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.

Taking lotteries G and H as specified above, and denoting the individual's utility function under his state of mind at time t as $u_t(.)$, G will be preferred to (less preferred than) H according to whether $[u_t(x_3)-u_t(x_2)]/[u_t(x_2)-u_t(x_1)]$ is greater than (less than) $[q-p]/p$. Representing the variability of an EU maximiser's judgment in terms of a set of $u_t(.)$ functions which vary from one state of mind to another amounts to allowing for that individual to have variable perceptions of how the utility difference between x_3 and x_2 compares with the utility difference between x_2 and x_1. On those occasions where $[u_t(x_3)-u_t(x_2)]/[u_t(x_2)-u_t(x_1)]$ is greater than $[q-p]/p$ – that is, on those occasions when it is as if he has selected at random a $u_t(.)$ for which that inequality holds – he chooses G; alternatively, on other occasions where it is as if he has picked a $u_t(.)$ where $[u(x_3)-u(x_2)]/[u(x_2)-u(x_1)] < [q-p]/p$, he chooses H. Thus the probability of an individual choosing G or H can be thought of as the probability that on the occasion when that choice is presented to him, he happens to be in one of the (less risk averse) states of mind where $[u_t(x_3)-u_t(x_2)]/[u_t(x_2)-u_t(x_1)] > [q-p]/p$ or else happens to be in a (more risk averse) frame of mind where $[u_t(x_3)-u_t(x_2)]/[u_t(x_2)-u_t(x_1)] < [q-p]/p$.

It has been shown elsewhere (Loomes, 2005, for example) that different variants of Fechner specifications of EU can have very different implications from one another, as well as from the RP form of EU. However, the implications for some of the questions posed in **(1)** to **(5)** above have not previously been established. So let us now address those questions.

1.1 Implications of the Fechner Approach

The implications of the Fechner approach for pairwise choices are straightforward. In its most general form, the approach allows different individuals to have different true subjective values for different lotteries, subject only to the restrictions imposed by the core theory that describes the structure of their preferences; and it also allows the ε's to vary from one individual to another and from one lottery to another, depending on how the characteristics of a particular individual interact with the characteristics of a particular lottery. Thus for any pair of lotteries G and H we should not necessarily expect an individual to choose the same lottery on every occasion when she is asked to make a choice: modelling the process as if fresh values of ε_G and ε_H are drawn (independently) on each occasion, the probability that G will be chosen on any occasion is simply the probability that $EU_G - EU_H$ is greater than $\varepsilon_H - \varepsilon_G$. We can extend this conclusion to any other non-EU core theory which, despite departing from EU in some respect, nevertheless entails some true subjective value (SV) for each lottery: in these cases, the probability of G being chosen from the pair {G, H} is simply the probability that $SV_G - SV_H > \varepsilon_H - \varepsilon_G$.

To extend the Fechner approach to MEs and PEs for any lottery such as G, we can proceed as follows.

First, MEs. Any sure amount of money M can be regarded as a degenerate lottery with its own distribution of ε, whose variance obeys some properties. Without being specific about how the variance of ε behaves for such degenerate lotteries[5], we can imagine 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. Equally, for some sufficiently high values of M there might be no consequential overlap at the other end of the $SV_G + \varepsilon_G$ distribution, so that he would be extremely unlikely to judge G to be as good as these high values. But as M is progressively increased over some intermediate range where $SV_M + \varepsilon_M$ and $SV_G + \varepsilon_G$ overlap, it becomes progressively more likely that M will be judged at least as good as G.

If the ε's are distributed symmetrically around zero means, the probability of M being judged at least as good as G rises to 50% at the point where $SV_M = SV_G$ and then increases further until M is so high that G is (almost) never preferred. As 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[6].

Two possibilities follow which are relevant to the questions posed under **(4)**. First, if different lotteries are associated with different degrees of noise, this may be reflected by the variance of their MEs, all other things being equal. So suppose that lottery H is such that $SV_G = SV_H$, but for some reason 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

[5] 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.

[6] Under EU, for example, it could be that u(.) is markedly concave so that a symmetrical distribution of perceived EUs 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.

variance of MEs for G. If this were the appropriate error model operating in conjunction with a deterministic 'core' theory which entails transitivity, it might allow us to gain insights into the features of lotteries that are associated with larger or smaller variances of ε.

The second possibility is that for any particular lottery, the variance of the PEs might be different from the variance of the MEs. Let us denote the 'yardstick' lottery (the one involving a payoff of 120€ in our earlier example) by Y. The variances of the ε_Y's associated with any of the yardstick lotteries Y that span the relevant range may be different from the variances of the ε_M's over the corresponding range. Thus the joint distributions of ε_G and ε_Y could have a different variance than the corresponding joint distributions of ε_G and ε_M.

However, although it is possible that the variance of MEs and PEs may be different across lotteries, and although the same lottery might be associated with a different variance of MEs than of PEs, one thing we should *not* expect under any Fechner model is that the variances of MEs *across lotteries* display a very different pattern from the variances of PEs *across the same set* of lotteries. For example, if for any individual the variances of the MEs were to decline progressively from G to H to J, we should suppose this to reflect a tendency for the variance of ε_J to be less than the variance of ε_H which, in turn, is less than the variance of ε_G. Thus the joint distributions of ε_Y with each of the lottery error terms should vary correspondingly, so that the variances of the PEs should not display a very different trend.

Finally, let us consider the implications of the Fechner approach for the questions listed under **(5)**. Since, for each individual, every lottery has its own ε distribution, the distributions of MEs and PEs might be thought to be proxies for the SV + ε distributions. So if, for example, the distribution of ME_Gs and ME_Hs were such that there is a 60% chance that an ME_G drawn at random from the first distribution would be greater than an ME_H drawn at random from the second 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 G would be chosen in roughly 60% of choice repetitions[7]. Similarly, we should expect the choice proportions to be broadly in line

[7] If the utilities of sure sums of money are perceived with no noise/error, as assumed in Blavatskyy (2007), the correspondence 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.

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 obviously out of alignment.

Moreover, if core preferences are essentially transitive, we should expect to see the same orderings of items expressed via all three types of task: for any individual for whom $SV_G > SV_H > SV_J$, we should not only expect to observe a majority of choices where $G \succ H$, $H \succ J$ and $G \succ J$, but we should also expect the median/mean values of the MEs and the PEs to be ranked in that same way[8].

Those, then, are the expectations we might have if people's preferences can be modelled in terms of some transitive core plus some form of Fechner error term. But the RP approach is an essentially different way of modelling imprecision/variability which has some very different implications.

1.2 Implications of the Random Preference Approach

The RP approach to pairwise choice is straightforward. For any {G, H} pair, an individual's set of preference functions can be partitioned into those that entail a preference for G and those that entail a preference for H (with the possibility that one or other subset may be empty). The probability of choosing G is then the same as the probability that a randomly-selected preference function will belong to the G subset.

This way of modelling the stochastic nature of preferences is easily extended to tasks such as judging equivalences. Suppose an individual is asked to state a sure payoff x_M such that he is indifferent between the certainty of that payoff and playing out lottery G. Under the RP version of EU, it is as if $u_t(.)$ is selected at random from the set of possible states of mind and on that basis the individual identifies the payoff x_M such that $[u_t(x_3)-u_t(x_M)]/[u_t(x_M)-u_t(x_1)] = [1-p]/p$. The value of x_M that satisfies this equation is liable to vary from one $u_t(.)$ to another, so the RP model entails some distribution of x_M derived from the distribution of $u_t(.)$ that constitutes the individual's set of possible states of mind.

[8] However, to follow up the qualification suggested in footnote 6, if nonlinearities in preference functions lead to these distributions being significantly asymmetrical, it is possible that the means will be less reliable indicators than medians.

Exactly the same reasoning applies to PE. For each $u_t(.)$ there will be some probability of the yardstick payoff (so long as that payoff is greater than x_3) which will make the individual indifferent between G and Y under that state of mind. Denote that probability by y_G. The distribution of $u_t(.)$ thus maps to some distribution of y_Gs.

Moreover, this applies not only to EU but 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 an individual is in a state of mind where his preference function entails a mapping between the lottery and that PE; and likewise for MEs.

Turning to the questions posed under **(4)**, consider first the case of an EU maximiser whose $u_t(.)$ functions can be ordered by some risk aversion parameter. Suppose that he is asked to undertake choice and valuation tasks involving two binary lotteries, G and H, where both the variance and expected value of G is greater. For some $u_t(.)$ 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 at the more risk-averse end of the spectrum, the $u_t(.)$ entail $H \succ G$. Let the proportion of $u_t(.)$ that entail $G \succ H$ be denoted by α: then α is the probability of observing $G \succ H$ on any occasion when he is asked to make a PC, while we can expect to observe $H \succ G$ with probability $1-\alpha$.

Now suppose we draw a representative sample of MEs for each lottery[9]. Were we to happen to draw the function where $G \sim H$, 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 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 $G \sim H$, 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 MEs would therefore exhibit a greater range and variance for ME_G than for ME_H.

If $\alpha > 0.5$, the median $u_t(.)$ 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 $u_t(.)$

[9] Notice that the assumption being made here is that the same probability distribution over preferences 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 due course we shall discuss what the relaxation of that assumption might signify.

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 ranges/variances of the MEs would be unaffected.

What about the PEs? Consider again the $u_t(.)$ where G ~ H. This entails some probability y* such that $PE_G = PE_H$. For all functions involving less risk aversion, the probability of the yardstick payoff would be lower than y* 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 $u_t(.)$ 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 $u_t(.)$ from the more risk averse part of the distribution, the probabilities of the yardstick payoffs would be greater than y*; 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 range/variance of a sample of PE_Hs to be greater than the range/variance of a comparable sample of PE_Gs. This relationship between variances is in the *opposite* direction to our expectation for the MEs and provides another 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.

However, if $\alpha > 0.5$, the median PE_G would be expected to be greater than the median PE_H, just as the median ME_G would be expected to be greater than the median ME_H, given that both are reflecting the median degree of risk aversion. Alternatively, if $\alpha < 0.5$ the median choice would be H $\succ$ G and for any core theory respecting transitivity we should expect both median ME_H > median ME_G and median PE_H > median PE_G.

Of course, if the underlying constituent functions were transitive but not capable of being ordered according to degree of risk aversion, it would be more difficult to derive such clear implications. And if the underlying functions did not respect transitivity, that would also complicate matters.

However, rather than try to work through all of the possible combinations and variants imaginable, our starting point is what many economists and decision theorists might be inclined to take as the default: some transitive core theory where preference functions can be ordered according to risk attitude. With such a core, a Fechner specification entails distributions of MEs and distributions of PEs that both reflect (at least approximately) the distributions of ε associated with each lottery: so if the ME_Gs display greater/equal/less variance that the ME_Hs, the PE_Gs should be expected to display correspondingly greater/equal/less variance than the PE_Hs. By contrast, an RP specification would involve the variances of the MEs increasing with the riskiness of the lottery, while the variances of the PEs are liable to move in the opposite direction.

When considering the questions asked under **(5)** about the correspondence between ME, PE and PC, we saw that the Fechner model implied that the probabilities of choosing each lottery in a PC *should* be broadly predictable on the basis of information about the MEs or the PEs: in particular, the probability that a randomly selected ME_G is greater than an independently sampled ME_H should be much the same as the probability of G being preferred to H in a pairwise choice; and likewise for the PEs. By contrast, RP allows very different patterns to emerge from the different tasks.

To see how striking this difference *could* be, consider an individual whose $u_t(.)$ are *all* concave: that is, she is risk averse in every state of mind, but the *degree* of risk aversion varies from one state of mind to another – let us say, from mildly risk averse at one end of the spectrum to quite markedly risk averse at the other end.

Now consider two lotteries G and H where both have the same EV but where the spread of G is unambiguously greater than the spread of H. When asked on a number of separate occasions to give MEs for H, our individual gives values ranging from just below the EV (when she is in a mildly risk averse frame of mind) to quite substantially below the EV (when minded to be quite markedly risk averse). Asked on a number of different occasions to give MEs for G, these too cover some distribution; and because G is unambiguously riskier than H, this distribution ranges from some value less than the highest ME_H down to some value rather lower than the lowest ME_H. If both samples of MEs are broadly reflective of the individual's distribution of $u_t(.)$, the mean and median ME_G will, respectively, be somewhat lower than the mean and median ME_H; but there might be considerable overlap between the two distributions, so that the chance of a randomly selected ME_G being greater than a randomly selected ME_H could be quite substantial, albeit less than 0.5.

However, this individual's pairwise choices will display a very different pattern: in this example, *every* $u_t(.)$ is concave, so she will exhibit some degree of risk aversion on every occasion of choice and *always* prefer H to G. Thus there could be a sizeable minority of instances where a particular ME_G will exceed a particular ME_H, while in direct pairwise choice, H will *always* be chosen[10].

Although this latter example may be somewhat extreme, it highlights another way in which the Fechner and RP approaches have different and potentially testable implications. We now describe our attempt to test this and the other implications identified above.

2. New Experimental Evidence

2.1 Design

As noted earlier, virtually all of the existing data pertaining to the variability of people's decisions consists of responses to being presented with the same pairwise choice on two or more occasions. The experiment we conducted sought to expand the database in two ways: first, by asking each respondent to repeat each PC on six rather than two occasions; and second, by asking those same individuals to repeat both the ME and the PE task six times for each lottery being considered. Thus we elicited all MEs and PEs under exactly the same conditions as the PCs, giving no feedback until all tasks had been completed, and then paying respondents on the basis of playing out one of those decisions picked at random at the very end of the experiment.

The experiment revolved around the six binary lotteries described in Table 1.

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

In each case, payoffs were in Euros. The six lotteries can be viewed as constituting three triples, {A, B, C} and {D, E, F}. Within each triple, safer lotteries

[10] Under RP alone, G would never be chosen by this individual; but if we allow for the very occasional 'tremble', we could observe G to be chosen now and then.

offer lower expected values (EVs); and D, E and F respectively offer the same EVs as A, B and C but each involve smaller spreads than their {A, B, C} counterparts.

Each participant in the experiment was required to attend two sessions, several days apart, each of which followed the same format[11]. Having signed in and read the instructions (the instruction/practice period was more formal at the beginning of the first session, but the instructions were available at the beginning of the second session for anyone who needed to refresh their memories), each participant answered 63 questions per session, organised in three successive 'phases', each consisting of the same 21 questions, as follows: an ME elicited for each of the six lotteries; a PE elicited for each of the six lotteries, with the yardstick payoff set throughout at 120€; and nine pairwise choices – {A, B}, {B, C}, {A, C}, {D, E}, {E, F}, {D, F}, {A, D}, {B, E} and {C, F}. Thus by the end of the second session, we had collected from each respondent six MEs for each lottery, six PEs for each, and six observations of every PC listed above.

All equivalences were elicited using an iterative choice format (see the Appendix for details and examples of displays) in order to make choice and equivalence as procedurally similar as possible. Standard incentive mechanisms were used (again, see the Appendix for details).

In terms of equivalents, this allows us to compute for each individual an estimate of the mean and median ME and PE together with the standard deviation and range of the responses given. Such data enable us to address the various issues raised in the previous section, as follows.

(1), (2) and (3): We will be able to see immediately any patterns of variability in choice and/or equivalence.

(4): If we do observe some variability in responses that goes beyond the occasional tremble, we can ask the following questions:

a) Can we accept the null hypothesis that there is no difference between standard deviations across lotteries? If we can accept this null, it would be consistent with a Fechner model assuming constant variance of ε – that is, the assumption made, for example, by Hey and Orme (1994) when testing the fit of various core theories to their data. On the other hand, if this null is rejected, we can ask:

[11] Participants were only paid at the end of the second session. Only data from those who completed both sessions will be used in the analysis.

b) If standard deviations differ systematically between lotteries, do the differences follow the same pattern for MEs as for PEs – as Fechner models would suggest – or do they operate in the reverse order, as RP would suggest?

(5): We can examine the relationship between equivalence and choice, and also between the medians/means of the various distributions:

a) Under Fechnerian assumptions, the degrees of overlap between any two sets of ME (alternatively, PE) responses should broadly correlate with the frequencies of choice in the repeated PC tasks. However, under RP there could be substantial differences. The pairs {A, D}, {B, E} and {C, F} may be particularly telling in this respect. Within each pair, EVs are the same and we might find considerable overlap between the ME and PE distributions, suggesting proportions of choice not very far from 50:50 on Fechner assumptions. However, because EVs are the same within each of those pairs, all risk averse preference functions would entail choosing the safer alternative: so if people operate in an RP manner and if their preferences with respect to these choices are predominantly risk averse, we might find a very different pattern of choice than the ME or PE distributions would entail under Fechner assumptions.

b) Under the assumption that preference functions can be ordered according to degree of risk aversion, and on the basis that underlying preferences are transitive, we might expect an individual's median MEs, median PEs and the majority of repeated PCs to give the same orderings, both under Fechner and under RP specifications. However, if the data suggests otherwise, we should need to consider the significance of this for assumptions about the nature of any potential core theory.

2.2 Results

Before reporting the results, some remarks about exclusions. Of those who participated in the experiment during a three-week period in March/April 2008, a total of 274 individuals completed two sessions[12]. 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 stochastic dominance, 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

[12] 45 others came to the first session but did not attend the second session within the time limit we imposed.

stating a PE at least as high as the probability of the high payoff in the {A, B, C} lotteries. 109 of the 274 respondents gave an answer that violated stochastic dominance on at least one occasion. In order to abstract from any such 'trembles' (or perhaps in some cases, confusion), we shall focus primary attention on responses by the 165 people who displayed no violations of stochastic dominance at all.

Consider first the patterns of pairwise choice reported in Table 2.

Table 2: Pairwise 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

For each respondent we observed the number of times out of 6 repetitions that he/she chose the riskier lottery within each pair. For each pair, Table 2 shows the 165 individuals categorised accordingly: that is, the column headed '6' shows how many respondents chose the riskier lottery every time; 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 they faced that pairwise choice.

It is immediately apparent that for the bulk of these data, no 'true-preference-plus-small-tremble' story is plausible. The only case which might appear compatible with such a story is the {C, F} choice, which might be reconciled with a hypothesis that 163 of the 165 respondents truly prefer F but have about a 1% chance of picking C by mistake. Otherwise, and especially in the first six rows, there are far too many individuals in the columns in the middle for any such story to be credible.

At the same time, there appear to be some very definite and quite intuitive trends. For the {A, B, C} pairs, clear majorities of choice 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.

Although the EV differences are the same for the {D, E, F} counterparts, the effect of raising the minimum payoffs to 8€ or 9€ and reducing the spreads 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.

However, when the EVs are equalised, as in the last three pairs, there are very substantial majorities favouring the safer alternatives, especially in the {C, F} choice. So despite the clear evidence of noise/imprecision in people's choices, there are also many signs of systematic tendencies underlying them. Is the same true for the equivalence responses? Table 3 shows some aggregate statistics.

Table 3: 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

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 3 reports the sample averages for these variables, plus the medians of the standard deviations, with those standard deviations in the middle rows for easier comparison between MEs and PEs.

Here too it is evident that there is considerable variability in respondents' evaluations of each lottery, with median coefficients of variation ranging between 9% and 19%[13]. Thus on the basis of both the choice data in Table 2 and the equivalence data summarised in Table 3, we can see that there is indeed compelling evidence of noise/imprecision that goes beyond the occasional tremble and requires some more substantial explanation. So we now turn to the questions listed under **(4)** and **(5)**.

The null hypothesis in **(4a)** is that there are no systematic and significant differences between the standard deviations of the MEs of the various lotteries; and likewise for the PEs. The data reject this null for both measures. In Table 3, both the averages and the medians of the standard deviations of the MEs follow a clear pattern: in every pairwise comparison within each triple, the sdevMEs associated with the safer lotteries are lower than for the riskier lotteries to a highly significant ($p < 0.001$) extent[14]. So even without consulting the PE data, the 'constant variance of ε' assumption is clearly rejected.

The next question is whether the differences move in the same direction for both MEs and PEs, as the Fechner framework would suggest, or whether they change in opposite directions, in keeping with an RP approach. At the aggregate level shown in Table 3, it is clear that both average and median standard deviations of PE responses change in the *opposite* direction to their ME counterparts, very much in line with the implications of the RP approach as set out in Section 1: of the six pairwise comparisons of sdevPEs within the two triples, only the increase from D to E fails to register as statistically significant, with the other five pairwise differences all being significant at the $p < 0.001$ level. It is hard to see how these patterns can be reconciled with *any* variant of the Fechner approach – and certainly not any in the existing literature of which we are aware.

The data relating to **(5a)** seem to provide further evidence against the Fechner framework. The most vivid example comes from comparing the relationships between

[13] These figures can be proxied by dividing the median standard deviations by the corresponding average medians. Much the same figures come from calculating individual coefficients of variation and taking the median values of those.
[14] The test here involved subtracting each individual's standard deviation for the safer lottery from his/her standard deviation for the riskier lottery and testing the null hypothesis that the mean of these differences was zero.

the MEs and PEs for lotteries C and F and the patterns of direct choices between those two lotteries.

Suppose, as the Fechner approach does, that each lottery is evaluated separately and subject to noise, and that the two independent evaluations are then compared, with the individual choosing whichever lottery has the higher evaluation on that occasion. For each individual, we proxy that by taking his/her six ME_C's and his/her six ME_F's and examining all 36 possible ways of matching each ME_C with each ME_F. Under Fechnerian assumptions, the proportion of times an ME_C is greater than an ME_F gives us, for each individual, a proxy for the number of times his/her $SV_C + \varepsilon_C$ exceeds his/her $SV_F + \varepsilon_F$, entailing the choice of C. Likewise, the proportion of times an ME_F is greater than an ME_C indicates the propensity to choose F over C[15]. We can follow exactly the same procedure using PEs instead of MEs to get a second estimate of the frequency with which each lottery will be picked in repetitions of the {C, F} choice if the Fechner approach is appropriate.

Consider first the ME data. 36 combinations per person for 165 individuals gives a total of 5940 comparisons: of these, $ME_F > ME_C$ in 2846 cases (47.9%), while $ME_C > ME_F$ in 2292 cases (38.6%) and $ME_C = ME_F$ in 802 cases (13.5%). For the PE data, the results were: $PE_F > PE_C$ in 2669 cases (44.9%); $PE_C > PE_F$ in 2363 cases (39.8%); and $PE_C = PE_F$ in 908 cases (15.3%). Thus even if all of the cases where $ME_C = ME_F$ or where $PE_C = PE_F$ were counted as favouring F, the Fechner framework would still entail C being chosen on at least 38% of occasions.

However, as the choice patterns reported in the bottom row of Table 2 show, C was actually picked on fewer than 2.5% of occasions. This result is quite incompatible with any Fechner model of which we are aware, but is entirely consistent with RP along the lines suggested by the illustrative example given at the end of Section 1.

On the evidence above, the RP model clearly outperforms the Fechnerian way of modelling variability of judgment. However, there is one possible *caveat*: the implications in **(4)** and **(5a)** were generated on the basis that 'core' preferences were transitive; but the data in Table 3 give grounds for questioning that assumption.

It was suggested in **(5b)** that if underlying preference functions could be ordered according to risk attitude, we should expect median responses to provide

[15] Of course, since many people tend to 'round' their ME responses to some extent, we can expect at least some instances where $ME_C = ME_F$, so there is potentially an issue about how to interpret those instances in terms of the choices people would make. However, as we shall see, the issue does not cause any serious difficulties in this particular case.

insights into the nature of those functions at the centre of the distributions; and if such functions entailed transitivity, we should expect that to be reflected at the level of the individual by the correspondence between median MEs, median PEs and majority choices.

Of course, Table 3 gives overall averages rather than individual-level data. Nevertheless, it suggests some very clear patterns: the median MEs order the lotteries A ≻ B ≻ C in the first triple and D ≻ E ≻ F in the second triple, while the median PEs produce exactly the opposite orderings within each triple. Meanwhile, Table 2 shows aggregate patterns of pairwise choices that do not fit either with MEs or with PEs: for {A, B, C} the majority choices favour the safer options, which is in line with the PEs but is contrary to the MEs; while for {D, E, F} the majority choices favour the riskier options, which tallies with MEs but runs counter to PEs.

The corresponding individual-level analysis is reported in Table 4, which categorises all 165 respondents according to their median ME, median PE and majority PC responses to each pair, with R and S referring, respectively, to the riskier and safer lotteries in each pairing.

Table 4: 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	26	32	39	52	63	62
R ≥ S	R ≥ S	R < S	10	31	7	7	18	10
R > S	R < S	R ≥ 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 ≤ S	0	11	0	1	4	2
R ≤ S	R ≤ S	R > S	4	3	1	19	1	1
R ≤ S	R ≤ S	R ≤ S	37	35	34	12	6	9

So, for example, the top cell in the {A, B} column shows how many of the 165 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, and 26 individuals fell into that category. The 10 in the next cell down favoured A both in terms of MEs and PEs but chose B on at least four of the six PC repetitions; while the 32 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 63 and 73 i.e. roughly 40% of the sample.

However, the frequencies of other patterns in the Table are not evenly or seemingly randomly distributed. If we go to the disaggregated data and focus on cases where there are just strict inequalities, we find the following patterns.

First, 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 – again, 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 PC 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 4, the ratios we observe are 69:4, 70:6, 68:2, 24:26, 50:3 and 40:5: 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.

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 only pair they investigated, they found that the conjunction of choosing the riskier option but placing a higher PE on the safer option outnumbered the opposite mix of choosing the safer option while placing a higher PE on the riskier one. For the six pairs in Table 4, the analogous ratios using majority choices and median PEs are: 24:10, 13:40, 17:7, 65:8, 32:21 and 36:11; again, with one exception, these ratios show the same kind of asymmetry reported in Butler and Loomes (2007), with those asymmetries becoming more pronounced for the {D, E, F} pairs where the riskier options were more often chosen.

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: 91:0, 53:14, 86:1, 70:8, 74:6 and 80:6.

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[16].

At the same time, our data are by no means chaotic. On the contrary, there are many respects in which aggregate patterns in the data move in systematic and quite intuitive directions: e.g. the broad shift of choices towards the higher-EV options in the {D, E, F} pairs compared to their {A, B, C} counterparts, alongside the fact that the overall ME and PE values for D, E and F all increased relative to those for A, B and C respectively, with this direction of change being reflected (though much more strongly and sharply) in the {A, D}, {B, E} and {C, F} pairwise choices.

3. Discussion

The propositions and implications in **(4)** and **(5)** were initially formulated on the basis of fairly general assumptions about core preferences that might apply to many non-EU models as well as to EU. However, the working assumption was that those core preference functions, whatever their particular form, entailed transitivity. On that basis, we considered some contrasting implications of Fechner and RP

[16] However, it has occurred to us that there is a somewhat different approach that might reconcile the data with a transitive core, in the following sense. 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 pairwise choices 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.

approaches for the distributions of ME, PE and PC 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, as discussed in the latter part of the previous section, other features of the data were actually very hard to reconcile with any transitive core. If that is the case – that is, if we should be thinking instead in terms of some core preference functions which allow systematic intransitivities – does this threaten to undermine our conclusions about the relative merits of RP and Fechner models?

We think not, for the following reasons. Consider what would be involved in any core theory that allowed 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[17].

Context, and the holistic evaluation of a decision task (rather than a particular item/lottery considered in isolation) is also central to the RP model, and is what distinguishes RP from any Fechner model, since the latter proceeds as if each lottery is evaluated separately and inclusive of noise, with that evaluation – and any noisiness surrounding it – being essentially a matter of the interaction between the characteristics of the lottery and the tastes/characteristics of the decision maker. 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 be involve a fundamental conceptual mismatch. If core preferences are intransitive, some RP way of specifying variability would appear to be more compatible with that.

It is not our intention to say much more here about the detailed nature of an intransitive core theory that might fit the data[18]. Rather, the main focus of the present paper is upon the appropriate specification of the variability in most of our participants' ME, PE and PC responses. When the database is expanded to encompass equivalence tasks, the evidence strongly suggest that Fechner specifications are simply inappropriate.

---

[17] So, for example, the reason why regret theory can accommodate the classic preference reversal pattern and the analogous pattern of intransitive choice (see Loomes, Starmer and Sugden, 1991) is that the way each lottery interacts with some sure sum of money modifies the utilities of the payoffs differently than happens when the two lotteries interact with each other.

[18] 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-PC 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-PC phenomenon – see Loomes (2008) – but it will require a much broader set of experiments to test more adequately the credentials of that model.

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 thereby 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), 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 insights 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 imprecision and noise in people’s judgments and decisions.

References

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

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.

Harless, D. and Camerer, C., (1994). The Predictive Utility of Generalized Expected Utility Theories, *Econometrica*, **62**, 1251-89.

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

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., (2008). Modelling Choice and Valuation in Decision Experiments, Working Paper available at http://www.uea.ac.uk/eco/people/add_files/loomes/

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.

Loomes, G., Starmer, C. and Sugden, R., (1991). Observing Violations of Transitivity by Experimental Methods, *Econometrica*, **59**, 425-39.

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.

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: 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.

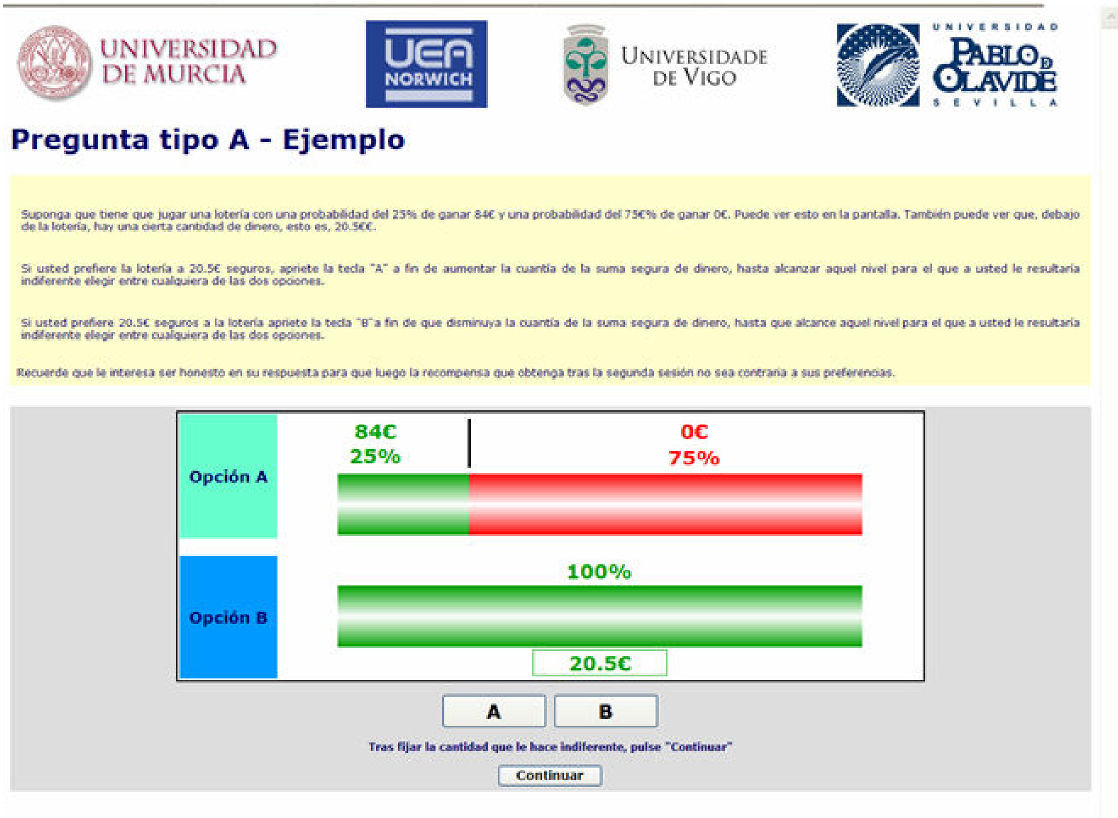
In the second stage, subjects were presented with three practice questions. Each of these questions illustrated the different types of question that participants would see in the third stage. The instructions for each type of question were displayed on the computer screen and they were also read aloud 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 Pairwise Choices (PC). 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.

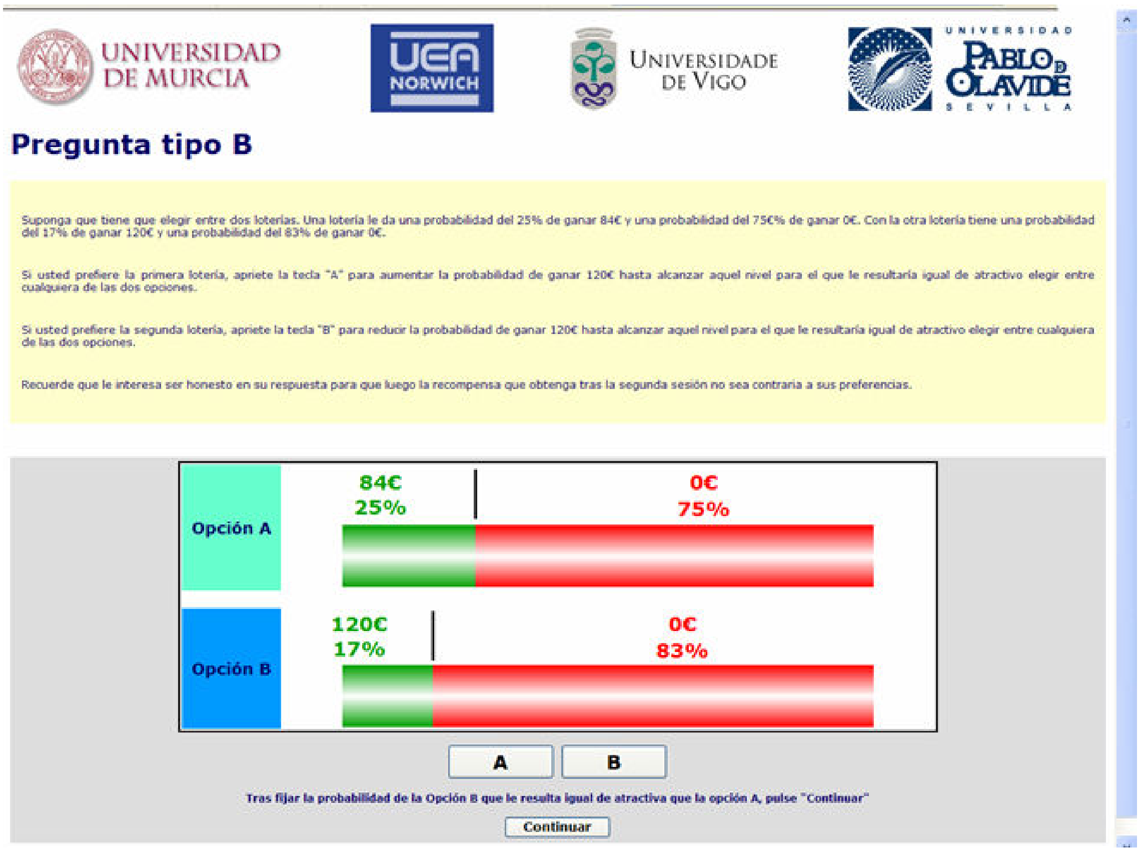
Figure A1: Screenshot of a Money Equivalence question



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.

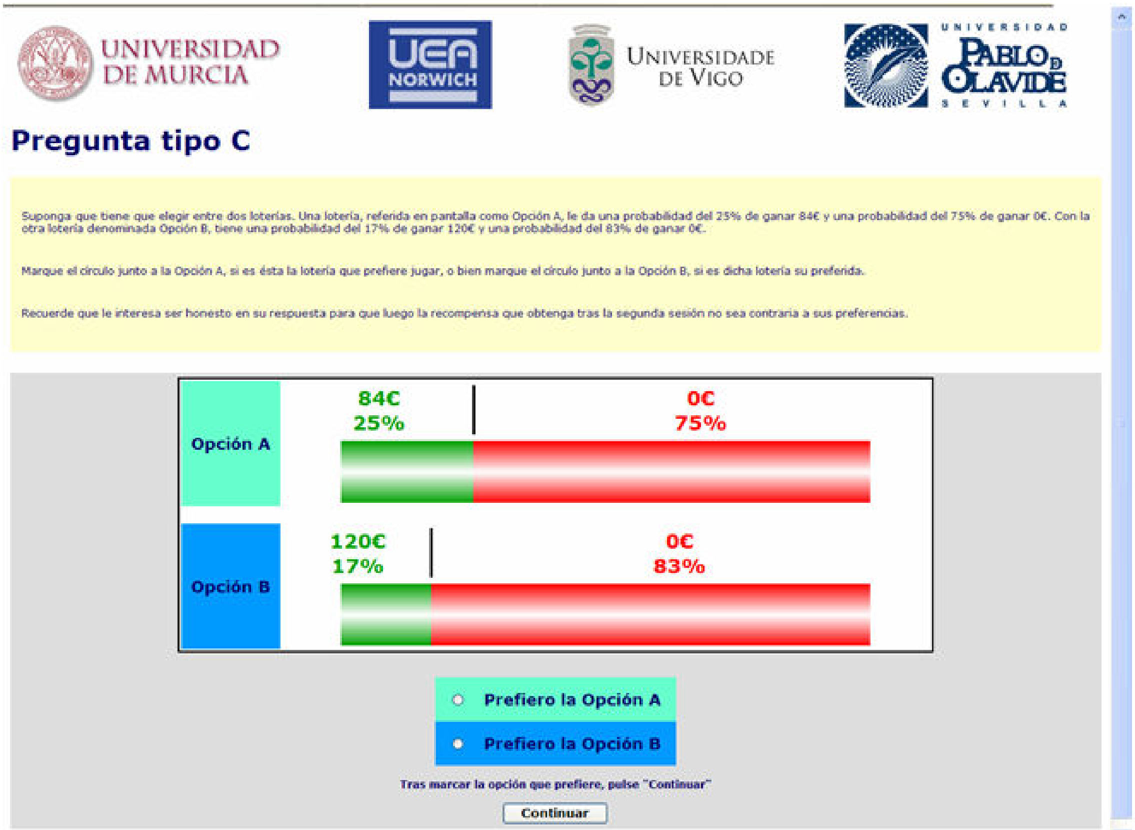
Figure A2: Screenshot of a Probability Equivalence question



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 PC method 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 Pairwise 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 PC 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.