THE UNIVERSITY of York

Discussion Papers in Economics

No. 1999/31

Which Error Story is Best?

by

John Hey and Enrica Carbone

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

Which Error Story is Best?

Enrica Carbone
Universities of York and Sannio

John D. Hey
Universities of York and Bari*

September 29, 1999

Abstract

Two recent papers, (Harless and Camerer, 1994) and (Hey and Orme, 1994) are both addressed to the same question: which is the 'best' theory of decision making under risk? As an essential part of their separate approaches to an answer to this question, both sets of authors had to make an assumption about the underlying stochastic nature of their data. In this context this implied an assumption about the 'errors' made by the subjects in the experiments generating the data under analysis. The two different sets of authors adopted different assumptions: the purpose of this current paper is to compare and contrast these two different error stories - in an attempt to discover which of the two is 'best'.

Keywords: errors, decision making under risk, experiments

JEL classification: D81, C91

Department of Economics, University of York, Heslington, YORK YO1 5DD, UK.
00 44 1904 433786

carbone@unisannio.it ecarbone@unina.it
jdh1@york.ac.uk

*We would like to thank the ESRC for a grant (R000 23 6636) "Experimental Investigation of Errors in Decision Making" which financed the research reported in this paper.

Two recent papers concerned with the empirical investigation of theories of decision making under risk, (Harless and Camerer, 1994) and (Hey and Orme, 1994), were both addressed to the same question: which is the 'best' theory of decision making under risk? As an essential part of their separate approaches to an answer to this question, both sets of authors had to make an assumption about the underlying stochastic nature of their data. In the context of deterministic theories of decision making under risk this involved an assumption about the 'errors' made by the subjects in the experiments generating the data under analysis. The two different sets of authors adopted different assumptions: the purpose of this current paper is to compare and contrast these two different error stories - in an attempt to discover which of the two is 'best'. A direct comparison of these two different error stories from the *original* two papers is not possible as they differed in a second crucial way - in terms of the way that the data was fitted: Harless and Camerer fitted the data *across all subjects* whilst Hey and Orme fitted the data *subject by subject.* Since the Harless and Camerer data does not allow us to fit the Hey and Orme error story (because the number of observations per subject is not sufficiently large) we are obliged to use the Hey and Orme data and fit both error stories to that. This note reports on the results of so doing.

1 Error Specifications

We begin by defining the two error stories proposed in the papers cited above. Both concern decision making in a pairwise choice task: subjects are asked which of two risky prospects they prefer. The first of these papers (Harless and Camerer, 1994) simply assumes that there is a constant probability θ that the agent will make a mistake[1] on any pairwise choice question - and that *this probability does not depend upon the nature of the pairwise choice question* itself. We call this the 'Constant Probability' (**CP**) error story[2]. The story

[1]By 'make a mistake' we mean that the agent says that he or she prefers the left (right) hand choice when in fact he or she prefers the right (left) hand choice.

[2]We also assume, as do Harless and Camerer (Harless and Camerer, 1994), that θ is constant across all questions and also across all subjects, but this, of course, is not necessary. We could, for example, assume that the probability varies across subjects, or across classes of questions, though of course some constancy is necessary to justify the spirit of the assumption.

proposed in (Hey and Orme, 1994) is quite different. It goes back to the primitive of the preference functional $V(.)$ implied by the theory: according to a theory with preferences given by $V(.)$, L is preferred to R if and only if $V(L) > V(R)$, that is, if and only if $V(L) - V(R) > 0$. However, to accommodate the empirical 'fact' that agents make errors when choosing, Hey and Orme interpret this as *measurement error*, suggesting that actual decisions are taken on the basis of whether $V(L) - V(R) + \epsilon > 0$ where ϵ is a *measurement error*. Obviously to make this operational one needs to specify the distribution of ϵ: it is fairly natural to specify its mean as being zero (assuming no left or right bias in the agent's answers) and possibly reasonably acceptable to assume that it has a normal distribution (appealing to the Central Limit Theorem). We call this the 'White Noise' (**WN**) error story. The magnitude of the error variance σ^2 can be taken as a measure of the magnitude of the error spread: the larger is σ the greater in general will be the measurement error.

2 Describing True Preferences

Although this paper is concerned with the appropriate modelling of the errors implicit in subjects' responses, this can not be carried out (at least with the data available to us) independently of the subjects' true preferences. Indeed to truly guarantee that we had correctly identified the true error structure we would have to know that we had also correctly identified the subjects' true preferences. But of course we do not know these true preferences. Therefore, we are forced to *jointly* fit an error story and a story of preferences. Unfortunately, there is no concensus as to what might be the best story of preferences - so we are obliged to fit a number of possible stories - and let the data tell us which might be best.

We fitted the following. We omit details as these are readily available elsewhere (see, for example, (Hey, 1997)). The two letter abbreviation is used to identify the functional. All the functionals, with the exception of Regret theory, are *holistic* and hence postulate a preference functional $V(.)$ which is used to rank prospects: hence $\mathbf{p}$ is preferred to $\mathbf{q}$ if and only if $V(\mathbf{p}) > V(\mathbf{q})$. The theories differ in their specification of $V(.)$[3] We shall limit the

[3]We normalise throughout with $u(x_1) = 0$ and $u(x_4) = 1$.

descriptions to the case relevant for the experiments described in this paper - where there are just 4 outcomes: x_1, x_2, x_3 and x_4, with respective probabilities p_1, p_2, p_3 and p_4.

rn : *Risk Neutrality* - subjects choose on the basis of expected value.

$$V(\mathbf{p}) = p_1x_1 + p_2x_2 + p_3x_3 + p_4x_4 \tag{1}$$

eu : *Expected Utility* - subjects choose on the basis of expected utility.

$$V(\mathbf{p}) = p_2u(x_2) + p_3u(x_3) + p_4 \tag{2}$$

da : *Disappointment Aversion* - subjects choose on the basis of expected (modified) utility - where utility is modified *ex post* to take account of any disappointment or delight experienced (see (Gul, 1991)).

$$V(\mathbf{p}) = \min(W_1, W_2, W_3) \tag{3}$$

where

$$W_1 = \frac{(1+\beta)p_2u(x_2) + (1+\beta)p_3u(x_3) + p_4}{1 + \beta p_1 + \beta p_2 + \beta p_3} \tag{4}$$

$$W_2 = \frac{(1+\beta)p_2u(x_2) + p_3u(x_3) + p_4}{1 + \beta p_1 + \beta p_2} \tag{5}$$

and

$$W_3 = \frac{p_2u(x_2) + p_3u(x_3) + p_4}{1 + \beta p_1} \tag{6}$$

pr : *Prospective Reference* - subjects choose on the basis of a weighted average of the expected utility calculated using the correct probabilities and the expected utility calculated using equal probabilities for all the non-null outcomes (see (Viscusi, 1989)).

$$V(\mathbf{p}) = \lambda(p_2 u(x_2) + p_3 u(x_3) + p_4) + (1-\lambda)(a_2 u(x_2) + a_3 u(x_3) + a_4) \quad (7)$$

where $a_i = |a_i|/(a_i n(\mathbf{p}))$ and $n(\mathbf{p})$ is the number of non-zero elements in $\mathbf{p}$.

ri : *Regret with Independence* - subjects choose on the basis of expected (modified) utility where utility is modified *ex post* to take account of any regret of rejoicing experienced and where the two prospects are considered to be independent of each other. Here the preference functional is not holistic, and we have instead the decision rule; $\mathbf{p}$ is preferred to $\mathbf{q}$ if and only if

$$\begin{aligned}(p_2q_1 - p_1q_2)\psi(x_2,x_1) + (p_3q_1 - p_1q_3)\psi(x_3,x_1) + (p_4q_1 - p_1q_4)\psi(x_4,x_1) + \\ (p_3q_2 - p_2q_3)\psi(x_3,x_2) + (p_4q_2 - p_2q_4)\psi(x_4,x_2) + (p_4q_3 - p_3q_4)\psi(x_4,x_3) > 0 \quad (8)\end{aligned}$$

where $\psi(.,.)$ is the Regret function.

rp : *Rank dependent with Power weighting function* - subjects choose on the basis of expected utility where the (cumulative) probabilities are distorted by a weighting function which takes the power function form.

$$V(\mathbf{p}) = w(p_2+p_3+p_4)u(x_2) + w(p_3+p_4)(u(x_3)-u(x_2)) + w(p_4)(1-u(x_3)) \quad (9)$$

where $w(.)$ is the *power* function $w(p) = p^\gamma$.

rq : *Rank dependent with Quiggin weighting function* - subjects choose on the basis of expected utility where the (cumulative) probabilities are distorted by a weighting function which takes the form advocated by Quiggin amongst others (see (Quiggin, 1982)).

$$V(\mathbf{p}) = w(p_2+p_3+p_4)u(x_2) + w(p_3+p_4)(u(x_3)-u(x_2)) + w(p_4)(1-u(x_3)) \quad (10)$$

where $w(.)$ is the *'Quiggin'* function[4] $w(p) = p^\gamma/[p^\gamma + (1-p)^\gamma]^{(1/\gamma)}$.

qu : *Quadratic Utility* - subjects choose on the basis of a utility function which is *quadratic* in the probabilities (rather than *linear* as in the case of Expected Utility theory).

$$\begin{aligned} V(\mathbf{p}) = p_1^2\psi(x_1,x_1) + 2p_2p_1\psi(x_2,x_1) + 2p_3p_1\psi(x_3,x_1) + 2p_4p_1\psi(x_4,x_1) \\ + 2p_2^2\psi(x_2,x_2) + 2p_3p_2\psi(x_3,x_2) + 2p_4p_2\psi(x_4,x_2) + \\ p_3^2\psi(x_3,x_3) + p_4p_3\psi(x_4,x_3) + p_4^2\psi(x_4,x_4) \end{aligned} \quad (11)$$

wu : *Weighted Utility* - subjects choose on the basis of expected weighted utility.

$$V(\mathbf{p}) = \frac{w_2p_2u(x_2) + w_3p_3u(x_3) + w_4p_4}{w_1p_1 + w_2p_2 + w_3p_3 + w_4p_4} \quad (12)$$

The generalisations of Expected Utility theory (**da**, **pr**, **ri**, **rp**, **rq** and **wu**) involve extra parameters over and above those specified by **eu**. In the context of the experiment from which the data for this note comes: **da**, **pr**, **rp** and **rq** all have one extra parameter; **da** has Gul's β parameter[5]; **pr** has Viscusi's λ parameter[6]; **rp** and **rq** have the weighting function's γ parameter. Further, in the context of the experiment reported in this note, **ri** has 5 extra parameters and **wu** 2 has extra parameters. These obviously influence the numbers of *degrees of freedom* involved with the fitting of these functionals.

3 The Experimental Data

We use the data[7] reported in (Hey and Orme, 1994): this consists of the individual responses of 80 subjects to 100 pairwise choice questions on two separate occasions, separated by a period of several days. We call the data from the first occasion Data Set 1, that from the second occasion Data Set 2, and the combined data from both occasions Data Set

[4] See (Quiggin, 1982).
[5] See (Gul, 1991).
[6] See (Viscusi, 1989).
[7] The data - along with further data reporting the time taken by subjects to answer each question - is available on request.

3. The responses of the subjects were either preference for one of the two choices in the pairwise choice or indifference between them. Subjects were motivated on each occasion by knowing that after expressing preference (or indifference) on all 100 questions, one of the 100 questions would be selected at random and the choice that they had said they preferred on that question would be played out for real and the subject paid accordingly; if the subject had expressed indifference on the chosen question then the experimenter would choose one of the two choices at random to be played out. The 100 questions on the second occasion were the same 100 as on the first occasion, though the order was randomised and left and right interchanged. For further details, see (Hey and Orme, 1994).

4 Fitting the Data

The procedure used for fitting the various preference functionals combined with the White Noise error story to the data is as described in (Hey and Orme, 1994): essentially the fitted model is either a linear or a non-linear ordered logit (or probit) model. We wrote our own software to carry out the maximum likelihood estimation. For fitting the various preference functionals combined with the Constant Probability error story to the data, a different procedure was required. Moreover, since, for the purposes of this paper we are fitting the various functionals *subject by subject* rather than *across all subjects* as in (Harless and Camerer, 1994), a procedure also different from theirs is required. Let us be more specific: following Harless and Camerer we assume a constant probability of making a mistake on any one pairwise choice question and we assume independence across questions. Now consider a particular preference functional with parameter vector α and suppose for a particular subject and a particular data set that the preference functional with parameter vector α predicts correctly[8] n of the subject's actual choices and predicts incorrectly the remaining $(N-n)$ questions (where N is the total number of questions in that data set). Then the

[8]We should state what we mean by this: (a) if the functional predicts Left (Right) is preferred then if the individual reports Right (Left) preferred or the two indifferent, then this is a mistake; (b) if the functional predicts strict indifference then Left or Right or Indifference is a correct response.

log-likelihood for this subject on this data set is:

$$LL(\alpha) = nln(\theta) + (N-n)ln(1-\theta) \tag{13}$$

Note that n will, in general, depend upon α. The value of θ that maximises this, for given n, is simply $\theta^* = n/N$ and the implied maximised value of the log-likelihood is then:

$$LL(\alpha) = nln(n) + (N-n)ln(N-n) - Nln(N) \tag{14}$$

Clearly, for $n < N/2$ the smaller is n the higher is $LL(\alpha)$. So, *in order to maximise the log-likelihood with this error model one should choose the paramer vector α so as to minimise the number of incorrect predictions*; on reflection, this is as one would expect.

Minimizing the number of incorrect predictions is precisely the same as *maximising the score* (Manski, 1975). We thus conclude that fitting the Constant Probability error story is achieved by use of the Maximum Score estimation procedure. Unfortunately, the score function is *not* smoothly concave in the parameter vector; on the contrary, the function to be maximised is a *step* function in the parameter space. Accordingly, standard maximisation routines (such as those found in the MAXLIK applications package associated with GAUSS) will not work - they require smoothly concave log-likelihood functions. There are a number of routines suggested in the case of non-smoothly concave functions - none of which are perfect. One that seems better than most is the Simulated Annealing algorithm implemented in GAUSS by E.G. Tsionas. This is what we used; we are most grateful to Dr Tsionas for making it available. We should however note that this algorithm is not *guaranteed* to find the maximum[9] - our results should be interpreted in that light.

5 Analysing the Results

We fitted each of the nine preference functionals listed above combined with each of the two error specifications to the data for each of the 80 subjects on each of the three data sets -

[9] An *exact* routine that is so guaranteed is being developed by Marie-Edith Bissey - see (Bissey, 1997).

giving a grand total of 4320 fitted models. We have parameter estimates for the respective preference functions for these 4320 fitted models. However, we do not report these[10] nor do we report the log-likelihoods for each of the 4320 models. Instead we give some relevant summary statistics. We start with Table 1 which simply gives, for each preference function and for each Data Set, the number of subjects (out of 80) for which each error story fits the data better than the other - by which we mean has the highest (maximised) log-likelihood. So, for example, the Risk Neutral preference functional with the White Noise error story fits the data better than the same preference functional with the Constant Probability error story for 28 (of the 80) subjects on Data Set 1 for 40 (of the 80) subjects for Data Set 2 and for 28 (out of 80) subjects for Data Set 3. It is clear from Table 1 that the question of the 'best' error story can not be answered independently of the question of the 'best' preference functional. However, there is the difficulty that any answer to the second question must first resolve the issue of the appropriate treatment of differing degrees of parsimony in the various models: put simply, the fitting of the different models involves differing numbers of fitted parameters and clearly those models which involve more parameters are bound to fit better and therefore need to be penalised in some fashion. There is no obviously correct way of doing this though one way that seems to command reasonable approval is that suggested by Aikake's Information Criterion. As we have discussed its merits elsewhere (Carbone and Hey, 1994), we will not rehearse the arguments further here. In essence this requires the penalisation of the goodness of fit of a model with k parameters simply by subtracting k from the maximised log-likelihood. Doing this provides us with 'corrected' log-likelihoods which can then be used to rank the various fitted models. Tables 2, 3 and 4 provide a summary of the consequences of carrying out such a ranking exercises. In these Tables, we list, for each Data Set, the number of subjects for whom each fitted model comes 1st, 2nd, 3rd, 4th or 5th in this ranking. The Tables also give an average ranking for each fitted model; clearly the lower the average ranking the better that fitted model performs - in an across-subjects analysis. It will be seen from these Tables that the Constant Probability error story (combined with some appropriate functional) generally fits better than the White Noise story. Particularly interesting is the very strong performance of the

[10]Though they are available on request.

rq[11] model combined with the Constant Probability error story.

This aggregate analysis hides some interesting individual subject effects, however: it is clear, if we look at the results subject by subject, that there are many subjects for whom the Constant Probability error story generally fits better than the White Noise story and there are many for whom the reverse is true. In other words, when we rank the models on the basis of the corrected log-likelihood, there are many subjects for whom a large number of Constant Probability specifications appear in the first 9 (out of 18) places, and many for whom the converse is true. Consider, for example, subjects 1 through 4 on Data Set 1. We have the following orderings based on the Corrected Log-Likelihood:

Subject 1 qu- rp- rq- da- wu- eu- pr- ri- rq+ pr+ eu+ da+ rp+ wu+ ri+ qu+ rn- rn+

Subject 2 pr+ rq+ wu+ rp+ eu+ da+ ri+ qu+ rq- qu- wu- rp- eu- da- pr- ri- rn+ rn-

Subject 3 rn- wu- rq- da- rp- eu- pr- qu- ri- ri+ qu+ wu+ rq+ pr+ rp+ rn+ eu+ da+

Subject 4 rq- wu- rp- eu- pr- qu- ri- da- rn- da+ qu+ rp+ wu+ ri+ rq+ pr+ eu+ rn+

In this the first two letters indicate the preference functional and '-' and '+' indicate which error story: '-' indicates the Constant Probability error story and '+' the White Noise error story. So, for example, for Subject 1, specifications involving the Constant Probability error story occupy 8 of the first 9 places; whilst for Subject 2, specifications involving the White Noise error story also occupy 8 of the first 9 places. With Subjects 3 and 4, the separation is even more complete: all 9 of the first 9 places are occupied by Constant Probability specifications and the last 9 places occupied by White Noise specifications. To give some indication of the average positions of the two error specifications, we calculate the mean position for each specification; this yields, for the four subjects above (where the first number is the average position for the White Noise specification and the second the mean position for the Constant Probability specification) the following:

- Subject 1: 13.1 5.9
- Subject 2: 5.9 13.1

[11]The Rank dependent preference functional with Quiggin weighting function.

- Subject 3: 14.0 5.0
- Subject 4: 14.0 5.0

One can do a formal statistical test of the null hypothesis that the 9 values of the corrected log-likelihood for each of the two error specifications appear in a random order: obviously, under this null hypothesis the expected mean position for each error specification is 9.5; significant departures in either direction indicate non-randomness. The results of such a test are given in Table 5. It is clear that the numbers of subjects for whom the test is significant is *considerably* greater than the expected number (namely 4 at 5% and 0.8 at 1%). The message seems to be very clear: for many subjects the Constant Probability story is 'best'; for many others the White Noise story is 'best'. Trying to get *one* error story as 'best' for all subjects would appear to be seriously misleading.

6 Conclusions

We could go further: it could be argued on the basis of the above evidence that each stochastic specification should be matched with the"right" preference functional. That is, that the well-specified fitted preference representation is a *joint* specification of the preference functional *and* the stochastic specification. This suggests that the preference functional and the stochastic specification are both part of the same decision process. Indeed, it seems reasonable to argue that the way people make mistakes *and* their 'underlying' deterministic decision rule both depend on the way that they process the decision problem. Hence, it could be argued, that if the decision rule chosen by a certain subject could be represented by the Expected Utility preference functional, he or she should make mistakes in a way that is not inconsistent with this preference functional. This suggests that for each individual the best fitting error model depends on the preference functional - and that the best fitting preference functional depends upon the stochastic specification. For example, consider the Rank Dependent model. Here there are three components: (1) the utility function; (2) the weighting function; and (3) the way the two are combined to produce the final preference functional. Randomness could enter into any of these three: there may be noise in the utility function, in the weighting function, or in the calculation of the functional. These would depend on the basic decision process that the individual was using. And these would determine the 'overall' stochastic specification of the complete model. But if all that was true this would suggest that the person who tried to act as an EU maximizer should have some stochastic specification whilst someone who processed decision problems in the way suggested by Rank Dependent EU would have another (and different) stochastic specification. In other words, and crucially, *the error is related to the way one chooses*. If this is true then the way that the preference functional and the stochastic specification are linked, through the underlying decision process, should be the next object of our investigations.

These considerations seem to take us very close to the way that *psychologists* view and model decision making. It may therefore useful to consider whether our approach, as it is evolving here, could learn from the psychological approach. Or more brutally, whether, in fact, we have simply ended up where the psychologists ended up some time ago.

At the risk of excessive simplification, we could summarise the psychology approach as based on a description of the decision *process*. This considers that the decision depends on the context, on the weight of each alternative, on various framing effects and so on. A useful reference is (Payne et al., 1993) in which a detailed discussion of all these elements, and others, can be found. Additional material can be found in (Slovic, 1995). A good illustration of this kind of approach is given by Prospect Theory (Kahneman and Tversky, 1979), in which it is postulated that decision-makers first edit the various prospects before evaluating them. More generally, the psychologists' approach to the modeling of decision making views it as a *process*, a process which is sensitive to the way that the decision problem is posed (framed), and in which the decision-maker breaks down the problem into a series of stages, in each of which some heuristic is invoked. These heuristics can take various forms, depending on the stage of the decision process. Examples of such heuristics include editing and evaluation, as in Prospect Theory (Kahneman and Tversky, 1979). The heuristics used could evolve through time in such a way as to make the decision-maker *adaptive* in the sense used by (Payne et al., 1993). At a first glance this whole approach seems far removed from the approach used by economists, though the latter could simply be regarded as a special case of the former, in which only an evaluation heuristic is invoked, and in which no adaption occurs.

The psychological literature makes frequent reference to potential errors in decision making, though there is little formal modeling of such errors. Given that the decision problem is solved in a sequence of stages it is clear that error could enter at any stage of the process. For example, errors could occur as a failure of the adaptivity process; being adaptive requires knowledge and ability to execute strategies. Deficits in either of these two categories can lead to the failure of the adaptivity process and therefore to an imprecise decision (a decision with error). Another example is errors in a measurement stage of the decision process: the decision-maker, when measuring something relevant to the decision process (for example a probability or a utility) measures it with error, just as when a decision-maker is asked to measure some physical entity - like the physical dimensions of some object. However, vary rarely is consideration given in the psychological literature to a formal modeling of this error process.

One exception to this general rule can be found in (Shugan, 1980), in which is proposed a "cost of thinking" theory. Shugan's basic story of decision making, restricted to a theory of binary choices, is that the decision maker makes a number of comparisons of the two choices. Further, he argues that the number of such comparisons depends upon three elements:

1) the difference in the mean utilities of the two options;

2) the confidence level at which the decision must be made;

3) the perceptual complexity in comparing the two options.

The basic idea behind this model is that people will continue to sample binary differences until the confidence that one option is best reaches a desired level; that is, until the probability of making a mistake falls below some critical value. This model is interesting in that it seems to provide a story behind the Harless and Camerer constant probability error assumption. Indeed, if one views the calculation of the the 'difference in the mean utilities of the two options' as being essentially the calculation necessary to determine the option with the highest expected utility, it would seem that we have arrived precisely at Expected

Utility combined with a Constant Probability error. Does this provide the necessary bridge between the economists and the psychologists' apparently conflicting approaches?

To an economist, unfortunately not. There is a logical flaw in the argument: if the number of comparisons that the decision-maker makes depends upon the 'difference in the mean utilities of the two options', then the decision-maker needs to know this difference in order to calculate the number of comparisons that he or she must make. But if the decision-maker knows this difference, then he or she already knows which option yields the highest expected utility, in which case, if the decision-maker wants to act as an expected utility maximiser, he or she can immediately, and without error, choose the best option. The error disappears. Moreover, there is the 'infinite regress' problem with any story that tries to incorporate a cost of thinking: if indeed there is a cost to thinking in general, and indeed a cost of thinking about the decision problem, there must also be a cost of thinking about thinking about the decision problem, and a cost to thinking about thinking about thinking about the decision problem, and so on. In such a world there can be no such thing as an optimal decision.

This brings us to the heart of the problem and the key source of conflict between economists and psychologists: the economist likes to build theories of decision making on top of a set of axioms of 'rational' behaviour. Almost by definition an 'error' must be irrational (and therefore difficult for an economist to model within their current paradigm) - unless it is chosen optimally. But for certain categories of error (such as those arising from a cost of thinking), the infinite regress argument is sufficient to count out optimally chosen errors. This perhaps leaves us with exogenous errors - that is, errors that are determined independently of the behaviour of the decision-maker. These are however singularly uninteresting, particularly so in the context used in the experiments discussed in this paper. Indeed, it could be argued, by economists and (some) psychologists alike, that if the magnitudes of the payoffs were to be increased then the frequency of errors would decrease. Moreover it is very clear from numerous experiments that there are more errors on certain kinds of questions than on others. The inference is clear: in practice, decision-makers do appear to adapt their behaviour, as psychologists would argue. In fact, economists would agree - justifying their agreement by arguing that on certain types of question the relative costs and benefits of trying to reduce the error are different from those on other questions. But economists often fail to recognise that this explanation conflicts with their basic underlying theory - as we have argued above. Unless these costs and benefits are specifically included in the theory, such arguments are an *ad hoc* modification of the theory, conflicting with its very structure, which assumed the satisfaction of certain axioms of rationality, and hence the absence of error in any genuine sense. We have reached a logical impasse.

So where does this leave us? In limbo, it would appear, half way between the economists' approach and the psychologists'. To get out of this limbo, we need to return to considerations of precisely what it is that we are trying to do. More importantly we need to recognise that economists and psychologists have different research agendas and different objectives. While psychologists may want to provide a detailed description of the process by which decision-makers arrive at decisions, economists are usually more pre-occupied with providing theories of decision making that are predictively useful. Moreover economists are usually much less interested in predicting individual behaviour, and more interested in predicting aggregate behaviour, at some appropriate level of aggregation. Economists, for good methododogical reasons, are often content with 'as if' explanations[12], as long as they have good predictive power. This latter is the key: if the error in some 'as if' theory is small and apparently genuinely random then there are good reasons for accepting the 'as if' theory if alternative theories are cumbersome to apply and add little to the predictive content. Nevertheless it is important - from an econometric point of view - to model the error process in a descriptively adequate way, for otherwise the econometric tests and estimates may be biased and lack power. This was the purpose of this paper. The fact that the error modeling depends upon the preference function does not detract from its usefulness. Indeed it provides an important warning that the search for the correct error specification and the search for the correct functional specification may not be independent.

It is clear that the psychologists' approach can help the economist search for appropriate error specifications - if the economic models can be interpreted using the psychologists'

[12] That is, explanations of behaviour based on behavioural assumptions which are not directly verifiable or verified.

perspective. For example, in the Rank Dependent model discussed above, we as economists could conceive of this model as the outcome of a process in which the decision-maker first ranks the various outcomes (possibly with error), then assigns utilities to these outcomes (possibly with error), then assigns cumulative probabilities to the outcomes (possibly with error) and then uses these various components to evaluate the prospects (possibly with error). In contrast, in the Expected Utility model we could conceive the choice process as involving just one stage - an evaluation stage. Thus different models have different numbers of stages and different kinds of stages - evaluation stages, ranking stages, comparison stages, and so on. The psychologists have expertise to inform us as to the types of errors that are typically present in each of these types of stages, and the economist can then incorporate these into his or her search for the descriptively-best error specification. For example, evidence from experiments conducted by psychologists tends to suggest that measurement error is white noise error, with a variance that typically depends upon the prospects (and which can therefore be estimated). It is clear from this discussion that the final error depends crucially, as we have argued above, on the number of decision stages and their type. So a different error structure may well be present with an Expected Utility maximiser (who goes through just one - evaluation - stage) as compared with a Rank Dependent Utility maximiser (who goes through three or four stages).

So the economist can learn from the psychologists. At the same time they should not abandon their basic approach - which is to postulate that decision-makers *try* to optimise something (possibly through an adaptive process). Without this, the economists would abandon one of the greatest strengths of their paradigm. Moreover, it is should not be forgotten that an 'as if' theory, even if not descriptively *precisely* accurate, may still have relatively strong predictive power (if the error is relatively small) and that is ultimately what the economist is interested in.

References

Bissey, M. E. (1997). Semi-parametric estimatiom of preference functionals. unpublished.

Carbone, E. and Hey, J. D. (1994). Discriminating between preference functionals - a preliminary Monte Carlo study. *Journal of Risk and Uncertainty*, 8:223–242.

Gul, F. (1991). A theory of disappointment aversion. *Econometrica*, 59:667–686.

Harless, D. and Camerer, C. (1994). The predictive utility of generalized expected utility theories. *Econometrica*, 62:1251–1290.

Hey, J. D. (1997). Experiments and the economics of individual decision making. In Kreps, D. M. and Wallis, K. F., editors, *Advances in Economics and Econometrics*, pages 171–205. Cambridge University Press.

Hey, J. D. and Orme, C. D. (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 decision under risk. *Econometrica*, 47:263–291.

Manski, C. (1975). Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics*, 3:205–228.

Payne, J., Bettman, J., and Johnson, E. (1993). *The Adaptive Decision Maker*. Cambridge U.P.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior and Organization*, 3:323–343.

Shugan, S. M. (1980). The cost of thinking. *Journal of Consumer Research*, 24:99–111.

Slovic, P. (1995). The construction of preferences. *The American Psychologist*, 50:364–371.

Viscusi, W. K. (1989). Prospective reference theory: Towards an explanation of the paradoxes. *Journal of Risk and Uncertainty*, 2:235–263.

Table 1: Summary of Winners for each Preference Functional

Preference Functional	White Noise error			Constant Proby. error		
	Set 1	Set 2	Set 3	Set 1	Set 2	Set 3
rn	28	40	28	52	40	52
eu	49	48	46	31	33	34
da	52	61	58	28	19	22
pr	55	57	53	25	24	27
ri	60	72	63	20	8	17
rp	37	42	45	43	39	35
rq	24	24	32	56	57	48
qu	25	37	42	55	43	38
wu	34	41	42	46	40	38

Table 2: Numbers of times each model is in first 5 positions; and average position. Data set 1.

Preference Functional	Error Specification	Position					Mean Position
		1st	2nd	3rd	4th	5th	
rn	WN	0	0	1	1	1	16.56
eu	WN	5	5	2	5	3	9.09
da	WN	0	1	1	3	10	9.95
pr	WN	5	9	6	6	3	7.44
ri	WN	2	5	4	4	3	8.72
rp	WN	1	3	6	9	8	8.55
rq	WN	5	9	9	4	8	7.06
qu	WN	1	3	6	2	2	10.14
wu	WN	1	1	6	5	4	8.24
rn	CP	6	3	3	2	0	13.25
eu	CP	1	1	5	5	6	9.85
da	CP	0	1	3	6	2	11.74
pr	CP	0	1	0	0	5	11.67
ri	CP	0	0	0	0	0	14.29
rp	CP	7	6	10	3	10	7.05
rq	CP	27	7	7	9	5	4.42
qu	CP	10	9	5	8	5	6.90
wu	CP	9	16	6	8	5	6.07

Table 3: Numbers of times each model is in first 5 positions; and average position. Data set 2.

Preference Functional	Error Specification	Position					Mean Position
		1st	2nd	3rd	4th	5th	
rn	WN	0	1	0	0	0	16.61
eu	WN	2	5	4	4	3	8.64
da	WN	2	3	7	4	5	8.27
pr	WN	3	11	9	7	7	6.71
ri	WN	2	5	8	5	7	8.09
rp	WN	2	6	4	5	5	8.35
rq	WN	4	10	7	4	7	6.60
qu	WN	3	6	5	4	4	8.87
wu	WN	6	2	3	6	10	7.59
rn	CP	3	2	0	0	0	15.11
eu	CP	1	4	1	10	6	9.90
da	CP	0	3	1	4	4	12.11
pr	CP	0	1	2	1	1	11.60
ri	CP	0	0	0	0	1	14.99
rp	CP	4	4	8	8	3	8.13
rq	CP	33	3	6	8	5	4.50
qu	CP	8	7	5	6	4	7.55
wu	CP	7	7	10	4	8	7.38

Table 4: Numbers of times each model is in first 5 positions; and average position. Data set 3.

Preference Functional	Error Specification	Position 1st	2nd	3rd	4th	5th	Mean Position
rn	WN	0	0	1	1	0	16.86
eu	WN	2	4	4	4	5	9.39
da	WN	1	4	2	5	8	8.99
pr	WN	8	11	8	4	2	6.94
ri	WN	5	5	6	10	2	8.16
rp	WN	2	4	2	7	7	8.55
rq	WN	6	9	8	4	10	6.84
qu	WN	5	6	8	2	4	8.21
wu	WN	2	3	4	9	8	7.95
rn	CP	7	3	1	0	2	14.02
eu	CP	4	0	7	2	7	9.46
da	CP	0	2	0	4	1	12.26
pr	CP	0	2	2	3	5	10.84
ri	CP	0	0	0	0	0	14.32
rp	CP	4	5	11	3	5	7.97
rq	CP	25	7	7	5	4	5.16
qu	CP	4	8	5	8	3	7.55
wu	CP	5	10	3	9	7	7.51

Table 5: Numbers of subjects for whom test of random order hypothesis is significant.

Data Set	Sig. level	In favour of White Noise	In favour of Const. Proby.	Total Significant
1	5%	19	23	32
	1%	8	18	26
2	5%	25	10	35
	1%	9	6	15
3	5%	32	21	53
	1%	19	18	37