The common consequence effect: testing a unified explanation of recent mixed evidence

Steven J. Humphrey *
School of Economics, University of Nottingham, Nottingham, NG7 2RD, UK
Received 13 July 1998; received in revised form 14 April 1999; accepted 29 April 1999

Abstract
The common consequence effect is perhaps the most investigated violation of expected utility theory, and yet no single generalised expected utility model convincingly organises the experimental data. This paper reports a test of a modified version of expected cardinality-specific utility theory with boundary effects [Neilson, W.S., 1992, Economics Letters 39, 275–278] which can explain some recent mixed evidence without recourse to subjective probability weighting models. The results do not convincingly support the boundary effect hypothesis, and neither support the more usual probability weighting arguments, but do indicate instability in preferences largely consistent with event-splitting effects. ©2000 Elsevier Science B.V. All rights reserved.

JEL classification: D81; C91

Keywords: Allais paradox; Common consequence effect; Boundary effects; Experimental economics

1. Introduction
The common consequence effect introduced by Allais (1953) is no doubt the most famous experimental violation of the expected utility axioms. It stimulated vast literature that produced many other violations and inspired various alternatives to the theory that could explain these apparent anomalies. In this paper I test an explanation of the common consequence effect inspired by Neilson’s (Neilson, 1992) hypothesis that individuals have preferences over the number of consequences in a lottery due to a dislike of complexity. My concern is with similar preferences but with an alternative explanation of their origin.

* Tel.: +44-115-951-5472; fax: +44-115-951-4159.
Studies by Conlisk (1989), Sopher and Gigliotti (1990) and Harless (1992) show that if lotteries with different numbers of outcomes are changed so that they contain identical numbers of outcomes then violations of expected utility theory diminish. Neilson interprets these data as indicating a preference for fewer probable outcomes and formalises his interpretation as the boundary effect hypothesis. Event-splitting effects reported by Starmer and Sugden (1993) and Humphrey (1995) are also consistent with preferences over numbers of outcomes, but suggest that preferences are over numbers of identical non-zero outcomes. Neilson’s model does not distinguish between numbers of different and identical outcomes and so cannot accommodate both common consequence and event-splitting effects. Humphrey’s (Humphrey, 1998a) modification of the model, however, can accommodate both, and provide an explanation of mixed evidence regarding the common consequence effect. This paper reports an experiment which discriminates between Neilson’s and Humphrey’s explanations of the common consequence effect, and investigates potential ambiguities in the nature of preferences over numbers of outcomes to explore the generality of boundary effects.

2. The common consequence effect

2.1. The mixed-fan hypothesis

Fig. 1 illustrates a general common consequence decision problem in a state-contingent payoff matrix.

In pairwise choice between $S$ and $R$ the former lottery offers a higher probability of a positive outcome, but lower expected value, and the latter lottery offers a lower probability of a positive outcome, but higher expected value. Money consequences are given by $a$, $b$ and $c$ where $a > b > c = 0$. There are five states of the world which occur with probabilities $p$–$t$ where $0 < p \leq q < 1$, $0 < r,s,t < 1$ and $p + q + r + s + t = 1$. Since the last three states of the world involve common consequences, the outcome experienced under these states is independent of the lottery chosen. Expected utility theory, therefore, considers choice between $S$ and $R$ to be independent of $r$, $s$ and $t$. The experiment reported here is concerned the four common consequence problems described in Table 1.

Within this set of four problems, three types of common consequence problem pairs are considered; horizontal common consequence pairs involve comparisons of problems 1 and 2 and problems 1 and 3, vertical pairs involve problems 2 and 4, and north-west pairs involve problems 3 and 4. Allais’s original common consequence effect entails a violation of independence embodied in the conjunction of preferences $S_1 R_3$ over problems 1 and 3.

<table>
<thead>
<tr>
<th></th>
<th>$p$</th>
<th>$q$</th>
<th>$r$</th>
<th>$s$</th>
<th>$t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S$</td>
<td>$b$</td>
<td>$b$</td>
<td>$c$</td>
<td>$b$</td>
<td>$a$</td>
</tr>
<tr>
<td>$R$</td>
<td>$a$</td>
<td>$c$</td>
<td>$c$</td>
<td>$b$</td>
<td>$a$</td>
</tr>
</tbody>
</table>

Fig. 1. The general common consequence problem.
Table 1

<table>
<thead>
<tr>
<th>Problem</th>
<th>Probabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( p = q )</td>
</tr>
<tr>
<td>2</td>
<td>( p = q )</td>
</tr>
<tr>
<td>3</td>
<td>( p = q )</td>
</tr>
<tr>
<td>4</td>
<td>( p = q )</td>
</tr>
</tbody>
</table>

and is often described by the ‘fanning-out’ of indifference curves in the unit probability triangle developed by Machina (1982).

In the unit probability triangle each of the three consequences is placed at a vertex, starting with the largest at the top and moving downwards with the smallest at the bottom right-hand corner. The axes measure the probabilities of the largest and smallest consequences, with the probability of the intermediate consequence given by unity minus the other two. Any line joining two points in the triangle represents a pairwise decision problem, with \( R \) being vertically closest to the \( a \) vertex and \( S \) closest to the bottom edge. Expected utility theory implies the existence of upwards sloping, linear, parallel indifference curves within the triangle with a north-westerly direction of increasing preference. Allais preferences (\( S_1 R_3 \)) can be explained if indifference curves ‘fan-out’, since \( S_1 \) (a double-boundary point offering \( c \) for sure) lies on a higher indifference curve than \( R_1 \) and the reverse is true in problem 3.

A prominent feature of experimental tests of the common consequence effect is that fanning-out does not describe behaviour in all areas of the triangle. Camerer (1995, p. 636) points out that an overall picture of indifference curves stemming from common consequence experiments would need to incorporate fanning-out along the bottom edge, but for pairs of problems located along the left-hand edge this seems to be reduced or reversed (to become fanning-in). Chew and Waller (1986), for example, report bottom-edge fanning-out which diminishes as problems are moved north-west, and Conlisk discovers a similar pattern but with fanning-in in the north-west corner of the triangle. This general pattern of behaviour has become known as ‘mixed-fanning’. In terms of the common consequence problem pairs illustrated in Fig. 2, mixed-fanning would imply fanning-out over horizontal comparisons.
and fanning-in over vertical and north-west comparisons. The observation of mixed-fanning casts doubt over generalisations of expected utility theory which seek to explain observed violations by allowing the universal fanning-out of indifferences curves throughout the triangle.\(^1\) Camerer (1995, p. 636) suggests that non-linear probability weighting models, such as Kahneman and Tversky’s (Kahneman and Tversky, 1979) prospect theory, provide the best explanation of mixed-fanning. The main aim of the experiment reported here is to discriminate between Neilson’s and Humphrey’s accounts of mixed-fanning. Since neither of these models involve subjective probability weighting, a brief contextualization of the experiment is warranted in terms of how probability weighting according to prospect theory accommodates this kind of behaviour.\(^2\)

2.2. Prospect theory and the mixed-fan hypothesis

Prospect theory assumes a binary choice between prospects \(P_i\) \((i = 1,2)\), with each prospect comprising a probability vector \(p_{ij}\) \((j = 1, \ldots, n)\) representing the probability that \(P_i\) yields consequence \(x_j\). Individuals choose the prospect assigned the highest overall value \(V(.)\) according to the decision rule in expression (1).

\[
\max_{i=1} \sum_{j=1}^{n} \pi(p_{ij})u(x_j) \tag{1}
\]

In expression (1), \(\pi(p_{ij})\) is the subjective decision weight attached to probability \(p_{ij}\), with \(\pi(0) = 0\) and \(\pi(1) = 1\). The utility function \(u(.)\) is assigned to increments or decrements of wealth relative to a reference wealth position, is unique up to multiplication by a positive constant with \(u(0) = 0\) at the reference point. If prospect theory is applied to the decision problem in Fig. 1, the overall values for each prospect are given by:

\[
V(S) = \pi(t)u(a) + \pi(p + q + s)u(b) + \pi(r)u(c) \tag{2}
\]

\[
V(R) = \pi(p + t)u(a) + \pi(s)u(b) + \pi(q + r)u(c) \tag{3}
\]

Note that in prospect theory \(u(c) = 0\) and so, normalising \(u(a) = 1\), the choice between \(S\) and \(R\) is governed by the sign of the decision rule in expression (4):\(^3\)

\[
S \succ R \iff [\pi(p + q + s) - \pi(s)]u(b) + [\pi(t) - \pi(p + t)] \geq 0 \tag{4}
\]

2.2.1. Horizontal common consequence problems

Horizontal common consequence problems involve moving along the bottom-edge of the triangle in Fig. 2 from Problem 1 towards the \(e\) vertex. This is achieved by increasing probability \(r\) and correspondingly reducing \(s\). Since expression (4) does not include \(r\), only

---

1 Theories in this category include those forwarded by Fishburn (1983) and Machina.
2 Rank-dependent theory (Quiggin, 1982; Yaari, 1987) is a probability weighting model which has attracted much attention in the context of the common consequence effect. Wu and Gonzalez (1998), for example, use this theory to explain choices in common consequence problems. Prospect theory is used here since it provides an account of the event-splitting effects with which this paper is also concerned.
3 Where ‘\(>\)’ denotes strict preference and ‘\(\sim\)’ denotes indifference.
the first term in square brackets is affected by this manipulation. As Starmer (1992) shows, if, as is generally assumed, the decision weighting function is convex, the reduction in $s$ will reduce the first term in square brackets and render the left-hand side of expression (4) smaller. This generates a tendency for choices to switch from $S > R$ to $R > S$ and the subsequent observation of horizontal fanning-out. If the less usual assumption of a concave weighting function is true, then prospect theory yields the opposite prediction of horizontal fanning-in.

2.2.2. Vertical common consequence problems

Vertical movements in the triangle involve increasing probability $t$ and correspondingly reducing $s$. Assuming convexity, reducing $s$ will cause the first term in square brackets in (4) to decrease as above. Since the second term in square brackets is less than zero, increasing $t$ causes this term to decrease and (4) to decrease further, generating the tendency for choices to switch from $S > R$ to $R > S$, and the observation of vertical fanning-in. If $\pi(.)$ is concave then vertical fanning-out is predicted.

2.2.3. North-west common consequence problems

North-westerly movements in the triangle involve increasing probability $t$ and correspondingly reducing $r$. This will impact on (4) through the second term in square brackets as in Section 2.2.2 above, and generate north-west fanning-in through the tendency for choices to switch from $S > R$ to $R > S$. Concavity will yield the opposite prediction of north-west fanning-out. Thus, assuming a convex weighting function, prospect theory provides an explanation of mixed-fanning.

2.3. The robustness of the mixed-fan hypothesis

Just as theories that predict universal fanning-out are refuted by observed mixed-fanning, theories which allow mixed-fanning are cast under some doubt by recent evidence which questions the robustness of horizontal (bottom-edge) fanning-out. Mixed evidence is reported by Chew and Waller (1986), Cubitt et al. (1998) and Camerer (1989) fail to replicate the common consequence effect in bottom-edge problem pairs, and Prelec (1990) and Starmer (1992) observe significant bottom-edge fanning-in. Thus, despite there being a wealth of theoretical accounts of behaviour in common consequence problems located in different areas of the unit probability triangle, it seems that no single theory provides an adequate explanation of all the evidence. For example, prospect theory could explain Prelec’s and Starmer’s (Prelec, 1990; Starmer, 1992) observation of horizontal fanning-in if the decision weighting function was assumed to be concave, but concavity is inconsistent with Starmer’s (Starmer, 1992) observation of vertical and north-west fanning-in within the same parameter set. Interestingly, Starmer’s (Starmer, 1992) evidence is singled-out by Wu and Gonzalez (1998 p. 126, note 5 and p. 131) as being particularly anomalous in that it, unlike the vast majority of other evidence compiled over the last 20 years, does not sit easily with their characterisation of the weighting function. A potential resolution to these anomalies lies in an explanation of mixed-fanning without recourse to probability weighting.
3. Preferences over consequence frequencies

3.1. The boundary effect hypothesis

Conlisk reinterprets Tversky and Kahneman’s (Tversky and Kahneman, 1986) ‘certainty effect’ explanation of the Allais paradox as a ‘boundary effect’. Since certainties involve only one consequence, they lie on a double-boundary point in the triangle diagram and this may impart special attractiveness in relation to single-boundary lotteries which entail more than one consequence if individuals value the simplicity of a sure-thing. Conlisk tests this hypothesis by displacing the Allais lotteries such that they lie marginally inside the triangle boundary and reports reduced incidences of, and no longer systematic, violations of expected utility theory. The importance of the boundary-interior distinction is substantiated by Harless and Camerer (1994). They conclude that since moving from the boundary to the interior of the unit probability triangle usually entails adding a small probability of an outcome to a lottery, the poor fit of expected utility theory on the boundary and its vast improvement on the interior lends support to the importance of the non-linear weighting of small probabilities in explaining observed decision-making behaviour. By contrast, Neilson’s interpretation of boundary effects provides an explanation of this evidence and allows mixed-fanning by emphasising the importance of adding a low probability consequence to a lottery in terms of the consequence itself, and not the perception of its associated likelihood.

Neilson formalises the boundary effect hypothesis within an expected cardinality-specific utility model. \( P \) is the set of probability distributions over the set of outcomes \( X = \{x_1, ..., x_m\} \), measured as gains and losses from reference wealth; \( p \) is an element of \( P \), and the number of outcomes in \( p \) assigned a strictly positive probability is \( n(p) \). Expected cardinality-specific utility maximisation postulates a set of utility functions \( u_n(x) \) such that \( p \) is preferred to \( q \) if and only if:

\[
\sum_{i=1}^{m} p_i u_{n(p)}(x_i) \geq \sum_{i=1}^{m} q_i u_{n(q)}(x_i) \quad (5)
\]

The boundary effect hypothesis stipulates \( u_1(x) > \ldots > u_m(x) \) for all \( x > 0 \), \( u_1(0) = \ldots = u_m(0) \), \( u_1(x) < \ldots < u_m(x) \) for all \( x < 0 \), and is interpreted as a preference for fewer outcomes. Expected utility theory arises as a special case where \( u_1(x) = \ldots = u_m(x) \) for all \( x \in X \).

3.1.1. Horizontal common consequence problems

An application of (5) to problems 1 and 3 outlined in Section 2 yields expressions (6) and (7), respectively:

\[
S_1 \sim R_1 \iff (p + q + s)u_1(b) - pu_3(a) - (1 - p - q)u_3(b) \geq 0 \quad (6)
\]

\[
S_3 \sim R_3 \iff (p + q)u_2(b) - pu_2(a) \geq 0 \quad (7)
\]

The certainty effect entails the over-weighting of certain outcomes, such as \( b \) under \( S_1 \) in Fig. 2, in relation to probable outcomes, such as \( b \) under \( S_3 \) in Fig. 2.
To observe the Allais paradox \((S_1R_3)\) the left-hand side of (6) must be positive and the left-hand side of (7) negative. On this basis, summing across (6) and (7), noting that \(p + q + s = 1\) in expression (6), yields expression (8):

\[
p[u_2(a) - u_3(a)] + [u_1(b) - (p + q)u_2(b) - (1 - p - q)u_3(b)] > 0 \quad (8)
\]

Since the sum of the coefficients on \(u_2(b)\) and \(u_3(b)\) in the second term in square brackets equals minus one, the boundary effect hypothesis renders it positive and expression (8) will hold. Similarly, if (5) is applied to problems 1 and 2, horizontal fanning-out requires \(S_1R_2\) and expression (9) to hold:

\[
\frac{1}{2}(1 - p - q)[u_1(b) - u_2(b)] > 0 \quad (9)
\]

Expression (9) holds by the boundary effect hypothesis and Neilson’s model explains the Allais paradox and, more generally, horizontal fanning-out.

### 3.1.2. Vertical and north-west common consequence problems

If (5) is applied to problems 2 and 4 and problems 3 and 4, then vertical \((S_2R_4)\) and north-west \((S_3R_4)\) fanning-in respectively require expressions (10) and (11) to hold:

\[
\frac{1}{2}(1 + p - q)[u_2(a) - u_3(a)] + \frac{1}{2}(1 + p + q)[u_2(b) - u_3(b)] > 0 \quad (10)
\]
\[
\frac{1}{2}(1 + p - q)[u_2(a) - u_3(a)] + (p + q)[u_2(b) - u_3(b)] > 0 \quad (11)
\]

Expressions (10) and (11) hold by the boundary effect hypothesis, and along with (9) show Neilson’s model to explain mixed-fanning.

### 3.2. Event-splitting effects

#### 3.2.1. Event-splitting effects and prospect theory

Event-splitting effects have been observed by Starmer and Sugden (1993) and Humphrey (1995) over pairs of decision problems of the following general form; where \(a, b\) and \(c\) are money consequences \((a > b > c \geq 0)\) occurring in states of the world with associated probabilities \(p\) and \(q\) \((p, q > 0\) and \(1 - p > p)\):

Three-state problem:

- \(S_3: c, p; b, q; 1 - p - q\)
- \(R_3: a, p; c, q; 1 - p - q\)

Two-state problem:

- \(S_2: c, p; b, 1 - p\)
- \(R_2: a, p; c, 1 - p\)

Each decision problem entails identical probabilities of identical outcomes, and so expected utility theory requires individuals to choose either \(S_3R_2\) or \(R_3S_2\) \(^5\). An event-splitting effect occurs if there are significantly more \(S_3R_2\) than \(R_3S_2\) choices. Since the only difference between the two decision problems is that the latter state of the world \((probability \ 1 - p)\) in the two-state problem is split into two states of the world \((probabilities q + (1 - p - q) = 1 - p)\)

\(^5\) Assuming indifference is not allowed.
in the three-state problem, event-splitting effects can be interpreted as indicating a preference for the higher number of (identical) \( b \) outcomes in \( S_3 \) in relation to \( S_2 \). \(^6\)

The prevailing explanation of event-splitting effects is a stripped-down version of prospect theory. Implicit in standard prospect theory is the assumption of combination which, by postulating that the probabilities associated with disjoint events which contain identical consequences will be added for simplification prior to the evaluation of the prospects, renders the three-state problem identical to the two-state problem and thereby rules-out event-splitting. If, however, standard prospect theory is stripped of combination such that \( \pi(.) \) is applied to each disjoint event separately, irrespective of whether there are two (or more) disjoint events containing the same outcome, an application of expression (1) to the two and three state problems yields expression (12).

\[
\pi(q)\upsilon(b) + \pi(1 - p - q)\upsilon(b) > \pi(1 - p)\upsilon(b)
\]  

(12)

Expression (12) will hold and event-splitting effects are explained if, as Kahneman and Tversky (1979, p. 281) assume, \( \pi(.) \) is subadditive in the relevant region.

Although Neilson’s model incorporates preferences over numbers of outcomes, its inability to accommodate event-splitting effects is attributable to the lack of distinction between preferences over numbers of identical and numbers of different outcomes. It may, however, be premature to discount Neilson’s model on the basis of event-splitting effects alone. First, although standard prospect theory allows mixed-fanning it needs to be stripped-down to explain event-splitting effects. Second, stripping-down prospect theory to explain event-splitting not only rules-out mixed-fanning, but rules-out any kind of fanning. \(^7\)

Third, it may be possible to modify Neilson’s model to shed light on both the common consequence effect and event-splitting effects.

### 3.2.2. Modified expected cardinality-specific utility theory

Humphrey’s (Humphrey, 1998a) modification to Neilson’s theory differs only in that the utility function employed to evaluate lotteries depends on the number of outcomes in a lottery of a particular type. Outcomes are assumed to be coded as gains or losses from a reference wealth level \( r_p \) where \( r_p = \min x \) if \( n(p) > 1 \) and \( r_p = 0 \) if \( n(p) = 1 \), and \( s(p) \) is the number of outcomes \( x > r_p \) listed in the representation of the decision problem to which \( p \) assigns a positive probability. \(^8\)

The decision rule in expression (5) then becomes:

\[^6\] Event-splitting effects are also consistent with an aversion to the higher frequency of the lowest outcome \( c \) in \( R_3 \) in relation to \( R_2 \). This interpretation might better accord with the loss-aversion literature (if the lowest outcome is coded as a loss). Starmer (1999), however, reports event-splitting effect driven non-transitive choices over sets of three decision problems which are consistent with the interpretation of event-splitting effects as a preference for the higher number of \( b \) outcomes, and cannot be explained by an aversion to the higher number \( c \) outcomes.

\[^7\] This can be seen by inspecting expression (4). Without combination the left-hand side of expression (4) becomes; \( [\pi(p) + \pi(q) + \pi(s) - \pi(s)]\upsilon(b) + [\pi(t) - \pi(p) - \pi(t)] \). The terms in \( s \) and \( t \) cancel and so the manipulations in \( r, s \) and \( t \) which describe movements in the unit probability triangle do not, according to stripped-down prospect theory, generate violations of expected utility theory.

\[^8\] Kahneman and Tversky (1979, p. 274) suggest that the coding of outcomes is influenced by lottery representation. If, for example, the evaluation of a lottery with two different outcomes involves more effort than evaluating a lottery represented with two identical outcomes, the reference point may be altered in the former case from current wealth to the smallest outcome to save the effort of one comparison.
Applying (13) to the three-state and two-state decision problems yields (14) and (15), respectively:

\[
S_3 \succ R_3 \Leftrightarrow (1 - p)u_2(b) - pu_1(a) > 0 < (14)
\]

\[
S_2 \succ R_2 \Leftrightarrow (1 - p)u_1(b) - pu_1(a) > 0 < (15)
\]

To observe an event-splitting effect \(S_3 \succ R_2\) condition (16) must hold:

\[
(1 - p)[u_2(b) - u_1(b)] > 0 (16)
\]

Since expression (16) requires \(u_2(b) > u_1(b)\), it is contrary to the boundary effect hypothesis. If, however, the boundary effect hypothesis is reversed such that \(u_1(x) < \ldots < u_m(x)\) for all \(x > 0\), \(u_1(0) = \ldots = u_m(0)\), and \(u_1(x) > \ldots > u_m(x)\) for all \(x < 0\), Humphrey’s (Humphrey, 1998a) model can accommodate event-splitting effects. The reversed boundary effect hypothesis is termed the frequency effect hypothesis 9. Given the premise of preferences over numbers of outcomes and the interpretation of event-splitting effects as indicating exactly that, the failure of Humphrey’s (Humphrey, 1998a) model with boundary effects to accommodate event-splitting might be regarded as a mixed result. It is not clear, however, that the source of the mixed result is the boundary effect hypothesis. This can be illustrated by applying (13) to the common consequence problems from Section 2.

3.2.3. Preferences over numbers of identical outcomes and the common consequence effect

An application of (13) to problem pairs 1–3, 1–2, 2–4 and 3–4, respectively, yields expressions (17), (18), (19) and (20) which are required to hold to observe mixed-fanning

\[
p[u_1(a) - u_2(a)] + (1 - p - q)[u_1(b) - u_2(b)] > 0 (17)
\]

\[
\frac{1}{2}(1 - p - q)[u_1(b) - u_2(b)] > 0 (18)
\]

\[
\frac{1}{2}(1 + p - q)[u_1(a) - u_2(a)] + \frac{1}{2}(1 + p + q)[u_1(b) - u_2(b)] > 0 (19)
\]

\[
\frac{1}{2}(1 + p - q)[u_1(a) - u_2(a)] + (p + q)[u_1(b) - u_2(b)] > 0 (20)
\]

Expressions 17–20 all hold by the boundary effect hypothesis and so Humphrey’s (Humphrey, 1998a) model with boundary effects accommodates mixed-fanning in a similar manner to Neilson’s original model 10. If the plausibility of the boundary effect hypothesis in common consequence problems is accepted, it is unclear whether the appropriate foundational

---

9 The frequency effect hypothesis might be regarded as one way in which the non-procedurally optimizing use of a frequency heuristic can be incorporated in a utility maximizing framework to capture the separate influences of stimulus frequency and stimulus probability reported by Humphrey (1999).

10 Note that if the boundary effects are replaced by frequency effects the opposite inequalities hold and the reverse predictions are made.
model is Neilson’s or Humphrey’s (Neilson, 1992; Humphrey, 1998a) version of expected cardinality-specific utility theory. It is important to discriminate between the two since the latter implies choices to be sensitive to minor changes in problem representation and, if different common consequence experiments have employed different descriptions of decision problems, this may account for some of the mixed results.

### 3.3. Mixed evidence of the common consequence effect

The mixed evidence with which this paper is concerned is reported by Starmer (1992), who employs a system of common consequence problems as outlined in Section 2. In the 13 tests he conducts Starmer (1992) observes no significant fanning-out, but eight instances of significant fanning-in.\(^{11}\)\(^{12}\) He concludes that the data would be best organised by a theory which predicts universal fanning-in, but rules this out as a sustainable hypothesis on the basis of the apparent robustness of horizontal fanning-out in previous experiments. Starmer’s (Starmer, 1992) data are also inconsistent with Neilson’s original boundary effect hypothesis which predicts mixed-fanning. Most tests of the common consequence effect employ decision frames where each *different* outcome in a lottery is represented only once. Starmer (1992), however, uses a state-contingent matrix display where each state of the world (as in Fig. 1) with a non-zero probability is represented as a column in the matrix containing the appropriate consequence. Figs. 3 and 4 illustrate problem 1 as it would appear in a state-contingent display and a ‘standard’ display, respectively.

Although problem representation doesn’t influence the predictions of the boundary effect hypothesis in Neilson’s original model, the potential divergence of predictions under Humphrey’s (Humphrey, 1998a) model is illustrated by Table 2, which lists the utility function used in the models for each decision problem under a state contingent matrix display.

Following from Table 2, expressions (21)–(24) provide the predictions of Humphrey’s (Humphrey, 1998a) model for the state-contingent matrix display. These expressions are the

\[^{11}\text{Three significant observations from five tests of horizontal fanning-in, two from three vertical fanning-in, and three from five north-west fanning-in.}\]

\[^{12}\text{Starmer (1992) uses small, but real, monetary payoffs whereas many previous tests have employed the large hypothetical payoffs with the extreme probabilities of the original Allais example. MacCrimmon and Larsson (1979) discover stronger effects when extreme probabilities and payoffs are used.}\]
Table 2
Utility functions under ECSU and modified ECSU

<table>
<thead>
<tr>
<th>Problem</th>
<th>Standard ECSU</th>
<th>Modified ECSU</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Option S</td>
<td>Option R</td>
</tr>
<tr>
<td>1</td>
<td>$u_1(.)$</td>
<td>$u_3(.)$</td>
</tr>
<tr>
<td>2</td>
<td>$u_2(.)$</td>
<td>$u_3(.)$</td>
</tr>
<tr>
<td>3</td>
<td>$u_2(.)$</td>
<td>$u_2(.)$</td>
</tr>
<tr>
<td>4</td>
<td>$u_3(.)$</td>
<td>$u_2(.)$</td>
</tr>
</tbody>
</table>

* ECSU refers to Neilson’s model and modified ECSU is Humphrey’s (Humphrey, 1998a) model.

Table 2
Utility functions under ECSU and modified ECSU

counterparts of expressions (17)–(20) deriving from the standard display, and are expressed in terms of requirements for mixed-fanning:

$$|u_3(b) - u_2(b)| > 0 \quad (21)$$

$$\frac{1}{2}(1 - p - q)[u_2(b) - u_2(b)] > 0 \quad (22)$$

$$\frac{1}{2}(1 - p - q)[u_2(a) - u_3(a)] + |u_3(b) - u_2(b)| > 0 \quad (23)$$

$$(p + q)[u_2(b) - u_3(b)] + \left[\frac{1}{2}(1 + p - q)u_2(a) - \frac{1}{2}(1 - p - q)u_3(a) - pu_1(a)\right] > 0 \quad (24)$$

Expressions (21) and (22) will not hold by the boundary effect hypothesis, but will hold by the frequency effect hypothesis. Starmer (1992), however, does not observe horizontal fanning-out, but horizontal fanning-in. An explanation of this observation requires the reverse inequalities in expressions (21) and (22). In this case the boundary effect hypothesis within Humphrey’s (Humphrey, 1998a) model explains horizontal fanning-in in the state-contingent matrix display. The situation is less clear for vertical and north-west comparisons. In expression (23) the terms in square brackets operate in opposite directions and so vertical fanning-in is consistent with boundary effects over outcome $a$ dominating frequency effects over $b$. By contrast, vertical fanning-out (the reverse inequality) is consistent with the frequency effects over $a$ dominating boundary effects over $b$. There is a similar ambiguity in the second term in square brackets in expression (24) relating to north-west comparisons. Nevertheless, expressions (21)–(24) make it clear that boundary effects can explain universal fanning-in in state-contingent matrix displays and the general observation of mixed-fanning in standard displays. The essence of the experiment described below is to establish whether this is a plausible explanation.

4. Experiment

4.1. Design

The experiment was designed around two sets of the four common consequence problems with parameters according to Table 3. The payoffs involved in the decision problems are
larger than those employed by Starmer (1992), but maintain a similar relationship between the expected values of the safer and riskier options.

The experiment employed two conditions. The control condition faced subjects with decision problems displayed with the outcome frequency in each option exactly as in Starmer (1992). Here Neilson’s model predicts mixed-fanning and Humphrey’s (Humphrey, 1998a) model with boundary effects allows universal fanning-in. The discrimination condition used displays where the frequency with which each outcome is represented in a lottery is varied as described in Table 4.

An application of Humphrey’s (Humphrey, 1998a) model (expression 13) to the discrimination condition problems yields no predicted violation of expected utility theory for horizontal problem pairs. Therefore, if Starmer’s (Starmer, 1992) horizontal fanning-in is explained by Humphrey’s (Humphrey, 1998a) model with boundary effects, horizontal fanning-in should be observed under control condition and not under the discrimination condition. For vertical and north-west comparisons an application of (13) yields (25) and (26), which must respectively hold to observe vertical and north-west fanning-in.

\[
\frac{1}{2}(1 + p - q)(u(a) - u_2(a)) > 0 \quad (25)
\]

\[
\frac{1}{2}(1 + p - q)(u_1(a) - u_2(a)) > 0 \quad (26)
\]

Expressions (25) and (26) demonstrate that the ambiguity under the control condition (expressions 23 and 24) is removed. If vertical and north-west fanning-in is observed under both conditions it cannot be due to frequency effects within Humphrey’s (Humphrey, 1998a) model, but can be explained by boundary effects. Table 5 summarises the predictions of each model. Convincing support for the boundary effect hypothesis would require the observation of vertical and north-west fanning-in, with discrimination between Neilson’s and Humphrey’s (Neilson, 1992; Humphrey, 1998a) models provided by horizontal compar-

---

**Table 3**

<table>
<thead>
<tr>
<th>Parameter set 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probabilities</td>
</tr>
<tr>
<td>( p )</td>
</tr>
<tr>
<td>0.15</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter set 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probabilities</td>
</tr>
<tr>
<td>( p )</td>
</tr>
<tr>
<td>0.2</td>
</tr>
</tbody>
</table>

---

**Table 4**

<table>
<thead>
<tr>
<th>Option</th>
<th>Display variation (relative to control condition)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control</td>
</tr>
<tr>
<td>( S_1 )</td>
<td>( p + q )</td>
</tr>
<tr>
<td>( S_2 )</td>
<td>( p + q )</td>
</tr>
<tr>
<td>( R_3 )</td>
<td>( p )</td>
</tr>
<tr>
<td>( S_4 )</td>
<td>( p + q )</td>
</tr>
<tr>
<td>( R_4 )</td>
<td>( p + t )</td>
</tr>
</tbody>
</table>

* Under parameter set 1 the split is 0.08 and 0.07, and under parameter set 2 the split is 0.1 and 0.1.
Table 5
Model predictions

<table>
<thead>
<tr>
<th>Display</th>
<th>EUT&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Pt&lt;sup&gt;b&lt;/sup&gt;</th>
<th>ECSU&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Modified ECSU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horizontal</td>
<td>None</td>
<td>FO</td>
<td>FO</td>
<td>FO</td>
</tr>
<tr>
<td>Vertical</td>
<td>None</td>
<td>FI</td>
<td>FI</td>
<td>FI</td>
</tr>
<tr>
<td>North-west</td>
<td>None</td>
<td>FI</td>
<td>FI</td>
<td>FI</td>
</tr>
</tbody>
</table>

<sup>a</sup> Expected utility theory. Under parameter set 1 the split is 0.08 and 0.07, and under parameter set 2 the split is 0.1 and 0.1.

<sup>b</sup> The predictions for prospect theory are conditional on the convexity assumption. If concavity is assumed the opposite predictions hold.

<sup>c</sup> The predictions for ECSU (Neilson’s model) are reversed if boundary effects (BEH) are replaced with frequency effects (FEH). Modified ECSU refers to Humphrey’s (Humphrey, 1998a) model.

4.2. Implementation

The experiment was conducted in a final year undergraduate microeconomics class, which had received no prior instruction regarding the theories being tested, at the University of Nottingham during May 1998. Participants were provided with a set of instructions, two envelopes (marked ‘A’ and ‘B’) each containing a booklet of ten decision problems (each page displayed one problem), and a consent form. There were two sets of booklets (Set 1 and 2) containing different problems which were allocated randomly to subjects. Each set was such that individuals would not face the same problems under the control and discrimination conditions. They might, however, face problems from the control condition under one parameter set and the same problems in the discrimination condition under the other parameter set. To minimise the likelihood of recognising what was being tested, despite the use of different parameters, the control and discrimination problems were presented in different booklets (booklet A or booklet B). Subjects were required to complete booklet A, place it in an envelope, seal the envelope and sign their name over the seal before proceeding to booklet B. Subjects did not face two problems (in either condition) involved in a direct problem pair comparison. Thus, all comparisons are between-subjects. Within each set of booklets there were two subsets which were identical in all respects, but to control for order effects presented problems in exactly opposite orders.

<sup>13</sup> Copies of the problem booklets can be obtained from the author. Each booklet contained four problems of concern here, with the remaining six problems testing other hypotheses.

<sup>14</sup> Since these tests are between-subjects they are weaker than which would emerge from a within-subject design since, here, it is impossible for any individual to actually violate expected utility theory by making inconsistent choices in a particular problem pair. It does seem reasonable, however, to assume that the subjects are drawn from the same population and, given the random allocation of questionnaires, there is no obvious reason why one would expect the sampling procedure alone to generate significant (one-tailed) violations of expected utility theory.
Fig. 5. Lottery representation.

The experiment was introduced by explaining that the normal class had been replaced by a session to assist in research concerning how individuals make decisions under risk. It was stressed that the session was entirely voluntary, was not a ‘test’, and would not compromise the progression of the class being replaced. Subjects were informed that they would be compensated for their participation with the chance to win cash prizes which would be paid on the spot. Subjects read through their instructions (as the experimenter read them out loud) which explained the task, the progression of the session, and provided opportunity for clarification. When the task was completed, for which there was no time limit, subjects were told they should proceed to a desk at the front of the room where their winnings would be determined. They would then be paid their winnings (if any) and could then leave. The experiment employed the random-lottery incentive system where, at the end of the experiment, each subject randomly selected one of 20 consecutively numbered discs and played-out the question number they selected for real money. It was emphasised that since the real-question would not be determined until after all questions had been answered, any question could be ‘for real’ and so all should be treated as if they were for real money.¹⁵ Decision problems were represented using a state-contingent display as in Fig. 5, with subjects marking their choice in the boxes on the left. The risk was resolved for the real question by drawing a disk from a bag containing one hundred consecutively numbered disks. For example, if the bottom lottery was chosen and the disk drawn was between 1 and 15, the subject would win £12. A total of 69 subjects participated in the experiment, of which approximately 60 percent were male. The session lasted 45 minutes in total (the first subject finished after 25 minutes), and average winnings were £6.12 (at the time equivalent to just under US$ 10).

5. Results

5.1. Boundary effects

The results of the experiment are presented in Tables 6 and 7 for the control and discrimination groups, respectively, and should be read in conjunction with Table 5.

¹⁵ Holt (1986) shows that experiments which employ the random-lottery incentive system may not reveal true preferences. For discussion and experimental evidence which negate this possibility see Starmer (1992) and Starmer and Sugden (1991), respectively.
Table 6
Results: control condition

<table>
<thead>
<tr>
<th>Pair</th>
<th>Parameter set 1</th>
<th>Parameter set 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Responses</td>
<td>Responses</td>
</tr>
<tr>
<td></td>
<td>$S$</td>
<td>$S'$</td>
</tr>
<tr>
<td>Horizontal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[1,2]</td>
<td>20/35 (57%)</td>
<td>29/34 (85%)</td>
</tr>
<tr>
<td>[1,3]</td>
<td>20/35 (57%)</td>
<td>24/33 (73%)</td>
</tr>
<tr>
<td>Vertical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[2,4]</td>
<td>29/34 (85%)</td>
<td>21/35 (60%)</td>
</tr>
<tr>
<td>North-west</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[3,4]</td>
<td>24/33 (73%)</td>
<td>21/35 (60%)</td>
</tr>
</tbody>
</table>

* $S$ is the proportion of subjects choosing the safer option in the former problem in the pair and $S'$ is the proportion of safer choices in the latter problem in the pair. $Z$ is the test statistic for a test based on the normal distribution for difference in sample proportions. $Z > 0$ is consistent with fanning-out and $Z < 0$ is consistent with fanning-in. An asterisk denotes a significant (one-tailed) violation of expected utility theory at the 5% level. The sample size of 69 can be obtained by adding the denominators in each response proportion. Note that for pairs 1–3 and 3–4 under parameter set 1 the sample size is 68 due to one subject failing to answer one problem.

Tests are based on the null hypothesis that there are no violations of expected utility theory so that the proportions of safer option choices should be approximately equal across each decision problem in a comparison. The alternative hypothesis is that there is a common consequence effect manifest in significant differences in the proportions with which the safer option is chosen for each problem in a comparison.

Under the control condition, Table 6 shows the horizontal comparisons under parameter set 1 to be in line with Starmer’s (Starmer, 1992) data. There is no replication of the Allais paradox (pair 1–3), but a tendency for horizontal fanning-in which reaches significance in pair 1–2. Under parameter set 2 the data show a tendency for horizontal fanning-out, but again there is only a significant effect under the 1–2 comparison.

In the discrimination condition, Table 7 reveals significant horizontal fanning-out under

Table 7
Results: discrimination condition

<table>
<thead>
<tr>
<th>Pair</th>
<th>Parameter set 1</th>
<th>Parameter set 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Responses</td>
<td>Responses</td>
</tr>
<tr>
<td></td>
<td>$S$</td>
<td>$S'$</td>
</tr>
<tr>
<td>Horizontal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[1,2]</td>
<td>24/34 (71%)</td>
<td>15/35 (43%)</td>
</tr>
<tr>
<td>[1,3]</td>
<td>24/34 (71%)</td>
<td>14/35 (40%)</td>
</tr>
<tr>
<td>Vertical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[2,4]</td>
<td>15/35 (43%)</td>
<td>25/34 (74%)</td>
</tr>
<tr>
<td>North-west</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[3,4]</td>
<td>14/35 (40%)</td>
<td>25/34 (74%)</td>
</tr>
</tbody>
</table>
Set 1, but not under Set 2. Overall, the horizontal comparisons reveal no general support for boundary effects under either Neilson’s model or Humphrey’s (Neilson, 1992; Humphrey, 1998a) model. Boundary effects under Neilson’s model are cast under doubt by the tendency towards horizontal fanning-in under some treatments and, in particular, significant horizontal fanning-in for Set 1 problem pair [1, 2] in Table 6. Evidence contrary to Humphrey’s (Humphrey, 1998a) model is provided by the observation of significant horizontal fanning-out in Set 1 under the discrimination condition, where no such prediction is made.

For vertical and north-west comparisons under the control condition the data are inconsistent with both Starmer’s (Starmer, 1992) evidence and Neilson’s boundary effect hypothesis. Significant vertical fanning-out is observed under both parameter sets and significant north-west fanning-in is observed under neither. The former observation is contrary to any model which predicts mixed-fanning, but would be consistent with Neilson’s theory if boundary effects were replaced with frequency effects. This would also be consistent with the tendency towards horizontal fanning-in observed under Set 1, but is inappropriate in providing an overall description of the data due to the tendency towards horizontal fanning-out under Set 2. Humphrey’s (Humphrey, 1998a) model with frequency effects allows the vertical fanning-out observed for both parameter sets under the control condition. In terms of expression (23), this would require frequency effects over consequence $a$ to dominate boundary effects over $b$. Yet, whilst frequency effects under Humphrey’s (Humphrey, 1998a) model are consistent with horizontal fanning-out observed under Set 1, it fails to explain horizontal fanning-out observed under Set 2.

The data in Table 7 for vertical and north-west comparisons are equally problematic to organise in terms of boundary effects. The tendency towards universal fanning-in and significant violations of expected utility theory under parameter set 1 are consistent with any theory which predicts mixed-fanning. Whether Neilson’s model is the correct explanation, however, is doubtful. First, Neilson’s model with boundary effects makes identical predictions under both the discrimination and control conditions, and this has been ruled-out above as explaining the data under the latter. Second, there are no significant violations of expected utility theory in parameter set 2 under the discrimination condition. This is contrary to Neilson’s boundary effect hypothesis. Boundary effects might be rescued by Humphrey’s (Humphrey, 1998a) model, since this predicts the observed tendency towards vertical and north-west fanning-in, and allows either fanning-in or fanning-out under the control condition. If this interpretation were true, however, significant horizontal fanning-out should not be observed in either Set 1 under the discrimination condition or in Set 2 under the control condition.

Overall, the data provide little support for either model. There appears to be an instability in preferences, epitomised by the switch from standard mixed-fanning observed in Set 1 under the discrimination condition to the opposite and less usual pattern of mixed-fanning in Set 1 under the control condition. Although the essential difference between the two conditions is the frequency with which particular outcomes are represented in decision problems, it seems that models which accommodate preferences over numbers of outcomes alone are not rich enough to explain behaviour over the common consequence problems employed in this experiment.
5.2. Prospect theory

At first glance, the data appear to be potentially better organised by probability weighting according to prospect theory. Prospect theory with a convex weighting function predicts the mixed-fanning observed in Set 1 under the discrimination condition. Also, whereas the boundary effect hypothesis does not explain the absence of common consequence effects (and tendency towards universal fanning-in) in Set 2, prospect theory would accommodate this if the weighting function is less convex over the range of probabilities in Set 2 than it is in Set 1. As Set 2 probabilities entail a smaller variance than those in Set 1, this explanation is consistent with Kahneman and Tversky’s (Kahneman and Tversky, 1979, p. 283) suggestion that the weighting function becomes less well-behaved near the end points. Moreover, previous findings reported by MacCrimmon and Larsson (1979) and Chew and Waller (1986) suggest that horizontal fanning-out, which requires convexity, is more prevalent under more extreme (greater variance) probabilities.

The support for prospect theory, however, is somewhat weakened by the control condition. Organisation of the data in Table 6 requires the weighting function to be concave for parameter set 1 and convex for parameter set 2. Whilst it is possible that that the greater variance in probabilities under parameter set 1 induces concavity and smaller variance under Set 2 induces convexity, this is contrary to previous findings outlined above and implies exactly the opposite assumption to that necessary to accommodate the mixed-fanning evident under Set 1 in Table 7. Thus, it seems that prospect theory or, more generally, non-linear probability weighting is no better placed to provide an overall description of the data than are boundary effects.

Although the data reported here are generally out of line with those reported by Starmer (1992), perhaps the conclusion that should be reached is not. That is, the data are not particularly well explained by any single model. Of particular concern is that differences in choices between the control and discrimination support neither boundary effects nor the more usual probability weighting argument. It is, however, possible to provide a reasonable organisation of the data in a manner consistent with the premise that the numbers of identical outcomes in lotteries are important to decision-makers.

5.3. Event-splitting effects

The data for Set 1 in Table 7 are consistent with prospect theory with a convex weighting function for set 1 parameters, which drives observed mixed-fanning, but a less (or not) convex weighting function for Set 2 which does not generate violations of expected utility theory. From this starting point, the requirements for an overall organisation of the data are described in Table 8.

Table 9, shows how the proportions of safer options chosen change between the control and discrimination conditions for each individual decision problem. Recall from Table 4 that, for problems 1 and 2, the display variation between the discrimination and control conditions involves joining two states of the world in which outcome \( b \) occurs in the safer option and, for problem 3, splitting one state of the world in which outcome \( a \) occurs in the riskier option. These display alterations are similar to the display changes in the
Table 8
Requirements for overall data organisation

<table>
<thead>
<tr>
<th>Pair</th>
<th>Observed under discrimination</th>
<th>Change required to get</th>
<th>Observed under control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter set 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[1,2]</td>
<td>HFO</td>
<td>Less HFO</td>
<td>HFI</td>
</tr>
<tr>
<td>[1,3]</td>
<td>HFO</td>
<td>Less HFO</td>
<td>No violation</td>
</tr>
<tr>
<td>[2,4]</td>
<td>VFI</td>
<td>Less VFI</td>
<td>VFO</td>
</tr>
<tr>
<td>[3,4]</td>
<td>NWFI</td>
<td>Less NWFI</td>
<td>No violation</td>
</tr>
<tr>
<td>Parameter set 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[1,2]</td>
<td>No violation</td>
<td>More HFO</td>
<td>HFO</td>
</tr>
<tr>
<td>[1,3]</td>
<td>No violation</td>
<td>No change</td>
<td>No violation</td>
</tr>
<tr>
<td>[2,4]</td>
<td>No violation</td>
<td>More VFO</td>
<td>VFO</td>
</tr>
<tr>
<td>[3,4]</td>
<td>No violation</td>
<td>No change</td>
<td>No violation</td>
</tr>
</tbody>
</table>

* HFO and HFI denote horizontal fanning-in and fanning-out, respectively. Similar acronyms apply to vertical and north-west comparisons.

Table 9
Event-splitting effects

<table>
<thead>
<tr>
<th>Problem</th>
<th>Parameter set 1</th>
<th>Parameter set 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control Discrim</td>
<td>Control Discrim</td>
</tr>
<tr>
<td></td>
<td>S S' Z</td>
<td>S S' Z</td>
</tr>
<tr>
<td>1</td>
<td>20/35 24/34</td>
<td>27/34 18/35</td>
</tr>
<tr>
<td>2</td>
<td>29/34 15/35</td>
<td>19/35 22/34</td>
</tr>
<tr>
<td>3</td>
<td>24/33 14/35</td>
<td>23/35 20/34</td>
</tr>
<tr>
<td>4</td>
<td>21/35 25/34</td>
<td>27/34 25/35</td>
</tr>
</tbody>
</table>

* Z is the test statistic for a test based on the normal distribution for difference in sample proportions. An asterisk denotes a significant (one-tailed) difference at the 5% level. Z > 0 corresponds to choices being proportioned according to an event-splitting effect.

event-splitting problem pairs in Section 3.2.1. An event-splitting effect would occur in problems 1, 2 and 3 if there were a significantly larger proportion of safer choices under the control condition than under the discrimination condition. Table 9 reveals that in four of six cases for problems 1, 2 and 3 (problems 2 and 3 under Set 1 and problems 1 and 3 under Set 2) choices are proportioned in the direction consistent with event-splitting (Z > 0). Of these four, three are significant. Of the two event-splitting inconsistent observations (Z < 0) neither are significant.

Starting with parameter set 1, Table 9 shows a significant increase in choices of the safer option for problems 2 and 3 between the discrimination and control conditions, but an insignificant decrease in safer options chosen for problems 1 and 4. The implications of these

16 Note that the display changes here are slightly different from the standard event-splitting problem pairs in that the state of the world is only split in one, as opposed to both, options. The interpretation of event-splitting effects, however, is that they are driven splitting the positive consequence. This is exactly what is occurring in problems 1, 2 and 3. Problem 4 is excluded from this discussion since is involves splitting events containing positive outcomes in both options between the discrimination and control conditions, and so there is no unambiguous prediction of an event-splitting effect.
changes for problem pairs 1–2 and 1–3 should be a net reduction in horizontal fanning-out between the discrimination and control conditions. Table 8 shows this to be exactly what the data reveal and, moreover, the larger of the two increases in choices of the safer option between the discrimination and control conditions (problem 2) actually reverses horizontal fanning-out such that it becomes fanning-in. For the vertical and north-west comparisons under parameter set 1, the difference between the discrimination and control conditions requires less fanning-in. The significant increases in safer choices in problems 2 and 3 coupled with the insignificant decrease in problem 4 (where event-splitting is ambiguous) explain this. The same logic applies to parameter set 2. 17

Thus, although it seems that models premised upon preferences over numbers of outcomes do not convincingly organise the data, there is evidence of subjects having preferences over numbers of outcomes and that this makes a contribution to choices in common consequence problems. The instability in preferences between the control and discrimination conditions can be attributed to significant event-splitting effects over some decision problems and not over others. Why event-splitting effects should only emerge in some comparisons is unclear, particularly in light of previous evidence indicating their robustness over a range of probabilities and outcomes. For example, despite the split events being identical for problems 1 and 2 under Set 1, an event-splitting effect occurs between the conditions in the latter but not in the former. This could be explained by the fact that the option containing the split/un-split event in problem 1 is a certainty, but so too is the safer option in problem 1 in Set 2 where an event-splitting effect does occur. Nevertheless, it is clear that starting from Set 1 under the discrimination condition, a reasonable organisation of the data is provided by non-linear probability weighting which allows mixed-fanning, but less mixed-fanning under the smaller variance probabilities in set 2, plus an account of event-splitting effects.

Explaining experimentally observed behaviour in this manner has a precedent. Starmer and Sugden (1998) report a test of non-transitive choices predicted by regret theory (Loomes and Sugden, 1987). They speculate that evidence supporting this prediction could be explained by problem representations which created the potential for event-splitting effects, and re-run the original tests with appropriately controlled displays. Replication of regret-cycles leads to the conclusion that event-splitting was not driving the original evidence. Importantly, Starmer and Sugden (1998) report a lower incidence of regret cycling in one of their displays than in another. They rule-out event-splitting effect as generating this asymmetry as it exists between two event-splitting-controlled displays. Humphrey (1998b), however, argues that although Starmer and Sugden (1998) are correct that event-splitting cannot explain within-subject regret cycling, it does provide an explanation of between-subject differences in the incidence of non-transitivities. In this respect, Starmer and Sugden’s (Starmer and Sugden, 1998) data is poorly explained by either regret theory or a theory which allows event-splitting effects in isolation, but is organised by regret theory plus event-splitting effects.

17 The only exception is problem pair 2 and 4 in Set 2 where no significant differences between safer choices are observed. Here, acceptance of the null hypothesis under the discrimination condition and vertical fanning-out under the control condition can be interpreted as two individually insignificant changes in proportions of choices contributing to an overall significant vertical common consequence effect.
Whether the contribution outcome frequency makes to decision-making behaviour is best explained by Humphrey’s (Humphrey, 1998a) model with frequency effects or the stripped-down version of prospect theory is unclear. The relevant probabilities being split in the experiment reported here are smaller under Set 1 than under Set 2, and so non-linear weighting of small probabilities might suggest the prevalence of event-splitting effects to be greater under set 1 than under Set 2. Table 9 indicates some inconclusive evidence of this in revealing two event-splitting effects under set 1 and only one under Set 2. Note that whilst Humphrey’s (Humphrey, 1998a) model with frequency effects explains all three significant observations in Table 9, there is no support for boundary effects.

More generally, the fact the event-splitting involves representing both actual outcomes and associated probabilities more frequently creates the potential for both probability weighting and the frequency of outcomes per se to influence choices. Discrimination between these explanations is difficult with the kind of decision-making task reported here. There is, however, evidence to suggest that stimulus frequency is important to decision-makers in the absence of risk and, therefore, where the potential for non-linear weighting of small probabilities is not present. Weber et al. (1988), for example, report attribute-splitting effects in a multi-attribute utility evaluation study which appear related to event-splitting effects. Humphrey (1999) reports further evidence which suggests stimulus frequency exerts an influence on choices independently of stimulus probability. Perhaps the appropriate conclusion is that although models incorporating preferences over numbers of outcomes do not provide an overall explanation of the behaviour observed in this experiment, the influences they seek to capture cannot be ruled out as systematically and predictably contributing to the data independently of non-linear probability weighting.

5.4. Decision error

The appendage of a stochastic term to a probability weighting model which allows mixed-fanning might negate the suggestion that the data are not organised by any single model. Here the interpretation would be that ‘true’ preferences are consistent with non-linear probability weighting, but differences in problem representations cause some choices to be made contrary to true preferences and thereby add noise to the data.18 Three models of stochastic choice have recently attracted some degree of attention in the literature. Harless and Camerer (1994), Hey and Orme (1994) and Loomes and Sugden (1995) propose different methods of generating stochastic versions of deterministic theories. The first two conceptions of decision error view mistakes as being ‘genuine’ in that they stem from miscalculations, slips in recording a choice, and so on. The third conception views the random component of choices to be in preferences themselves. That is, choices are generally made according to a deterministic core theory, such as one which allows mixed-fanning, but rather than there being true preferences plus error, preferences themselves are inherently imprecise.

The two possible conjunctions of preferences which violate expected utility theory over problems 1 and 2 are $S_1R_2$ (conjunction 1) or $R_1S_2$ (conjunction 2). Assume, for example, that true preferences are described by a deterministic probability weighting model which

---

18 This interpretation is due to an anonymous reviewer.
allows mixed-fanning under both conditions. Since, under such a theory, mixed-fanning requires horizontal fanning-out, the probability – denoted \( c(2) \) – of observing conjunction \( 2 \) is zero. If, however, a stochastic element is introduced to the deterministic theory, \( c(2) \) is no longer zero and a degree of horizontal fanning-in might be observed despite true preferences dictating horizontal fanning-out. To explain the data observed in this experiment, however, a theory of stochastic choice would need to provide an account of why \( c(2) \) is greater under the control condition than under the discrimination condition. If errors are ‘genuine’, it is unclear why this asymmetry in decision probability would arise independently of event-splitting effects.\(^{19}\)

If the assumption made in Hey and Orme that decision errors are normally distributed with constant variance is followed, then an organisation of the data is unlikely; the errors, if they are captured by Table 9, are clearly not normally distributed. Similarly, if the error variance was assumed to be a function of the number of outcomes in a decision problem in order to capture the influence of stimulus frequency, it is unclear why errors seem to be occurring in one direction and not the other. It is possible that the distribution mean shifts between the control and discrimination conditions, but when an explanation of this shift is sought, one would be drawn back towards event-splitting effects. Another alternative would be to reject the notion of normally distributed errors, as is implied in Ballinger and Wilcox (1997), in favour of a highly-skewed distribution which would accommodate (from a starting point of mixed-fanning in Set 2 under the discrimination condition) the difference in behaviour between the conditions. More generally, standard econometric techniques would ascribe errors to unexplained causal influences and typically seek to incorporate known causal influences in the deterministic part of the model. In this respect, although recent research on decision error suggests that when variations induced by factors such as different problem representations are controlled, a stochastic element of choice remains, it seems reasonable to suggest that the variation described by Table 9 is systematic, predictable, clearly a (necessary) property of the experimental design and should not be relegated to a stochastic term whatever the structure of that term may be.

There is a point of principle, rather than one of modelling the error term, at stake here. Even if it is accepted that sources of instability in data such as event-splitting effects should not be treated as econometric errors, they might still be treated as decision-theoretic errors if they are interpreted as representing irrational behaviour with little economic meaning. Violations of expected utility theory such as event-splitting effects are, from the perspective of expected utility theory, irrational. But so too are other violations such as the common consequence effect. If an explanation of particular violations is sought in the context of consistent (but non-expected utility) preferences, as is certainly the case for common consequence violations, but not for other violations, then the implication is that some violations are more important than others. A proponent of mixed-fanning might argue that the mixed-fan hypothesis works perfectly well apart from irrational behaviour induced by problem representation, which should be treated as decision error. Thus mixed-fan plus errors is appropriate. Why then, should the proponent of mixed-fanning reject the similar

\(^{19}\) In the Harless and Camerer (1994) model, decision error is analogous to a game theoretic tremble and is assumed to be constant across individuals and decision problems. In this respect, the asymmetry required to organise the data would not be allowed.
argument that expected utility works perfectly well, apart from some irrational behaviour in particular (common consequence) decision problems, and so expected utility plus noise is appropriate? Rejection of this argument would introduce a somewhat arbitrary distinction between ‘reasonable’ and ‘unreasonable’ violations of expected utility theory. Moreover, the evidence mentioned in the previous section which documents the importance of stimulus frequency in both risky and riskless choice, suggests that phenomena such as event-splitting effects do have behavioural meaning. If prediction, description and prescription are important in an economic sense, then stimulus frequency must also have economic importance. Finally, a failure to consider and seek a theoretical account of ‘unreasonable’ violations might entail foregoing alternative, and possibly empirically superior, explanations of the ‘reasonable’ violations.

6. Conclusions

In related studies Starmer (1992) and Conlisk both conclude that their results do not sit easily with standard economic theory. The same is true of the results presented here. The data are not explained by any of the generalised expected utility models which were developed to explain observed violations of expected utility theory in decision-problems of exactly the type used in this experiment. More worrying, perhaps, is that minor changes in problem representation seemingly impart large changes in choices. It is not surprising, therefore, that any single model is descriptively inadequate. Starmer (1992, p. 829) suggests that individual choice behaviour is ‘more subtle and complex’ than decision theorists have generally conveyed in their models. If so, this may render the induction of theories from sub-sets of experimental evidence problematic. If a single theory was to be induced from the data reported here, it would need to incorporate mixed-fanning and event-splitting effects. In light of evidence reported by Humphrey (1999), which suggests that stimulus frequency and stimulus probability exert independent influences on choice, a model which incorporates both non-linear probability weighting and a utility function indexed on numbers of particular types of outcomes might be descriptively appropriate.

This conclusion depends upon the perceived role of theory. If a single theory should explain as much (as parsimoniously) as possible, the volumes of diverse observed influences on decision-making behaviour seemingly condemn this task to inevitable failure. If, however, risky choice is recognised as being too complex to be captured by any single theory and that the role of a single theory is to capture a facet of behaviour in a specific context, then it may be necessary to accept that slightly different contexts will invoke additional facets of behaviour and overall explanations of data will require more than one model. Although the behaviour observed in this experiment might, with sufficient ingenuity, be explained by a single model involving a complex probability weighting function, experience suggests that any such function will be limited in its application to other types of decision problem. Given this unpredictable (lack of) generality, it may be preferable to explain the data relatively easily with any theory which allows mixed-fanning plus an account of event-splitting effects.

There is growing acceptance that preferences are not fixed and immutable as generally assumed in economic theory, but are constructed according to the nature and demands of
the decision-making task at hand. In the context of a constructed preferences framework, the existing spectrum of theory and experimental evidence has a role to play in terms of suggesting what form such preferences might take and when they might arise. In terms of this experiment, it seems that if preferences are constructed the nature of that construction is influenced by outcome frequency in a predictable manner. Although the extent to which this influence has pervaded studies of risky choice other than those discussed here is unclear, it seems uncontroversial to suggest that it should be controlled in future work. The real-world importance of stimulus frequency is illustrated by the activities of marketers and advertisers, and this alone suggests that models incorporating preferences over numbers of outcomes independently of probability weighting are worthy of further investigation.

Acknowledgements

I am grateful for the comments of an anonymous reviewer and the editor. I would also like to acknowledge the assistance of Matt Ayres and Peter Talks in running the experiment, and Peter Wright for allowing the experiment to be conducted in place of one of his lectures. The research reported in this paper was financially supported by the University of Nottingham, award no. 97NLRG865.

References

Humphrey, S.J., 1998b. Non-transitive choice and the prediction of frames, unpublished manuscript (School of Economics, University of Nottingham).
Expected utility hypotheses and the Allais paradox (Reidel, Dordrecht).
Sopher, B., Gigliotti, G., 1990. A test of generalized expected utility theory, Mimeo, (Department of Economics,
Rutgers University).
of Economic Studies 59, 813–830.
42, 139–157.
6, 235–254.
Uncertainty 16, 115–139.