

On preference imprecision

Robin P. Cubitt · Daniel Navarro-Martinez ·
Chris Starmer

Published online: 20 February 2015

© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract Recent research invokes preference imprecision to explain violations of individual decision theory. While these inquiries are suggestive, the nature and significance of such imprecision remain poorly understood. We explore three questions using a new measurement tool in an experimental investigation of imprecision in lottery valuations: Does such preference imprecision vary coherently with lottery structure? Is it stable on repeat measurement? Does it have explanatory value for economic behaviour? We find that imprecision behaves coherently, shows no tendency to change systematically with experience, is related to choice variability, but is not a main driver of the violations of standard decision theory that we consider.

Keywords Preference imprecision · Imprecision intervals · Noisy preferences · Violations of expected utility theory · Construct validity

JEL Classifications D81 · C91 · D01 · D03

1 Introduction

Imagine a setting in which you are asked to compare an object against a range of fixed sums of money. For each such sum, you can say one of three things: “I’m sure I prefer the object”; “I’m sure I prefer the money” or “I’m not sure about my preference”. Someone who responds with “I’m not sure” for some contiguous range of money

R. P. Cubitt (✉) · C. Starmer

School of Economics, The University of Nottingham, Nottingham NG7 2RD, UK
e-mail: Robin.Cubitt@Nottingham.ac.uk

D. Navarro-Martinez

Department of Economics and Business, Universitat Pompeu Fabra, 08005 Barcelona, Spain

D. Navarro-Martinez

Barcelona Graduate School of Economics, 08005 Barcelona, Spain

D. Navarro-Martinez

Barcelona School of Management, 08005 Barcelona, Spain

values has revealed what we will call an *imprecision interval*. This paper presents an experimental investigation of imprecision intervals.

Our primary motivation flows from an accumulating literature (including Cohen et al. 1987; Butler and Loomes 1988, 2007, 2011; Dubourg et al. 1994, 1997; and Morrison 1998) in which procedures like the one just described¹ are used to measure preference imprecision and which provides convincing evidence that individuals will often express imprecision in their preferences, when allowed to do so. Recent contributions to this literature also make the striking claim that preference imprecision may co-vary with, and help to explain, some notable observed behaviours, including particular classic violations of standard individual decision theory. Butler and Loomes (2007) provide the most prominent and the most positive evidence for this. Their findings suggest that imprecision contributes importantly to the notorious preference reversal phenomenon. More recently, Butler and Loomes (2011) argue that imprecision may explain other well-known violations of expected utility theory (EUT) too; and they conjecture (p.521) that preference imprecision might provide a linking explanation of violations of standard decision theory that have typically been seen as separate. This conjecture is important because it conflicts sharply with a long tradition of research, comprising some of the best-known contributions of behavioural economics, which explains violations of standard decision theory using models of precise but non-standard preferences, with different non-standard elements accounting for different violations.² Yet, radical and suggestive as the conjecture is, the direct evidence for it is still limited.³

That said, the idea that imprecision may play a pervasive role in choice behaviour is also suggested by recent theoretical literature on ‘noisy’ preferences. This literature has its roots in stochastic choice models first proposed in the 1950s and 1960s which have been further developed and applied by economists in the last two decades.⁴ Some contributions to this literature refer to ‘imprecision’ as a closely related, or even synonymous, concept to ‘noise’ (for example, Loomes 2005). Yet, notwithstanding such passing remarks, little has been done to explore the relationship between noise and imprecision explicitly. Clearly, noisiness of choice and imprecision of preference might be connected—for example if one is the cause of the other or some underlying psychological mechanism drives both. But, it is also possible that the two concepts, properly understood, are distinct. For example, stochastic elements in decisions might reflect errors in choice (Harless and Camerer 1994; Hey and Orme 1994), while imprecision reflects, say, ‘incommensurability’ or difficulty of reaching a decision.

¹ The similarity is greater in some cases than in others because some of these studies focus on imprecision in a different trade-off than the one between objects and money that we have described.

² See Camerer (1995), Starmer (2000) and Wakker (2010) for three reviews of different vintages of this line of research.

³ For example, the earlier literature does not all support the idea that imprecision contributes importantly to violations of standard decision theory. Dubourg et al. (1994) and Morrison (1998) find little evidence that imprecision explains the well-known gap between willingness-to-pay and willingness-to-accept. Yet, this finding does not undermine the claim derived from Butler and Loomes (2007, 2011) that an important subset of violations of EUT flow from a common influence of imprecision.

⁴ The older literature includes prominent papers by, for example, Quandt (1956), Luce (1958, 1959), Block and Marschak (1960) and Becker et al. (1963). The newer literature was sparked by Harless and Camerer (1994), Hey and Orme (1994), Loomes and Sugden (1995, 1998) and includes Loomes et al. (2002), Hey (2005), Loomes (2005), Blavatsky (2007) and Wilcox (2008).

For these reasons, our study is motivated in part by important open questions about the role of preference imprecision in violations of standard decision theory and in choice variability. However, it is also motivated by a more basic concern about imprecision intervals: that of *construct validity*.

The various tools that have been developed for eliciting imprecision intervals all have a feature that is rather unconventional in experimental economics. They allow subjects to reveal imprecision, but without any explicit material incentive for them to do so accurately. Since there is no obvious way to incentivise revelation of imprecision with material rewards, it seems inevitable that direct measurement of preference imprecision will rely on subjects' self-reports. A sceptic might then wonder whether such 'measurements' are artefacts of the elicitation tools and subjects' good will. For example, if a subject guesses that the point of an experiment is to measure 'imprecision', she might report a bit of it, either side of her perfectly precise preference, if it is not costly to do so.⁵ Then, though she would report imprecision, this would not be very informative about her behaviour when more is at stake or imprecision harder to reveal. Such concerns may explain why still only a few studies in experimental economics have measured imprecision intervals. But, especially given the compelling reasons for being interested in preference imprecision described above, it is important that experimental economists do not jump to conclusions about the potential of un-incentivised measures of it. In other social and behavioural sciences—and even in other areas of economics, such as economics of self-reported happiness (surveyed, for example, by Clark et al. 2008)—material incentives are not seen as a *sine qua non* of construct validity. Rather, the latter is seen as a matter to be investigated empirically. We follow that approach here.⁶

To be more precise, we develop a new tool for measuring preference imprecision and deploy it in an experiment designed to investigate three sets of questions about imprecision intervals, motivated by the above discussion.

The first set concerns *coherence*. If imprecision intervals do not vary comprehensively with the structure of the objects over which preferences are considered, that would detract from their credentials as empirically meaningful entities. Existing literature contains some hints about how imprecision intervals vary with the structure of objects considered, but typically in the form of stylised findings from experiments designed to explore other matters. For example, Butler and Loomes (2007) consider imprecision in relation to simple gambles of the form that each yields some strictly positive prize with some probability, and zero otherwise. When imprecision is measured on a money scale, the interval for a low probability (high payoff) bet is wider than that for a high probability (low payoff) bet. This would be consistent with a more general tendency for imprecision measured in this way to be negatively related to some concept of 'distance' from some particular certainty, but the existing data is not rich enough to assess this more general conjecture.⁷ Another stylised observation, consistent with the evidence reported in Butler and Loomes (2007, 2011) is that, for gambles with common

⁵ That said, it would be an over-simplification to think that experimenter demand effects can only arise with unincentivised tasks. For more discussion of such effects, see Zizzo (2010).

⁶ For further argument that scope for material incentives should not be seen as defining the legitimate territory of experimental economics, see Read (2005) and Bardsley et al., (2010, Ch. 6). For specific discussion of incentivisation in relation to imprecision intervals, see Butler and Loomes (2007, pp. 293–6; 2011, pp. 516–17).

⁷ Butler and Loomes (2007) compare intervals for just one pair of such bets.

support, the width of imprecision intervals, measured on a probabilistic scale, is an approximately constant fraction of the range of responses in the associated choice tasks that would not violate stochastic dominance. This could, however, be a by-product of there being relatively little variation in that range in their tasks.⁸ Finally, Butler and Loomes (2011) report a tendency for imprecision intervals to widen when the best payoff in the support of the gamble is increased but, again, this observation rests on very few data points.⁹ While these patterns and commentary from the previous literature suggest the possibility of coherence, we present the first systematic and sustained attempt to study and separate the effects on imprecision of variation along different dimensions of lottery structure.

Our second set of questions concerns *stability*. It is not far-fetched to suppose that imprecision might change with experience. For example, repeated measurement might alter the amount of deliberation brought to bear on a particular task, which might either increase or reduce imprecision. The more transient imprecision is, the less convincing it would be as an explanation of non-transient preference phenomena. To our knowledge, stability of imprecision intervals has not been considered in previous literature, a gap we address in this paper.

A construct could, in principle, satisfy coherence and stability and yet still be of limited value, if it provided no new insights into any other concerns. Hence, our third set of questions, which concerns explanatory and predictive *value-added*. Clearly, in one paper, we can only consider value-added in relation to a limited domain. We choose the domain of monetary valuation of monetary lotteries for this because, in addition to fitting well with our objectives on stability and coherence,¹⁰ it is suggested by our opening discussion. We study the association between measured imprecision and (i) noise (or variability) in certainty equivalents elicited by standard incentivised techniques; and (ii) violations of EUT analogous (in a sense made precise below) to well-known violations of the independence axiom and its weaker relative, ‘betweenness’. Choice of (i) is motivated by the fact mentioned above that, while the literature on noise refers in passing to imprecision, little systematic exploration of the link between the two concepts has occurred. Choice of (ii) allows us to evaluate whether the concept of imprecision adds usefully to existing explanations of violations of standard decision theory, by studying cases where previous findings are not conclusive.

Whether violations of independence or betweenness are associated with preference imprecision is the main issue addressed by Butler and Loomes (2011). They sketch a theoretical model which predicts such an association, at least for their measure of imprecision, and they report some experimental findings consistent with that prediction. Yet, as Butler and Loomes (2011, p. 521) acknowledge, their data analysis cannot ultimately identify whether imprecision is integral to the violations of EUT they observe or whether imprecision intervals are simply epiphenomenal around stable non-

⁸ Specifically, all their $M_1 - M_4$ lotteries induce the same range of “permissible” (i.e., non-dominated) responses. The variation in that range only comes from comparing these with their M_5 lotteries.

⁹ The observation is based only on comparing imprecision for a set of lotteries where the best payoff is increased from 40 to 60 Australian dollars.

¹⁰ Subjects can easily be faced with the same lottery more than once; and it is straightforward to vary the structure of lotteries in ways that might plausibly be expected to alter the precision of subjects’ valuations of them. The hints from existing literature that we wish to pursue in relation to coherence all derive from papers on monetary lotteries.

EU preferences. For example, being entirely aggregative, their analysis cannot explore whether subjects who are less precise are more prone to violate EUT. Our design allows us to dig deeper and to investigate exactly that question.

Our main findings are as follows. Our measure of preference imprecision varies across lotteries in an intelligible and systematic way, but does not show any sign of systematic change with repetition or experience. We find strong evidence of association between imprecision intervals and variability in incentivised certainty equivalents, but little evidence for association between imprecision and the violations of EUT that we consider. Overall, our results support the view that imprecision is a meaningful construct, with some prospect for aiding understanding, at least of variability in economic behaviour. But, they do not lend weight to the suggestion that imprecision of preference is a unifying explanation, in the sense of being the main driver behind a broad range of violations of standard decision theory.

The rest of the paper is organized as follows: Section 2 explains the experimental design and procedures, Section 3 presents the results, Section 4 sets them in the broader context of research on preference imprecision and Section 5 concludes.

2 Experimental design

2.1 The elicitation mechanism

Our study uses a new tool to elicit, for each subject, imprecision intervals and certainty equivalents for each of various lotteries. Figure 1 illustrates one such lottery in the format used to describe it to participants.

The figure describes a binary lottery with a 60% chance of winning £35 and a 40% chance of winning £17. Participants knew from that start that, at the end of the experiment, they could end up playing one of the lotteries for real. The numbers along the top of the figure refer to the physical randomisation device used to resolve these lotteries. This was a bag containing 100 balls, numbered from 1 to 100. When a subject played a lottery for real, they drew a number from the bag and were rewarded according to the number drawn. So, in the case of the lottery of Fig. 1, they got £35 so long as they drew a number below 61; otherwise, they got £17.

For every lottery presented to them, each subject was required to complete a lottery-specific *response table*. Figure 2 illustrates the response table for the lottery described in Fig. 1. Each row of the first column contains a different sum of money with these increasing, in £0.50 intervals, from the lowest possible lottery outcome in the first row, rising to the highest possible lottery outcome in the final row. The column headings and these features of Fig. 2 applied to all response tables, but the highest and lowest certain sums of money varied across response tables depending on the lottery outcomes.

1	60	61	100
£35		£17	
60%		40%	

Fig. 1 Presentation of lotteries

amounts at which they switch from being sure that they prefer the lottery to being sure they prefer the money provide a measure of their certainty equivalent (CE) valuation for the lottery. In contrast, individuals who are unsure about what they prefer for at least some rows were asked to draw another line in the centre column to confirm the set of money amounts for which they were unsure. The three vertical lines that we have marked in the figure illustrate the case of an imaginary individual whose responses indicate that they are sure they prefer the lottery for any certain amount shown up to and including £23; sure they prefer any certain amount of at least £27; but unsure about which they prefer in between. When an individual reports a positive range of values in the middle column, we interpret them as revealing uncertainty about their CE valuation of the lottery and refer to this range as their measured *imprecision interval* (in this case, from £23.50 to £26.50, so an interval size of £3).

Our experiment included monetary incentives for individuals to reveal CEs for lotteries implemented by using the following easily understood variant of the Becker et al. (1964) procedure. Subjects were told that one of the tables they completed would be selected at random and used to determine the payoff for them in the experiment. Once a table was selected, a second random device would select a row from that table and the subject would receive what they had indicated they preferred in that row.

Subjects were told that in order to be able to complete this procedure, for every response table where they chose to express an imprecision interval, they would also have to indicate a point at which they chose to switch from having the money amount to having the lottery so that we would know how to reward them in the event that application of the BDM procedure chose a row located in one of their imprecision intervals. We asked them to indicate a switch point for each table by drawing a line separating the last row for which they wanted to choose the lottery from the first row for which they wanted to choose the certain amount. We interpret this line as the individual's own best estimate of their CE.¹¹ In Fig. 2, we illustrate the case of an individual who has drawn a horizontal line between the amounts £24 and £24.50; in such a case, we impute a CE of £24.

This elicitation mechanism inherits features from some others used in the literature but, to the best of our knowledge, is novel in bringing them together in one tool. Like us, Cohen et al. (1987) presented subjects with choice-lists in which each row was a choice between some lottery and a certainty, with the latter rising across rows. Besides strict preference, they permitted subjects to express either definite indifference or uncertainty about their preference. For a subject who declared themselves either indifferent or uncertain over a range of certain sums for a given lottery, they constructed the CE as the mid-point of the “indecision interval” obtained by combining the ranges of indifference and uncertainty, rather than having subjects set their switch-points directly, as we do. The analysis reported by Cohen et al. (1987) concerns the CEs constructed by their method, rather than the ranges of uncertainty or indecision per se.

¹¹ Some readers may think that an alternative to requiring each subject to give a switch-point in every response table, even when they had indicated a non-zero imprecision interval, would have been to flip a coin if a row in an imprecision interval is selected, and give the subject either the sure sum or the lottery depending on its outcome. We see this alternative as worse than our procedure because it confounds imprecision with preference for (or against) two-stage lotteries.

There are several other studies, including those of Dubourg et al. (1994, 1997) and Butler and Loomes (2007, 2011), that do investigate properties of preference imprecision, as we do, but which have used iterative procedures, rather than response tables, to identify the boundaries of imprecision.¹² Like our procedure, the monetary equivalence tasks of Butler and Loomes (2007) elicit boundaries of imprecision and a certainty equivalent via responses to a set of pairwise comparisons between lotteries and certain amounts of money and, for each comparison, the respondent can indicate either a clear preference or express doubt.¹³ The feature that CEs are incentivised but there is no material reward for revealing imprecision is also common to their procedures and ours. However, there are two primary differences. One is that Butler and Loomes use a four category response scale with two forms of imprecise response: In each comparison of a gamble and a sure amount in their procedure the options were described as ‘lotteries’ (with one of them being degenerate) and subjects responded by selecting one of four options: “I definitely prefer lottery A”; “I think I prefer lottery A but I’m not sure”; “I think I prefer lottery B but I’m not sure”; “I definitely prefer lottery B”. The second difference, already noted above, is that Butler and Loomes apply an iterative search procedure in which comparisons are presented one at a time.

The particular combination of features we employ in our elicitation device is driven by our research objectives. We think a tabular procedure, as in Fig. 2, is more appropriate for our purposes than an iterative one because a table allows subjects to consider more directly the thing we want them to reveal: their imprecision interval. If a range in which they are uncertain is a meaningful notion for subjects, our format provides a convenient framework for them to express that. It is simpler and more direct to ask subjects to reveal this as one range, rather than as two parts delineated, respectively, by the ranges in which two different imprecision-revealing statements apply. Our procedure also separates out the elicitation processes for imprecision and certainty equivalents, respectively, and places them in a definite order: an individual first identifies a range (if any) in which they are not sure what they prefer; having done this, they reflect on where they will switch from choosing the lottery to the certain amount. Finally, but perhaps most importantly, our tabular presentation facilitates reflection on and adjustment to the whole *set* of responses that a subject might provide in relation to a given lottery. For a subject who really has an imprecision interval, it permits them to reflect on whether they have accurately revealed it and to adjust if they

¹² In presenting their designs, Dubourg et al. (1994, 1997) focus on an elicitation procedure in which discs are rotated iteratively to present a series of tasks one at a time, but Dubourg et al. (1994, p.124) also refer to a tabular presentation of tasks more akin to ours. As with the rotating discs, these presentations elicit prices at which subjects are sure or unsure they would be willing to trade, rather than choices. Hanley et al. (2009) describe a procedure in which an interviewer asks a respondent whether he “would definitely pay” each in an ascending series of sums for an environmental improvement, and then whether he “would definitely NOT pay” each in a descending series, using the responses to build a two-way “payment ladder” in which the ranges “definitely pay” and “definitely NOT pay” may or may not meet.

¹³ Butler and Loomes (2007) also report a probability equivalents treatment which, like Butler and Loomes (2011), measures imprecision intervals on a probabilistic scale. In the latter case, the intervals are potentially constructed from segments on different edges of a probability triangle, meaning that subjects may have to consider variation in one probability followed by variation in another probability, with both these dimensions being combined additively by the analysts (see Butler and Loomes 2011, pp. 515–516).

think necessary. It also allows subjects to adjust their responses to reflect less imprecision if, through the process of completing a task, they come to feel that they are less unsure about some of their choices than they initially felt (for example, if the process of choosing a switch-point reduces initial unsureness).

2.2 The lotteries

In the experiment, we collected response tables for 33 different lotteries which, for the purpose of analysis, we organise into seven sequences. The lotteries and sequences are described in Table 1.

All of the lotteries have either two or three monetary consequences with $x_1 > x_2 > x_3 = 0$, and associated probabilities of p_1, p_2 and p_3 . The first five rows of the table (after the header) show the lotteries in sequence 1, the next five those of sequence 2, and so on. Consequences are given in UK pounds. The first column of the table gives a unique identifier for each lottery where, for example, the label “1.3” indicates lottery 3 in sequence 1, and so on. A letter “R” in the second column indicates a ‘repeated’ lottery: for all such lotteries, the individuals who evaluated them completed two identical response tables, separated by some intervening tasks. A letter “D” in the second column indicates a *degenerate* lottery: we did not gather data for these lotteries since we should expect there to be no imprecision, but we include them in the discussion below as meaningful ‘endpoints’ of various sequences.

Sequences 1 and 2 were designed to investigate how imprecision varies with two dimensions of the structure of consequences: *scale* and *spread*. Moving down Sequence 1, the mixing probabilities are held constant but the consequences fall in regular steps, keeping the distance between them constant. In contrast, moving down Sequence 2, the distance between consequences is progressively increased, creating a mean-preserving spread, while again keeping probabilities constant. In this sequence, the range of certainty equivalents that would not violate stochastic dominance varies by a factor of six, from £4 to £24. This allows us to test the extent to which this range is a crucial determinant of the size of imprecision intervals (as does variation in the range across sequences).

In contrast, within each of sequences 3–7, we hold the set of consequences fixed but vary the probabilities attached to them. The nature and purpose of these sequences can be seen by locating them in ‘unit probability triangles’.

Figure 3a locates sequences 3 and 4 in such a triangle. These sequences involve the three consequences £24, £20 and £0. The vertical axis measures the probability of the best consequence, increasing from zero to one moving bottom to top. The horizontal axis measures the probability of the worst consequence, increasing from zero to one moving left to right. The probability of the middle consequence is given implicitly as $(1 - p_1 - p_3)$.

Figure 3b locates sequences 5 and 6 (which involve the consequences £30, £22.50 and £0) in another triangle in the same way; and Fig. 3c locates sequence 7 (which involves the consequences £20, £10 and £0) in another. The lotteries presented to subjects from sequences 3 to 7 are marked by black circles in the relevant triangles; the degenerate lotteries 4.1; 5.1 and 7.1, which give the middle outcome of each triangle for sure, are marked by white circles. Repeated lotteries are indicated by an ‘R’ appended to their labels.

Observing imprecision across the lotteries in Fig. 3 allows us to address a number of questions. One relates to the conjecture (MacCrimmon and Smith 1986; Butler and

Table 1 The lotteries

Label		x_1	x_2	x_3	p_1	p_2	p_3	Sequence
1.1	R	40	22	–	0.6	0.4	–	1. Payoff Scale
1.2		35	17	–	0.6	0.4	–	
1.3		30	12	–	0.6	0.4	–	
1.4		25	7	–	0.6	0.4	–	
1.5	R	20	2	–	0.6	0.4	–	2. Mean preserving spread
2.1	R	18	14	–	0.4	0.6	–	
2.2		19.5	13	–	0.4	0.6	–	
2.3		21	12	–	0.4	0.6	–	
2.4		24	10	–	0.4	0.6	–	
2.5	R	30	6	–	0.4	0.6	–	3. Common Consequence (Risky)
3.1		24	20	0	0.05	0.93	0.02	
3.2		24	20	0	0.05	0.90	0.05	
3.3		24	20	0	0.05	0.60	0.35	
3.4	R	24	20	0	0.05	0.30	0.65	
3.5	R	24	20	0	0.05	0.00	0.95	4. Common Consequence (Safe)
4.1	D	–	20	0	–	1.00	0.00	
4.2		–	20	0	–	0.97	0.03	
4.3		–	20	0	–	0.67	0.33	
4.4	R	–	20	0	–	0.37	0.63	
4.5	R	–	20	0	–	0.07	0.93	5. Common Ratio (Safe)
5.1	D	–	22.5	0	–	1.00	0.00	
5.1*	R	–	22.5	0	–	0.99	0.01	
5.2		–	22.5	0	–	0.95	0.05	
5.3		–	22.5	0	–	0.75	0.25	
5.4	R	–	22.5	0	–	0.25	0.75	6. Common Ratio (Risky)
5.5		–	22.5	0	–	0.05	0.95	
6.1	R	30	–	0	0.80	–	0.20	
6.2		30	–	0	0.76	–	0.24	
6.3		30	–	0	0.60	–	0.40	
6.4	R	30	–	0	0.20	–	0.80	7. Betweenness
6.5		30	–	0	0.04	–	0.96	
7.1	D	20	10	0	0.00	1.00	0.00	
7.2		20	10	0	0.08	0.90	0.02	
7.3	R	20	10	0	0.20	0.75	0.05	
7.4		20	10	0	0.60	0.25	0.15	7.5
7.5	R	20	10	0	0.80	0	0.20	

Loomes 2007) that imprecision in valuation of lotteries increases with ‘distance from certainty’. The distance of a lottery from its nearest corner in the triangle is a natural measure of its proximity to certainty. Inspection of Fig. 3a–c shows that our set of lotteries contains considerable variation on this dimension.

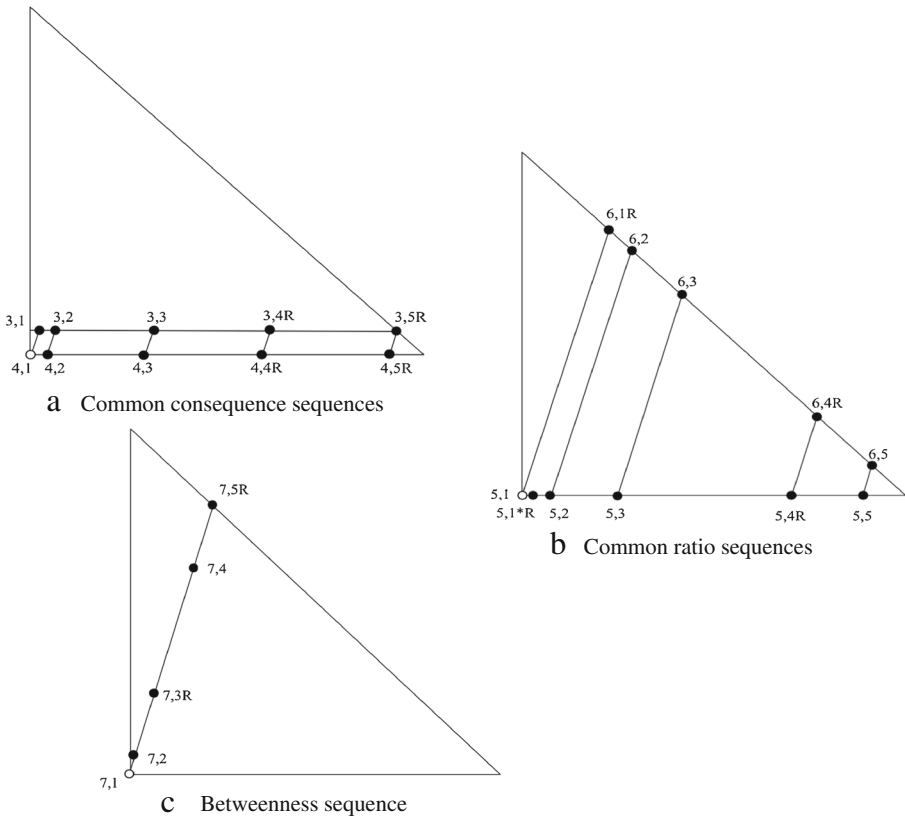


Fig. 3 a–c: Sequences 3 to 7 in triangle diagrams

Within each triangle, the lotteries are also selected to permit tests for the analogues of well-known violations of EUT, using preferences over lotteries inferred from comparisons of CEs. In each case, we can explore whether those violations are associated with imprecision.

To understand this in more detail, note that in Fig. 3a (and similarly 3b) we have constructed lines connecting pairs of lotteries with each such pair containing one relatively safe lottery (the lower left of the pair) and one relatively risky lottery (with higher expected value and range). Sequences 3 and 4 together constitute a set of so-called ‘common consequence’ lotteries because, starting with the left-most pair in Fig. 3a (i.e., 3.1 and 4.1), each pair further to the right is constructed by increasing the probability of the worst consequence at the expense of the middle consequence, with probability shifted by the same amount for the safe and risky options. EUT implies that these adjustments will not affect an individual’s ranking of the safe alternative versus the risky alternative. There is, however, well documented evidence of a tendency for individuals to switch from preferring the safe options in pairs like (3.1; 4.1) near bottom left corners of triangles to preferring risky options in pairs like (3.5; 4.5) near right hand corners. This is the so-called ‘common consequence effect’ first discovered by Allais (1953). Sequences 3 and 4 are based on, and extend, sequences used in studies by Wu and Gonzalez (1996, 1998) that identified a common consequence effect.

Sequences 5 and 6 in Fig. 3b are constructed with a similar strategy in mind. Ignoring lottery 5.1* for the moment and starting with the leftmost pair (5.1; 6.1), each subsequent pair is constructed by scaling down the probability of all non-zero outcomes in either lottery by a common factor. This generates a series of ‘common ratio’ pairs (so called because the ratio of winning probabilities for the safe and risky lotteries is constant across pairs). According to EUT, this common scaling of probabilities should not affect any individual’s preference of the safe option, as against the risky option, but there is a well-known empirical tendency for individuals to switch from preferring safe alternatives to risky alternatives as the probability of winning is scaled down (equivalently, as we move left to right across pairs in the triangle). This is the so called ‘common ratio effect’. Our sequences borrow and extend the probabilistic structure of lotteries used in a classic demonstration of that effect due to Kahneman and Tversky (1979).

One lottery in sequence 5 does not follow this logic. This is lottery 5.1*, a slightly displaced version of the degenerate 5.1. It is the closest to certainty of all the lotteries for which we observe decisions; we include it to see what happens to imprecision in the *very* close proximity of certainty. We append a star to its label in Fig. 3b to highlight that its construction does not follow the pattern of other lotteries in the common ratio sequences.

Finally, the lotteries in sequence 7 are constructed to lie on a single straight line in the triangle of Fig. 3c. EUT implies that indifference curves in the triangle are upward sloping, straight (and parallel) lines, linearity being implied by a property called ‘betweenness’. Given standard regularity conditions, any betweenness-consistent subject must either be indifferent between the lotteries of sequence 7 or, if not, rank them either in the order they appear in the sequence or in the reverse order. This sequence extends a series of lotteries shown to exhibit violation of betweenness in a study by Chew and Waller (1986).

The bulk of the previous evidence for common consequence effects, common ratio effects and betweenness violations is derived from studies in which preferences are inferred from choices. In contrast, we will infer the presence of the corresponding violations of EUT from binary comparisons of CEs. As it is an implication of EUT that binary preferences are revealed by relative CEs,¹⁴ searching for patterns in binary comparisons of CEs that match patterns of choice that would violate EUT is a valid test of EUT. As we discuss in Section 4, we do not presuppose that the relative CEs we elicit would coincide with choices. In view of this, if the patterns we observe in relative CEs match classic patterns of EUT violation previously observed in choices, we will say that our observed patterns are *analogous* to the classic effects.

We make repeated observations for two lotteries in each sequence. For sequences 1 and 2, we select the first and last lotteries of each sequence (i.e., the most extreme in terms of payoff size or range, respectively). This allows us to examine whether variability of imprecision is associated with either of these dimensions. For sequences 3–7, we repeat particular lotteries which have (or in the case of 5.1* are very close to ones which have) demonstrated violation of EUT in past research. These are cases where we have relatively strong priors for expecting EUT violation. Where we observe violation, data on repeated tasks will allow us to examine whether EUT violation is associated with variability of CEs as well as with, or perhaps instead of, imprecision.

¹⁴ This is an immediate implication of the ordering assumption of EUT. Assuming that more money is preferred to less, one lottery is preferred to another if and only if it has a higher certainty equivalent.

This selection of repeat tasks also achieves a mix of lotteries from sequences 3 to 7 in terms of proximity to certainty.

2.3 The sessions

Including the 14 repeats, we wished to collect responses for a total of 47 lotteries. We thought this too many for an individual subject, so we divided lotteries between two different types of session. In one type, subjects responded to sequences 1, 3 and 4 and, in the other type, to sequences 2, 5, 6 and 7. Note that this means that the same subjects faced both common consequence sequences, and likewise for both common ratio sequences; so we have within-subject data on the violations of EUT analogous to common consequence and common ratio effects.

In each session, we followed the same procedure. At the start, subjects were provided with general instructions. Then, we explained the concept of a lottery, the response format, and the incentive system. The explanation of the response format included an example in which the experimenter illustrated how to complete a response table displayed on an overhead transparency. Exact written instructions are available from the authors upon request.

Before proceeding with the main tasks, subjects completed a practice response table and we checked to make sure that participants appeared to have understood what we had asked them to do.¹⁵ After the practice, booklets containing half of the lotteries in the session were given to subjects to complete. When subjects had done that, they completed some filler tasks (not reported here) for about 15 min. These were partly intended to reduce the chance that subjects would recall the details of earlier tasks when they faced repeated tasks in the second block. As the filler tasks were rather different, we also expected them to help maintain subjects' interest. After the filler tasks, subjects were given another booklet with the rest of the tasks for their session. The repeated lotteries were always at the end of the second booklet, partly to allow a maximum of experience before facing any repeats and partly to prevent subjects realizing that some lotteries were repeated before having responded to the whole set once. The order of tasks in the first booklet was randomised for each subject, as were the orders of the tasks preceding the repeats in the second booklet and the block of repeats.

Once all subjects had completed all tasks, public random draws were completed to determine which task would be for real for each subject. Subjects then played out their 'real' task: for the selected response table, a random draw determined a row and the subject then received their chosen alternative for that row. If the chosen alternative was a lottery, a final draw determined their payoff. Subjects were paid, on the spot, in cash and the average payout was £20.09. Sessions lasted about 90 min.

3 Results

The experiment was run at the University of Nottingham. Following a pilot, we ran six sessions (three of each type) involving 82 participants: 42 in type 1 sessions and 40 in

¹⁵ We used a protocol under which specific response patterns that clearly suggested misunderstanding triggered further explanation.

type 2 sessions. Following input of the data, we removed the responses of three participants (two from sessions of type 1 and one from a type 2 session), whose responses could not be interpreted. The analysis that we present is based on the responses of the remaining 79 participants.

3.1 Coherence and stability of imprecision

Table 2 presents means and standard deviations for a series of variables that will play a key role in our analysis. Here, data are presented at lottery level (aggregating across individuals), but exclude observations based on repeat tasks. (We analyze individual responses and data for repeats later.) For each lottery and subject, CE denotes the certainty equivalent; INTSIZE denotes the *size* of the imprecision interval expressed as a monetary sum, while INTMID denotes the *location* of the imprecision interval, measured for each individual as the mid-point of their interval. Before considering the data, it is worth noting two points. First, for a given subject and a given lottery, INTSIZE and INTMID jointly fully characterize the imprecision interval. Second, INTMID and CE are quite different concepts in terms of our elicitation: INTMID is constructed *by us* to be the midpoint of the subject's imprecision interval; CE is selected *by the subject* to be the point where they switch between choosing the lottery and choosing the sure amount. In view of this, Table 2 also reports ASYM, which measures deviation between CE and INTMID and is defined more precisely below.

3.1.1 Properties of certainty equivalents

Before analysing measured imprecision, we start by examining the behaviour of the more familiar concept of certainty equivalent. We ask two questions: whether CEs covary sensibly with the objective attractiveness of lotteries; and whether they have plausible magnitudes. Broadly speaking, the answers to both questions are positive.

There are 5 sequences where the lotteries become unambiguously worse moving down the table within a sequence: sequence 1 and sequences 3–6. Additionally, for sequence 2, lotteries get worse moving down the table, conditional on risk aversion. CEs behave coherently by falling consistently as we move down each of these sequences. For most lotteries, the average CE also appears unsurprising, given previous studies. One simple metric for this is the ratio of mean certainty equivalent to expected value (CE/EV) which is also reported in Table 2. With the exception of four lotteries, CE/EV lies in the range 0.65 to 1.34 and in most cases the ratio is less than one, consistent with a modest degree of risk aversion, typical of many experiments.¹⁶

¹⁶ There are four lotteries for which CE/EV > 3, suggesting extreme risk seeking. These lotteries all lie very near bottom right corners of the relevant triangles in Fig. 3, with extreme probabilities of zero outcomes. (They are lotteries 3.5, 4.5, 5.5 and 6.5 with probabilities of zero of, respectively, 0.95, 0.93, 0.95 and 0.96.) The observation of unusually risk-seeking behaviour near bottom right corners of triangles is consistent with previous evidence of common ratio and common consequence effects. Overvaluation of lotteries with a low chance of winning is also well-known as a major contributor to the preference reversal phenomenon (Tversky et al. 1990).

Table 2 Summary results

Lottery	INTSIZE		INTMID		CE		CE/EV	ASYM		Paired <i>t</i> -test CE vs INTMID
	Mean	Sd	Mean	Sd	Mean	Sd		Mean	Sd	
1.1	3.63	2.22	30.02	2.14	29.26	2.14	0.89	0.63	1.62	0.03
1.2	3.73	2.13	24.70	2.45	24.76	2.24	0.89	0.05	1.38	0.81
1.3	4.74	2.15	19.76	2.31	19.56	2.56	0.86	0.53	2.48	0.19
1.4	4.14	1.99	14.26	2.22	14.26	2.41	0.80	0.20	1.73	0.49
1.5	4.74	2.06	9.07	1.83	9.08	2.26	0.71	0.17	2.17	0.63
2.1	1.00	0.98	16.25	2.6	15.91	0.70	1.02	0.62	0.47	0.00
2.2	1.66	1.00	16.37	1.56	15.96	1.03	1.02	0.63	0.73	0.00
2.3	2.23	1.05	16.22	1.35	15.87	1.36	1.02	0.59	0.70	0.00
2.4	3.82	1.77	16.33	0.95	15.81	1.59	1.01	1.00	1.41	0.00
2.5	6.55	3.05	14.94	0.53	14.24	3.42	0.91	1.40	3.10	0.01
3.1	3.45	2.66	17.38	3.88	17.79	3.63	0.87	-0.32	2.46	0.64
3.2	3.45	2.44	16.82	3.45	17.12	3.37	0.89	-0.13	2.49	0.78
3.3	5.64	3.12	10.85	2.77	11.15	2.92	0.84	-0.09	3.79	0.88
3.4	5.01	3.00	7.93	2.35	7.61	3.02	1.06	0.28	2.99	0.57
3.5	3.56	2.99	5.22	3.52	4.79	3.95	3.99	0.45	3.21	0.40
4.2	2.74	2.26	15.26	3.76	15.78	3.67	0.81	-0.11	1.53	0.69
4.3	4.85	2.24	9.31	2.10	9.40	2.54	0.70	-0.14	2.39	0.71
4.4	4.40	2.11	7.18	2.63	6.94	3.08	0.94	0.44	2.20	0.22
4.5	3.31	2.30	4.96	2.58	4.49	2.95	3.21	0.33	2.24	0.39
5.1*	2.80	3.73	17.77	3.75	19.42	2.70	0.87	-1.54	3.20	0.03
5.2	3.99	3.74	16.84	3.10	17.44	3.10	0.82	-0.56	3.13	0.31
5.3	5.40	3.11	12.37	2.42	12.27	2.76	0.73	0.19	2.66	0.65
5.4	4.78	3.02	7.76	2.58	7.03	3.33	1.25	0.77	2.56	0.08
5.5	3.94	2.86	5.30	3.42	4.44	3.33	3.95	1.59	2.44	0.00
6.1	6.72	3.07	17.12	4.29	16.83	4.98	0.70	0.25	3.46	0.66
6.2	6.77	3.53	15.42	3.66	14.88	4.00	0.65	0.95	2.62	0.03
6.3	7.15	3.59	12.46	3.41	12.26	4.30	0.68	0.41	3.40	0.46
6.4	5.71	4.14	9.05	4.03	8.06	4.08	1.34	1.53	2.60	0.00
6.5	4.23	4.62	5.97	5.06	4.78	4.56	3.98	1.91	2.82	0.00
7.2	4.08	2.33	11.66	1.78	11.53	1.75	1.09	0.22	1.62	0.42
7.3	3.67	2.66	11.61	1.90	11.56	2.42	1.01	0.46	2.36	0.24
7.4	4.96	2.34	11.66	2.38	11.56	2.39	0.80	0.14	2.28	0.70
7.5	4.51	2.56	12.11	2.17	12.18	2.49	0.76	0.44	2.34	0.26

3.1.2 Properties of imprecision intervals

We now turn to the behaviour of imprecision intervals. One striking feature of our data is that imprecision is pervasive. Across all lotteries, the average percentage of subjects reporting positive imprecision intervals is 87%; the percentage ranges from 61 to 100%

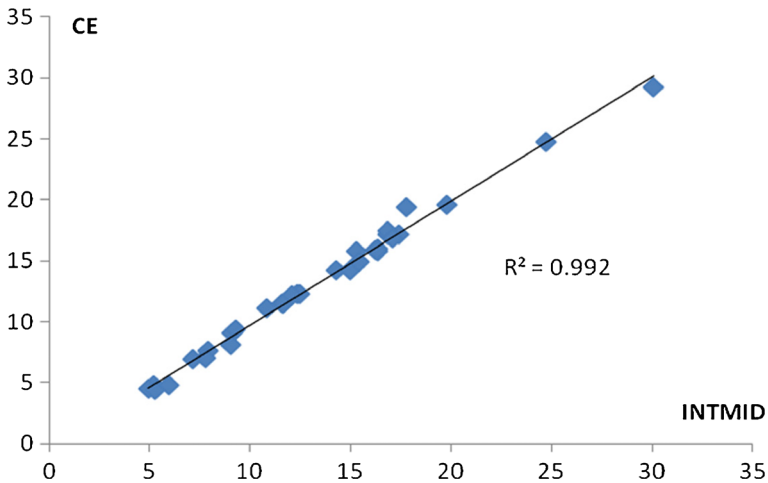


Fig. 4 The association between INTMID and CE

depending on the lottery. As well as cross-lottery variation which we analyse below, there is considerable cross-individual variation: some individuals express more imprecision than others (as measured by their mean INTSIZE across tasks).¹⁷ We find no association between this and individual risk attitudes (proxied by an individual's mean CE across tasks) and we find no evidence that the propensity for imprecision is related to gender.

One of our main questions is whether INTMID and INTSIZE, the two measures that characterise an imprecision interval, vary in a coherent and intelligible way with the structure of lotteries. We start by examining INTMID. At lottery level, CE and INTMID are very highly correlated. This is illustrated in Fig. 4 which plots the average CE for each lottery against the average INTMID and shows a positive correlation of 99%.

While it is unsurprising that there is a positive association in Fig. 4, the strength of it brings another question to the fore: Does INTMID have *distinct* properties from CE? To investigate this, we explore two sub-questions: (i) Do CE and INTMID have different distributions? (ii) If so, do the differences depend on lottery structures?

The statistic reported in the last column of Table 2 addresses the first of these issues. For each lottery, this reports a paired *t*-test comparing the means of CE and INTMID, using a two-tailed procedure. Of the 33 tests reported, 11 show significant differences ($p < 0.05$) between means of CE and INTMID for the relevant lottery. For 10 of these cases, CE is below INTMID. The significant differences in means do not seem to be randomly scattered across sequences: for some sequences (3, 4 and 7), the difference is never significant; for one (sequence 2) it is always significant; and for others it is significant in places (1, 5 and 6).

¹⁷ For Type 1 (Type 2) sessions, individual imprecision (mean across tasks) varied from 0 (1.76) to 8 (9.37) with a mean of 4.05 (4.5) and a standard deviation of 1.51 (1.76).

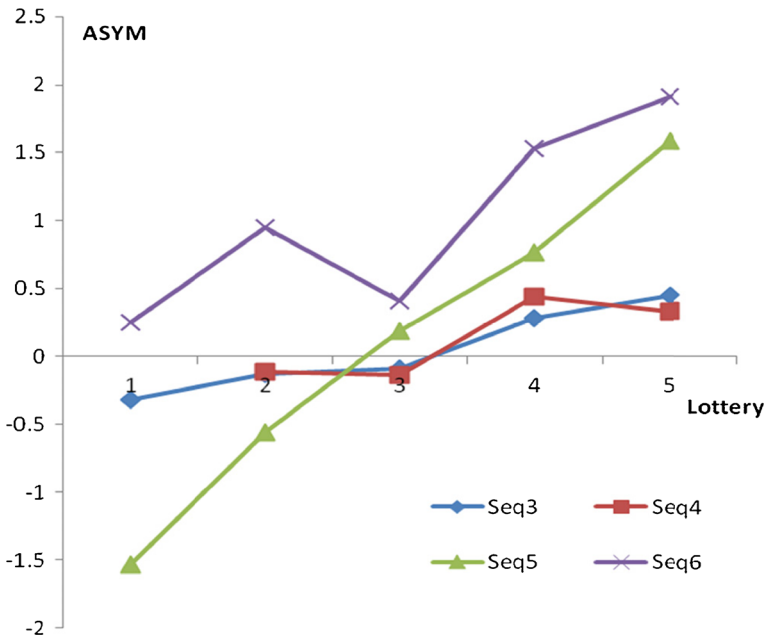


Fig. 5 ASYM, sequences 3–6

The variable ASYM (reported in Table 2) reveals more about patterns in those differences. It measures deviation between INTMID and CE and is calculated as $ASYM = (\text{upper boundary of imprecision interval} - CE) - (CE - \text{lower boundary of imprecision interval})$.¹⁸ Hence, $ASYM > 0$ indicates a tendency for CE to be skewed towards the lower boundary of the interval; $ASYM < 0$ indicates skew to upper boundary. A particularly striking pattern is evident in the behaviour of ASYM for the common consequence and common ratio sequences (sequences 3 through 6). To illustrate this, Fig. 5 plots mean ASYM for these four sequences.

At the eyeball level, ASYM rises along each of the sequences. Recall that the final lottery of each sequence is in the extreme bottom right hand corner of the relevant triangle. For these lotteries with a very low chance of winning, mean ASYM is always positive, indicating skew of CE towards the zero-payoff end of the imprecision interval; and, in the cases of sequences 3, 4 and 5, it flips from negative to positive along the sequence. Since these sequences share the characteristic of having quite extreme probabilities at both ends, this pattern is consistent with a tendency for CE to be skewed away from the midpoint of the interval towards an outcome that is highly

¹⁸ For a given lottery and a given subject who reveals a CE and a non-zero imprecision interval for the lottery, the definition of ASYM implies $ASYM = 2(\text{INTMID} - CE)$. But, it is not true of the values in Table 2 that, for each lottery, $\text{mean ASYM} = 2(\text{mean INTMID} - \text{mean CE})$. The reason is that for each variable the average (for a given lottery) is taken over those subjects for whom the variable is defined; and the set of such subjects is not always the same for all three variables. INTMID and ASYM are defined, for a given lottery, only for subjects whose INTSIZE is non-zero. Moreover, for a few lotteries, we observed some subjects who indicated a non-zero imprecision interval but did not indicate a switch-point as requested. CE is defined only for the remaining (and large majority of) cases.

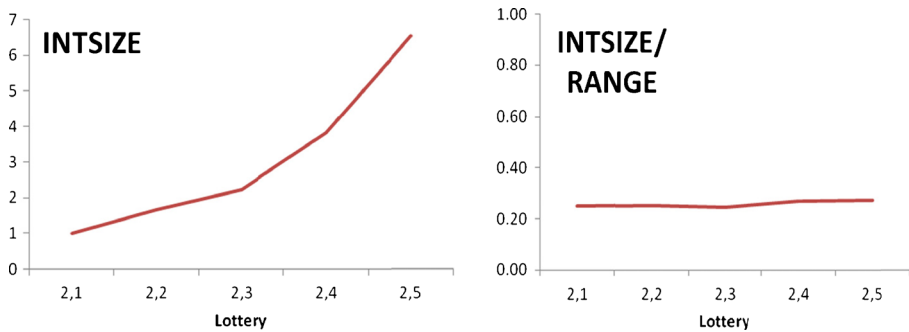


Fig. 6 INTSIZE, sequence 2 (mean preserving spread)

probable, when there is one. Sequences 5 and 6 show significant changes from beginning to end of sequence (Z-value (p-value) on Wilcoxon is 3.41 (0.00) for sequence 5 and 2.02 (0.04) for sequence 6). For sequences 3 and 4, changes are not significant at the conventional levels.¹⁹ So, filtering according to whether CE is significantly different from INTMID, we have clear evidence that the tendency for CE to be skewed to the lower outcome (ASYM>0) gets stronger as you move towards the bottom right hand corner of a triangle and we have more limited evidence (sequence 5) of the mirror image effect as you approach the bottom left hand corner of triangle: in the lottery 5.1* the CE is significantly skewed towards the better payoff.

We now examine cross-lottery variation in the *size* of imprecision intervals. We first examine sequences 1 and 2, each of which varies consequences holding probabilistic structure constant. Along sequence 1, there is a substantial decline in attractiveness, measured by CE, but little change in mean INTSIZE, suggesting that the extent of imprecision is not much affected by the overall size of lottery outcomes. By contrast, sequence 2 generates large changes in INTSIZE. Mean INTSIZE for this sequence is plotted in the left hand panel of Fig. 6.

The right hand panel of Fig. 6 plots mean INTSIZE for the same sequence, but this time scaled by a variable that we will call RANGE. RANGE records the distance between the best and the worst outcomes for each lottery; it can also be interpreted as measuring the range of CEs which would not violate stochastic dominance. It is apparent that, for this sequence, INTSIZE is essentially a constant proportion of RANGE (it is always in the tight interval of 25–27% of RANGE). This result is reminiscent of the conclusion reached in Butler and Loomes (2011) that, for their lotteries, imprecision intervals are of “much the same magnitude” when expressed as a proportion of the response range which would not violate stochastic dominance.²⁰ But, as noted above, we have much more variation in that range.

¹⁹ We also ran parallel tests using a measure of asymmetry scaled by size of imprecision interval. This produced the same qualitative conclusions: sequences 5 and 6 having the only statistically significant end-to-end differences.

²⁰ When scaled by their corresponding range measure, all imprecision intervals reported in the two studies by Butler and Loomes (2007, 2011) lie in the range 21–28%.

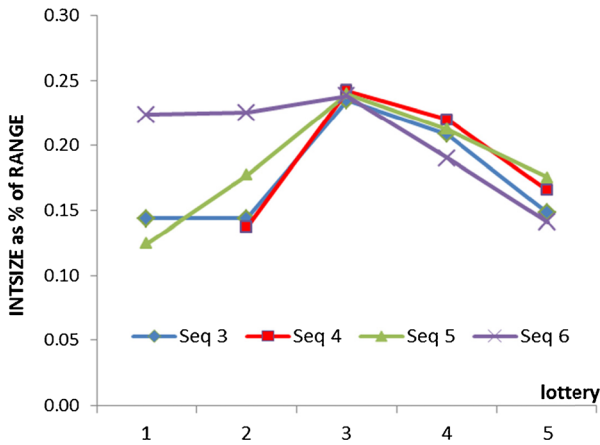


Fig. 7 INTSIZE scaled by RANGE, sequences 3–6

Our data also point to another significant determinant of interval size. There is a strong tendency for imprecision intervals to increase with the ‘distance’ of a lottery from certainty. As a natural measure of this distance, we construct a variable DISTCERTAINTY. For the lotteries displayed in Fig. 3, this variable is calculated as the length of the straight line connecting the lottery to the nearest corner of the relevant triangle; whereas, for the (two-outcome) lotteries in sequences 1 and 2, it is calculated as the probability of the least probable possible outcome. A simple regression of INTSIZE/RANGE against DISTCERTAINTY results in an R^2 of 0.747.

A specific aspect of this association is illustrated in Fig. 7, which plots INTSIZE/RANGE for the common ratio and common consequence sequences. (Note that RANGE is constant along each of these sequences, but varies across them.) There is an apparent (and always significant) decrease in the size of imprecision intervals every time that lotteries in these four sequences approach certainty (i.e., at the start and end of sequences 3, 4, and 5 and at the end of sequence 6). There is also a striking regularity in the behaviour of INTSIZE/RANGE: it hovers around 15% in the vicinity of either bottom edge corner of the relevant triangle but increases to around 25% for lotteries most distant from the corners. So, to summarise: our data allow us to reject the hypothesis that interval size is simply proportional to the range of responses that would not violate stochastic dominance because INTSIZE varies systematically with proximity to certainty, holding RANGE constant.

3.1.3 Stability across repeat measurement

As the final step in the lottery-level analysis, we examine how imprecision varies across repeat measurement.

Figure 8 reveals that changes in CE and INTMID follow a similar pattern to each other. For each lottery, the figure plots the percentage changes in mean CE against percentage changes in mean INTMID. Eyeballing this graph suggests some tendency for downward drift in both variables. There is, however, only one case (the most extreme point) where CE changes significantly at the 5% level and no case where INTMID does so.

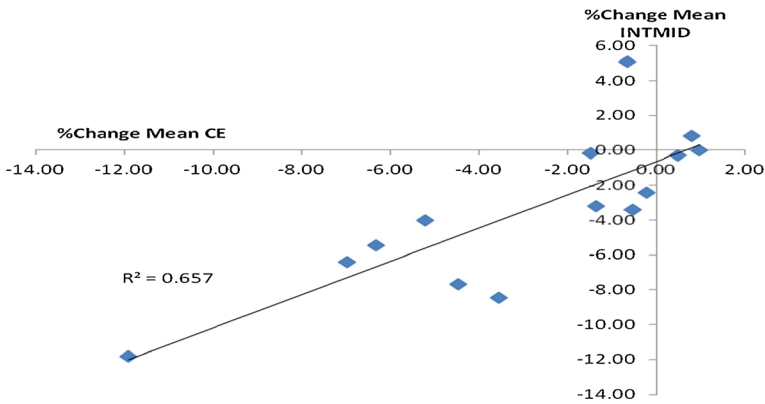


Fig. 8 Percentage changes in CE and INTMID

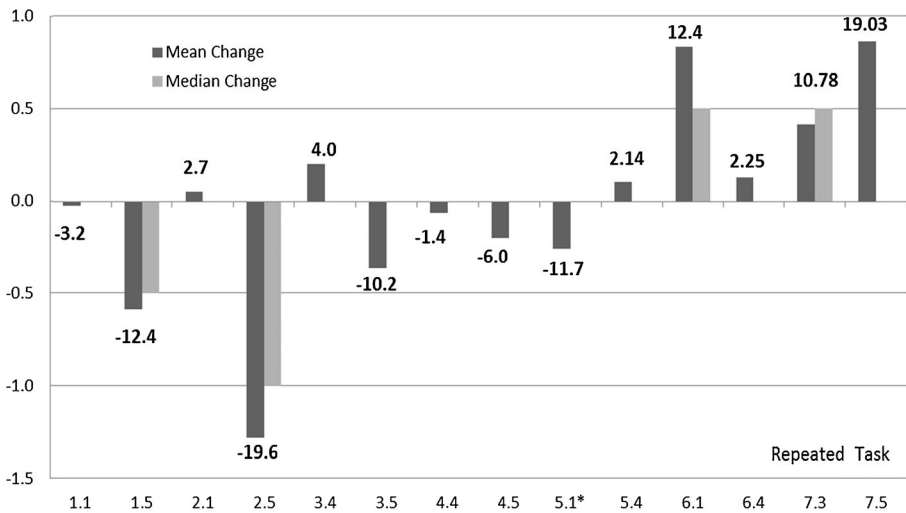


Fig. 9 Changes in INTSIZE (Mean, median and percentage change in mean)

Figure 9 shows the mean and median changes in INTSIZE for each of the 14 repeated lotteries. Percentage changes in the mean are recorded above or below the relevant bars. Median changes are zero in all but four cases. There is no evidence that INTSIZE varies systematically with repeat measurement. For 7 lotteries, the mean change is positive and for the other 7 it is negative but in only one case is there a significant change (lottery 2.5).²¹ Across all lotteries, the average change is about half of 1%. The stability in these data is notable given the degree of separation between repeat observations: recall that repeat tasks were separated both by a variable number of intervening elicitation tasks and by a set of filler tasks. Thus, it would seem implausible to attribute stability to effects of memory and/or anchoring.

²¹ Wilcoxon signed rank tests and t-tests produce the same conclusions.

3.1.4 Individual level analysis of imprecision

To complement the analysis presented so far, we developed a set of regression models to explore the determinants of imprecision intervals. This analysis allows us to exploit to greater effect the richness of the individual level data for the full set of tasks across all sequences, and also to add additional controls.

Table 3 presents estimated equations for each of INTMID, INTSIZE and ASYM. For each regression, we use a series of variables to represent lottery and task characteristics. Every lottery in our study can be thought of as a point in some probability triangle. We have already interpreted sequences 3–7 in this way (in Fig. 3) and a similar interpretation can be provided for the two-outcome lotteries of sequences 1 and 2 by thinking of each of them as having a “third outcome” x_3 of zero, with probability zero. On this interpretation, each of the lotteries in sequences 1 and 2 is located on the vertical edge of some triangle; and all lotteries in all sequences have $x_3=0$. We can then

Table 3 Regressions for INTMID, INTSIZE and ASYM

Dependent Variable	INTMID		INTSIZE		ASYM	
	Full	Full	Stepwise	Full	Stepwise	
x_1	0.528** (0.076)	0.058 (0.079)		-0.232** (0.067)	-0.228** (0.035)	
RATIO x_2x_1	5.018** (1.398)	0.870 (1.65)		-3.378* (1.352)	-3.360** (0.819)	
p_1	2.195* (1.081)	2.816* (1.132)	2.211** (0.280)	-3.194** (0.925)	-3.197** (0.520)	
p_3	-7.640** (1.959)	-2.262 (2.386)	-0.841* (0.377)	4.908* (1.927)	4.984** (1.042)	
RANGE	-0.256** (0.021)	0.203** (0.019)	0.206** (0.013)	0.011 (0.015)		
EV	0.245* (0.105)	-0.122 (0.286)	-0.049** (0.016)	0.300** (0.091)	0.303** (0.050)	
REPEAT	0.395 (0.266)	0.081 (0.186)		-0.033 (0.196)		
ORDER	-0.037 (0.022)	-0.014 (0.018)		-0.006 (0.014)		
Const	-0.166 (1.281)	-0.285 (1.329)	0.294 (0.315)	2.783* (1.178)	2.775** (0.583)	
Observations ^a	1674	1830	1830	1660	1660	
R ²	0.779	0.162	0.161	0.044	0.043	
F	397.88	48.36	87.27	4.05	14.78	
(dof)	(8, 77)	(8, 78)	(4, 1825)	(8, 77)	(5, 1654)	
Prob>F	0.000	0.000	0.000	0.001	0.000	
Clusters (subject)	78	79		78		

^a Differences in the number of observations are mainly due to imprecision intervals of size zero. In those cases, INTMID and ASYM are undefined. In a small number of cases, subjects provided an imprecision interval but not a CE, making ASYM undefined. In this and subsequent tables, standard errors are in parentheses and * and ** denote significance at the 5 and 1% levels, respectively

characterise a lottery by two features: the first is ‘which triangle’ the lottery is in (i.e., what is the support of the lottery); the second is ‘where in the triangle’ the relevant lottery is located. We identify the triangle from which a lottery comes using two variables: ‘ x_1 ’ which (as defined above) is the best payoff in the triangle and ‘ $\text{RATIO } x_2/x_1$ ’ which is the ratio of the middle to the best consequence in the triangle. We identify the position of a lottery in its triangle using the probability of its best (p_1) and worst (p_3) consequences. We include the variable RANGE which, as above, is the distance between the best and worst²² consequences of the lottery and corresponds with Butler and Loomes’ (2007, 2011) concept of a range of permissible responses. We also include EV, the expected value of each lottery. Finally, we include two variables to capture the experience of the decision maker. The variable ‘ORDER’ is a measure of how many tasks the subject had previously completed while ‘REPEAT’ is a dummy with value 1 for tasks where the subject had completed an identical task before.

The models labelled ‘Full’ include all of the regressors just explained and parameters are estimated via linear regression (in STATA). Since there might be correlations between some subsets of our regressors (and hence possible issues of multicollinearity), as a robustness check, we also estimated via STATA’s stepwise regression procedure. Stepwise estimation did not eliminate any variables in the equation for INTMID but does generate alternative equations (labelled ‘stepwise’ in Table 3) for INTSIZE and ASYM.²³

The main functions of this analysis are to probe our earlier claim that imprecision varies coherently with lottery characteristics; and, especially, to give a clearer indication of whether experience affects imprecision, controlling for those characteristics. We comment on these points in turn, in the light of Table 3.

First, each of the objective features of the lotteries has the expected effect on INTMID; and, for both INTSIZE and ASYM, key lottery characteristics survive the stepwise procedure and have intelligible directions of effect. This supports the conclusion that all three of these features of imprecision intervals are systematically related to objective features of the lotteries. Moreover, the results broadly cohere with the lottery-level analysis. For example, RANGE is the key lottery outcome variable for determination of INTSIZE. In addition, keeping in mind that the combination of a high p_3 and a low p_1 is associated particularly with lotteries in the bottom right corner of their triangle, Table 3 confirms that such lotteries will tend to have low INTMID and low INTSIZE (reflecting the fact that, other things equal, they are unattractive and close to certainty) and high ASYM (corroborating the eyeball impression from Fig. 5).

Second, we find no evidence that any of INTMID, INTSIZE or ASYM is affected by experience. Hence this analysis supports the conclusion from the lottery level analysis that measured imprecision is apparently rather stable. This feature of our findings is striking as it suggests imprecision is not just a temporary phenomenon confined to subjects’ early encounters with given lotteries and tasks.²⁴

²² For purposes of this statement, we interpret the ‘worst outcome’ as the worst of the outcomes with strictly positive probability (i.e., we ignore the zero outcomes with zero probability introduced above to allow characterisation of each lottery as lying in a triangle).

²³ Stepwise deletion (p -values ≤ 0.2) and stepwise addition (p -values ≥ 0.2) produce identical results.

²⁴ Some readers have suggested that the process of selecting CE might help subjects make their preferences more precise, even if they had initially revealed imprecision. If so, one might have expected a systematic tendency for INTSIZE to fall when tasks are repeated.

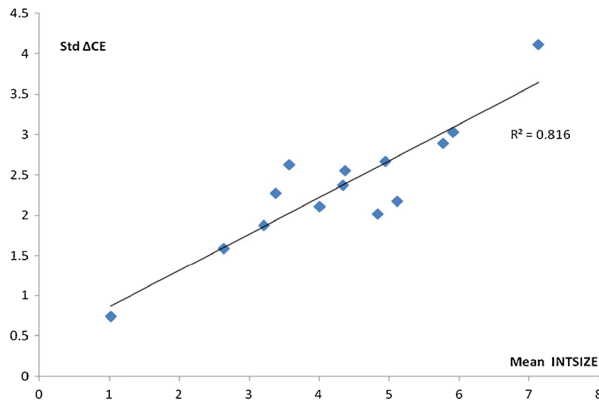


Fig. 10 Association of INTSIZE and variability of certainty equivalents

3.2 Imprecision measures: explanatory value

In this section, we examine the usefulness of the imprecision concept focussing on two questions: Does imprecision help explain variation in measured certainty equivalents? Does imprecision contribute to the explanation of violations of expected utility theory?

3.2.1 Imprecision and variation in certainty equivalents

To explore the relationship between imprecision and variation in certainty equivalents, we use data from the fourteen repeat tasks and begin with lottery level analysis. For each task repeated by an individual, we observe the change in their reported certainty equivalent, which we denote ΔCE . If, for a given lottery, variation in certainty equivalents across repeat tasks is generated by a stochastic process that is common to all individuals, we may interpret the standard deviation across individuals of ΔCE for the lottery as a simple measure of the variability of that process.²⁵ This measure of lottery level variability of certainty equivalents shows a very strong positive association with mean INTSIZE, as illustrated in Fig. 10. The best fitting regression produces an R^2 of 0.82. For the repeated lotteries, we also find a significant and strong positive association between the percentage of subjects who express positive imprecision for a given lottery and the percentage who give different CEs for the lottery in the two elicitations (Pearson’s correlation for the two variables is 0.66).

One possible interpretation is that these associations reflect a causal relationship between imprecision and variability in CEs. If true, that could be an important discovery. But, alternatively, the association could just reflect co-variance of imprecision with other structural features of lotteries that also determine variability of CEs. To explore this further, Table 4 reports regressions, testing whether imprecision predicts changes in CEs, controlling for objective characteristics of lotteries. The analysis uses individual level data and the dependent variable is the absolute value of ΔCE . The variable ‘INTSIZE(ave)’ is the size of the imprecision interval for the individual/lottery,

²⁵ An alternative interpretation is that the process generating variation in certainty equivalents across the repeat tasks for a given lottery is deterministic, but varies across individuals. Then, the measure captures the heterogeneity of subjects in the variation of certainty equivalents.

Table 4 Modelling changes in certainty equivalents

Dependent variable	abs(Δ CE)	
	Full	Stepwise ^a
Procedure		
INTSIZE(ave)	0.113* (0.048)	0.113** (0.030)
x_1	0.003 (0.409)	
RATIO_ x_2x_1	-0.625 (0.849)	
p_1	0.678 (0.657)	0.955** (0.255)
p_3	0.385 (1.114)	
RANGE	0.039* (0.017)	0.044** (0.014)
EV	0.015 (0.055)	
Const	0.200 (0.745)	0.013 (0.302)
Observations	538	538
R ²	0.104	0.102
F	12.43	20.17
(dof)	(7.78)	(3, 534)
Prob>F	0.000	0.000
Clusters (subject)	79	

^aforward and backward stepwise procedures produce identical results

averaged across the pair of repeated tasks. The remaining regressors are as in Table 3. Since the data involve multiple observations generated by each individual subject, we report robust standard errors calculated using regressions clustered on individuals. Alongside the full specification, we report a restricted specification generated by stepwise regression in STATA. The primary result from this analysis is that average imprecision interval size has a significant positive effect on variability in certainty equivalents, controlling for objective features of lotteries. The effect is significant at 5% level in the ‘full’ specification and at 1% in the stepwise specification, suggesting that imprecision does have predictive value in relation to variability in certainty equivalents.

3.2.2 Imprecision and violations of expected utility theory

We begin this sub-section by looking for evidence of the analogues of common consequence and common ratio violations of EUT, using sequences 3 and 4 in the first case and sequences 5 and 6 in the second. Since the same subjects faced sequences 3 and 4 (respectively 5 and 6), we can detect instances of the relevant type of EUT-violation at the level of the individual subject.

For example, consider the common consequence sequences of Fig. 3a. These sequences feature five pairs of common consequence lotteries, each of which contains

one relatively safe lottery on the horizontal axis and its relatively risky counterpart, connected by a straight line in the figure. We refer to the leftmost pair as common consequence ‘pair 1’, the next as ‘pair 2’, and so on across to ‘pair 5’ near the bottom right hand corner. For each individual who responded to the common consequence problems, we infer their ranking of the lotteries in each pair by comparing their stated certainty equivalents (in the case of repeat tasks, we use CEs revealed the first time). By doing this for each individual, we calculate the proportion of subjects who prefer the safer lottery in a common consequence pair, the proportion that give equal valuations, and the proportion who prefer the riskier lottery.²⁶ Figure 11 plots the first two of these statistics for the common consequence pairs (left hand panel); and the corresponding statistics for the common ratio pairs (right hand panel); in each case, the proportion preferring the riskier lottery is the residual.

The patterns of binary preference revealed by this procedure are similar to those of the studies which have elicited them through pairwise choice (Starmer 2000). If every subject behaved in line with EUT, these proportions would be stable across the five pairs in each panel but, in fact, there is marked change. Looking at the left hand panel first, in line with existing evidence there are highly significant falls in the proportion choosing safe, comparing the leftmost pair 1 with pairs further to the right in the triangle. For example, comparing pair 1 with any of pairs 2, 3 or 4, we can confidently reject the null hypothesis that the aggregate split between safe and risky preferences is unchanged ($p < 0.001$, Fisher’s exact test).²⁷ The change from 1 to 5 is only significant at the 10% level. This non-monotonic pattern of ‘switching’ is also seen in previous studies (Starmer 2000).

The right hand panel of Fig. 11 presents the corresponding data for the five pairs of lotteries that form the common ratio sequences (sequences 5 and 6, see Fig. 3b). In line with the evidence supporting the common ratio effect, but contrary to EUT, there is clear evidence that preference for the safe alternative decays from left to right across the triangle (79% prefer safe in the first pair, falling to 31% for pairs 4 and 5, and the change is significant at 1% based on Fisher’s exact test).²⁸

Our main interest in this sub-section is whether changes in preferences revealed by relative CEs across the common ratio and common consequence sequences are related to individual level imprecision. For example, are those who deviate from EUT more prone to imprecision than those who don’t? We explore this using the following statistical strategy. We work with adjacent pairs of (common consequence or common

²⁶ For the first pair in the sequence, we determine implicit preferences by comparing each individual’s valuation of lottery 3.1 against the certain amount associated with degenerate lottery 4.1 (i.e., £20).

²⁷ We implemented these tests in two ways. First, we excluded observations with equal valuations for safe and risk and then based the analysis on the subsample of subjects whose valuations revealed strict preferences in each pair. Our second procedure was to allocate the equal valuation cases between the safe and risky categories in each pair so that the allocation minimised the difference between them. The results we report in the text are based on results from the second, more conservative, procedure which always produced slightly higher (i.e., less significant) p -values.

²⁸ These aggregate figures actually understate the extent of individual level violation because even for some pairs of problems with relatively small aggregate changes in ranking, individual switches are still quite common (e.g., around 30 % of subjects switch moving between common consequence pairs 2 and 3 or between pairs 3 and 4 but some switches cancel out in the aggregate data). The analysis that follows below will take account of the individual level variation.

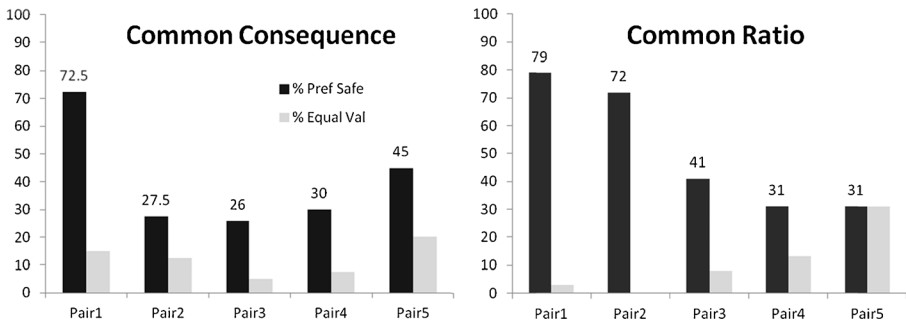


Fig. 11 Implicit preferences in tests of independence

ratio) lotteries and with a notation such that the two leftmost pairs of common consequence problems from Fig. 3a are called common consequence ‘pair1+2’ and so on. For each such pair, we identify individuals who violate EUT by switching their implicit preference either from safe to risk or from risk to safe across the adjacent pairs.²⁹ We then use logistic regression to examine whether the probability that an individual switches in a given case is associated with their measured imprecision (the dependent variable is 1 if an individual switches, zero otherwise). As an initial pass, for each pair we use three simple specifications with one independent variable which is either the individual’s mean (or ‘Global’) imprecision (calculated as their mean INTSIZE across all tasks they completed), their *maximum* INTSIZE among the four lotteries in the pair being analysed (‘Max_Local_INTSIZE’), or their *average* INTSIZE across the four lotteries being analysed (‘Mean_Local_INTSIZE’). This exercise generates 12 regressions, for each of the common consequence and the common ratio series.

Table 5 reports p-values associated with the independent variable in each of these regressions. For the common consequence pairs, none of these regressions produces an association that is significant at the 5% level between the measure of imprecision and the tendency to switch (in two cases, there are effects of local imprecision at the 10% level). For the common ratio pairs, there is more evidence of an association, but only for pair2+3 and only for local measures of imprecision. This might reflect a genuine local effect of imprecision but spurious correlation is another possibility so, to probe this, we report further analysis focussed on just this pair.

Table 6 reports eight regressions, each of which models switching across common ratio pair2+3. As independent variables, we use either Mean_ or Max_ Local INTSIZE across the four tasks in pair2+3 and, as an additional control, we introduce individual level ‘NOISE’, measured as variability of certainty equivalents on repeat tasks either globally or locally (we calculate the latter as the mean of variability for the three repeated tasks which form part of the common ratio sequences). These regressions show a consistent positive effect (at the 5% level) for local imprecision in five of six cases; the effect of mean local imprecision becomes insignificant with the inclusion of

²⁹ Individuals who give equal valuations for at least one pair are classified as EUT consistent.

Table 5 p-values for logistic analysis of common consequence/ratio sequences

INTSIZE Measure:	Common Consequence Pairs				Common Ratio Pairs			
	Pair1+2	Pair2+3	Pair3+4	Pair4+5	Pair1+2	Pair2+3	Pair3+4	Pair4+5
Global	0.193	0.968	0.416	0.139	0.604	0.074	0.454	0.169
Max_Local	0.082	0.331	0.896	0.329	0.302	0.018	0.984	0.120
Mean_Local	0.065	0.425	0.208	0.142	0.885	0.037	0.669	0.120

global noise which itself has a significant impact in two of the three specifications in which it occurs.

Overall, this suggests that imprecision and noise may both play some role in determining switching in specific common ratio pairs. The effects of imprecision are, however, apparently rather isolated to specific cases and so it seems clear that imprecision alone cannot explain the broader patterns of common ratio and common consequence effects found in our study.

We now conduct similar analysis for the betweenness tasks (sequence 7). EUT implies that, for any individual, CEs should vary monotonically, if at all, along sequence 7. Though the mean CEs for sequence 7 (reported in Table 2) are consistent with monotonicity in that they rise slightly moving from the triangle corner to the hypotenuse, the story is different at the individual level: almost 90% of subjects (35 out of 39) violate monotonicity somewhere across the sequence. To test for an association between monotonicity violation and imprecision, we conduct an analogue of the preceding analysis using tests based on adjacent triples of tasks from sequence 7.

Table 6 Effects of imprecision and noise for common ratio pair2+3

N=40,Logistic Dependent variable: Switching in common ratio pair2+3								
Mean_Local	0.274*	0.268*	0.228					
INTSIZE	(0.131)	(0.133)	(0.139)					
Max_Local				0.271*	0.301*	0.296*		
INTSIZE				(0.115)	(0.122)	(0.128)		
Global NOISE			1.000			1.285*		1.176*
			(0.547)			(0.594)		(0.541)
Local NOISE		0.408			0.543		0.419	
		(0.288)			(0.309)		(0.273)	
Const	-2.143*	-2.941**	-3.675**	-2.79**	-4.15**	-5.314**	-1.384*	-2.660*
	(0.868)	(1.079)	(1.283)	(1.05)	(1.394)	(1.732)	(0.670)	(1.067)
LogLikelihood	-24.02	-22.95	-22.09	-23.20	-21.49	-20.27	-25.20	-23.55
LRChi ² (dof)	4.89 (1)	7.02 (2)	8.74 (2)	6.51 (1)	9.95 (2)	12.38 (2)	2.53 (2)	5.83 (1)
Prob>Chi ²	0.027	0.030	0.013	0.011	0.007	0.002	0.112	0.016
PseudoR ²	0.092	0.133	0.165	0.123	0.188	0.234	0.048	0.110

Local imprecision measures are based on responses to the four tasks that form pairs 2 and 3. The local noise variable measures, for each individual, their average variation in CE across the three repeated tasks that form part of the common ratio sequences (i.e., tasks 5.4; 6.1; 6.4)

Table 7 Effects of imprecision and noise for intermediate betweenness triple

N=39, Logistic		Dependent variable: MTest2-4					
Global INTSIZE	-0.268 (0.190)						
Max_Local INTSIZE	-0.226 (0.154)						
Mean_Local INTSIZE		-0.406*	-0.450*	-0.413*			
		(0.207)	(0.212)	(0.208)			
Global NOISE			0.678		0.475		
			(0.519)		(0.475)		
Local NOISE				0.167		0.121	
				(0.268)		(0.243)	
Const	1.585	1.690	2.107*	1.100	1.891	-0.480	0.180
	(0.936)	(0.975)	(0.957)	(0.208)	(1.005)	(0.894)	(0.487)
LogLikelihood	-25.33	-25.28	-24.18	-23.17	-23.91	25.88	-26.27
LRChi ² (dof)	2.13 (1)	2.24 (1)	4.57 (1)	6.46 (2)	4.98 (2)	1.05 (1)	0.25 (1)
Prob>Chi ²	0.144	0.134	0.032	0.040	0.083	0.305	0.616
PseudoR ²	0.040	0.043	0.086	0.122	0.094	0.020	0.005

Max and Mean local imprecision are based on tasks 7.2; 7.3 and 7.4; local noise is change in CE across repeat of task 7.3.

For this purpose, we construct three new variables. ‘MTest1-3’ records, for each individual, whether they violate monotonicity across tasks 7.1; 7.2 and 7.3 (coded 1 for violation, zero otherwise)³⁰; 44% of individuals violate across this triple. MTest2-4 is constructed analogously using data from the task triple 7.2; 7.3 and 7.4 (59% violate on this triple); and, finally, MTest3-5 is constructed in the same way using tasks 7.3; 7.4 and 7.5 (56% violate on this triple). These three variables are then used as independent variables in a series of logistic regressions to test whether the probability of violating monotonicity is related to imprecision. We estimate specifications for each of these triples using measures of global and local (mean and max) imprecision as independent variables.

For MTest1-3 and MTest3-5 we find no evidence of association with any measure of imprecision (even at the 10% level of significance). However, we find some evidence for an effect of imprecision for the intermediate triple MTest2-4. Table 7 reports estimates for models of MTest2-4 including variants adding controls for (global or local) noise. There is a consistent effect (significant at 5%) for just one measure of imprecision (Mean_Local_INTSIZE) and it is robust to the inclusion of measures of noise. While the sign of the effect is surprising—local imprecision is associated with reductions in the likelihood of monotonicity violation—the overall picture emerging from the analysis of betweenness violation coheres with our investigation of independence violation: there may be some localised association between imprecision and violations of expected utility theory, but imprecision does not appear to be a primary causal

³⁰ We do this by computing the change in CE between the first and second task (assuming CE=10 for degenerate lottery 7.1) and similarly for the change in CE between the second and third task. We code an individual as violating monotonicity when the two changes have different signs.

variable capable of explaining the extensive and familiar patterns of violation of EUT observed in our data.

4 Discussion

The purpose of our experiment was to explore the three sets of questions posed in Section 1 about the empirical properties of preference imprecision, in relation to monetary lotteries, using a tool tailored to our purposes that measures imprecision on a money scale. Subject to the usual caveats about caution in generalising too far from the results of a single experiment, we believe we have delivered clear answers to these questions: measured imprecision intervals vary coherently with the structure of lotteries; there is no systematic tendency for their central location to move, nor for their size to either grow or shrink, with experience; and they correlate closely with variability in CEs, but not with the observed tendencies to violate EUT.

It may clarify our contribution to discuss in this section some related questions which, though interesting, we do not tackle elsewhere in this study. A first set of such questions concerns the underlying nature or causes of imprecision in preferences. Does imprecision reflect some underlying stochastic element in preferences? Or might it be understood better as a reflection of some quite different psychological state, yet to enter the vocabulary of standard economics, such as ‘vagueness of judgement’, ‘indecisiveness of reasons’ or ‘lack of confidence’?

We viewed it as premature to tackle these questions directly at this point in the development of the preference imprecision literature as, prior to our research, the evidence on empirical characteristics of imprecision was only fragmentary, as explained in Section 1. By systematically studying some basic empirical questions about the behaviour of measured imprecision, we have sought to establish clearer stylised facts about the behaviour of imprecision in a given domain. We hope that our research will help to inspire and shape broader conceptual debates about, and theoretical analysis of, the nature of preference imprecision.

Our finding of a strong empirical association between the size of imprecision intervals and the variability of measured certainty equivalents gives empirical support to past conjectures suggesting a close connection between noise and imprecision, but still leaves open how such an association might be understood. For example, consider two rather different possibilities.

One approach would be to model measured imprecision as a manifestation of underlying noisiness. It is well-established that people’s choices, as observed in experimental settings, are noisy in the sense that people quite frequently give different responses when presented with essentially the same decision task on more than one occasion (see Hey 2005 and Loomes 2005 for reviews). Given this, preference imprecision might be characterised in terms of people having some awareness of the variability of their own choices. Thus, for example, while considering a particular row of one of our response tables, an individual might appreciate the possibility that they would choose differently if offered the choice on this row on multiple occasions. In this case, individuals with relatively wide imprecision intervals in a given response table might be those who anticipate relatively high variability in repeated choices between the relevant lottery and amounts of money. Another possibility is that anticipated

variability might fall systematically moving away from the central point of the imprecision interval. While our data do not allow us to test these conjectures because we have no data on individuals' self-reports of anticipated choice variability, this modelling strategy combined with tests for association between imprecision intervals and anticipated choice variability may be an interesting route to pursue in further research.

A different conception of preference imprecision would be to suppose that it reflects a state of being unsure about what to choose, or of lacking confidence in one's grounds for a particular choice. Recall that our imprecision elicitation tool and others like it specifically refer to the concept of 'unsureness' about preference. Thus, one might expect responses to such tasks to have properties in common with forms of unsureness in other realms.

For example, suppose someone is asked in December 2014 if he is sure that Francois Hollande is President of France? He reflects on what evidence he has. He may have heard reports from different sources that agree on this point; he may recall events leading to Hollande's becoming president and reflect that he would have heard of any sudden fall from power. And so he may report that he is sure. In contrast, a different respondent to the same question might reflect that he had little recall of any French politicians or that, though aware of Hollande, he did not have clear grounds for thinking him President rather than Prime Minister, or French rather than Dutch. This respondent may conclude that he is unsure.

In the context of judgment about the facts of the world such as this, an account of 'unsureness' as originating from indecisiveness or paucity of reasons may be more plausible than one of it as originating from self-awareness-of-variability of response to the question. It does not necessarily follow that the same is true of unsureness of preference but, to the extent that preference imprecision has properties in common with unsureness of a fact, it may be worth exploring the modelling of it as indecisiveness of reasons. This approach could connect with accounts of preference as deriving from reasons (see Shafir et al. 1993, or Dietrich and List 2013a, b, for development of such an account and discussion of its philosophical antecedents). It could also connect with aspects of the psychological literatures on confidence judgement (see Keren 1991 for a survey, and Pleskac and Busemeyer 2010 for a more recent contribution) and constructed preference (see for example Slovic 1995).

As this discussion illustrates, there is more than one way of thinking about the sources of preference imprecision. But for our purposes in the main part of this paper, namely building stylised facts about empirical characteristics of measured imprecision, we did not need to commit to a particular theoretical account of imprecision. Indeed, one might suggest that our approach should be judged all the more successful the *less* it depends on such theoretical pre-commitments.

A second set of questions not intended to be addressed by our experimental design concern comparative empirical properties of different measures of preference imprecision. While our measurement tool has proved useful in pursuit of our objectives, we do not suggest that it provides a uniquely meaningful metric of imprecision. In fact, there is evidence to think that features of measured imprecision may vary considerably, and systematically, across different measurement tools. For example, Butler and Loomes (2007) find differences in the patterns of imprecision revealed for lotteries with money outcomes between measurements of imprecision made on a money scale and those made on a probabilistic scale. This suggests that we are not entitled to interpret

imprecision measured in our study (or others) as reflecting features of lotteries per se, but just as a representation of their imprecision *on a specific scale*.

In the realm of preference measurement, it is common to find that measured preferences vary with the procedure used to elicit them. A classic example of procedure-dependent preferences is the well-known case of preference reversal (Seidl 2000) where the ranking of a pair of alternative options varies systematically according to whether it is revealed through a pairwise choice task or by comparing valuations of the alternatives, reported separately, on some common monetary scale. As failures of procedure invariance like preference reversal are well-established in the literature on preference elicitation, it would be no surprise to see them reflected into the domain of measuring preference imprecision.

Our approach has been to look for coherence in the patterns of measured imprecision holding constant the scale of measurement within our study. Holding this scale constant was also important to our strategy for investigating whether violations of EUT are associated with imprecision. As noted in Section 2, in classic demonstrations of common ratio and common consequence effects, those effects have typically been identified by examining within-subject patterns of decisions across sets of pairwise choices. Had we followed this approach to identify individual-level violations of EUT, our failure to find association with imprecision might then have been a consequence of using different measurement procedures to observe preferences (i.e., pairwise choice data) compared with our procedure for evaluating individual-level imprecision (via our money valuation scale). We avoid this potential confound by testing for EUT violations in preferences inferred from CEs recorded on the same money scale as our imprecision measurements. So, while we are not entitled to assume that what we observe coincides with what would be observed using different elicitation techniques, we have found unambiguous evidence of EUT violations in CEs. Moreover, these violations reproduce the qualitatively familiar patterns of EUT violation typically observed in pairwise choice behaviour. Our failure to find much evidence of association between these violations and our measure of imprecision cannot plausibly be explained by a failure of procedure invariance. While our findings do not rule out the possibility of association between EUT violations observed in choice data and imprecision measured on some other scale, they obviously do not advance the case that the broad patterns of EUT violation are associated with imprecision in money valuations.

5 Conclusion

We began by posing three questions about measured imprecision intervals: their coherence; their stability; and their potential value added to economics. We have addressed them in the context of monetary valuation of lotteries, using a measure in which the imprecision intervals are also expressed on a monetary scale.

On the first two questions, our conclusions are largely positive. Imprecision intervals behave very coherently, in terms of location and size, in relation to objective characteristics of the lotteries being valued. We find support for the ideas suggested in different contexts by Butler and Loomes (2007, 2011) that the size of imprecision intervals is related positively to the range of non-dominated responses and to distance from certainty. But, the combination of these points, of course, implies that we do not find the size of imprecision intervals to be a simple fraction of the range of non-dominated responses,

except across a sequence of lotteries in which the probabilistic structure does not change. Our evidence suggests that imprecision intervals are quite stable: there is no significant tendency for their central location to move, nor for their size to either grow or fall, with experience. Taken together, these findings support the construct validity of our measure of imprecision. On the third question, our results are mixed. We find little evidence of any association, at the level of the individual, between imprecision and patterns in relative valuations that are analogous to classic violations of expected utility theory (despite clear evidence that those phenomena are exhibited by many of our subjects). This is ‘bad news’ for the idea that preference imprecision may provide a unifying explanation of a wide range of well-known choice anomalies, but ‘good news’ for the long tradition of research that has sought explanations of these phenomena in precise, but non-EU, preferences.

Our finding that measured imprecision *is* associated with variability in certainty equivalents is ‘good news’ in a different way. At the conceptual level, it provides empirical support for the conjecture, sometimes expressed in the literature, that imprecision of preference is connected to a feature of choice that is attracting increasing attention, namely its noisiness. The existence of this association may have some useful practical applications too. For example, as it is much easier to elicit an imprecision interval from a subject than to offer them repeated, incentivised, valuation tasks, it suggests that elicited imprecision intervals might play a useful instrumental role in experimental investigation of theories of stochastic choice. Finally, our results suggest that imprecision might be a useful construct to predict variability in contexts, such as surveys conducted with members of the general public, where repeated data are more problematic to obtain.

Acknowledgments We are grateful for the support of CeDEx at the Nottingham School of Economics; Cubitt and Starmer acknowledge the support of the ESRC funded Network for Integrated Behavioural Science (ES/K002201/1); and Navarro-Martinez acknowledges support from the Generalitat de Catalunya (2011BP-B00170) and the Spanish Ministry of Economy and Competitiveness (MINECO-PSI2013-41909-P). We thank Mohammed Abdellaoui, Tim Cason, Thomas Epper, Helga Fehr-Duda, Graham Loomes, Andrew Oswald, Larry Samuelson, Robert Sugden, Frans van Winden and Peter Wakker, as well as a referee and the Editor, for helpful comments. We would also like to thank participants at conferences, seminars and workshops where we have presented this work.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Allais, M. (1953). Le comportement de l’homme rationnel devant le risque: Critique des postulats et axiomes de l’École Américaine. *Econometrica*, 21, 503–546.
- Bardsley, N., Cubitt, R., Loomes, G., Moffatt, P., Starmer, C., & Sugden, R. (2010). *Experimental economics: Rethinking the rules*. Princeton: Princeton University Press.
- Becker, G. M., DeGroot, M. H., & Marschak, J. (1963). Stochastic models of choice behavior. *Behavioral Science*, 8, 41–55.
- Becker, G. M., DeGroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science*, 9, 226–232.
- Blavatskyy, P. (2007). Stochastic expected utility theory. *Journal of Risk and Uncertainty*, 34, 259–286.
- Block, H. D., & Marschak, J. (1960). Random orderings and stochastic theories of response. In I. Olkin et al. (Eds.), *Contributions to probability and statistics* (pp. 97–132). Stanford: Stanford University Press.

- Butler, D. J., & Loomes, G. (1988). Decision difficulty and imprecise preferences. *Acta Psychologica*, *68*, 183–196.
- Butler, D. J., & Loomes, G. (2007). Imprecision as an account of the preference reversal phenomenon. *American Economic Review*, *97*(1), 277–297.
- Butler, D. J., & Loomes, G. (2011). Imprecision as an account of violations of independence and betweenness. *Journal of Economic Behavior & Organization*, *80*, 511–522.
- Camerer, C. F. (1995). Individual decision making. In J. H. Kagel & A. E. Roth (Eds.), *The handbook of experimental economics*. Princeton: Princeton University Press.
- Chew, S.-H., & Waller, W. S. (1986). Empirical tests of weighted utility theory. *Journal of Mathematical Psychology*, *30*, 55–72.
- Clark, A. E., Frijters, P., & Shields, M. A. (2008). Relative income, happiness and utility: An explanation for the Easterlin Paradox and other puzzles. *Journal of Economic Literature*, *46*, 95–144.
- Cohen, M., Jaffray, J.-Y., & Said, T. (1987). Experimental comparison of individual behavior under risk and under uncertainty for gains and for losses. *Organizational Behavior and Human Decision Processes*, *39*, 1–22.
- Dietrich, F., & List, C. (2013a). Where do preferences come from? *International Journal of Game Theory*, *42*, 613–37.
- Dietrich, F., & List, C. (2013b). A reason-based theory of rational choice. *Noûs*, *47*, 104–134.
- Dubourg, W. R., Jones-Lee, M. W., & Loomes, G. (1994). Imprecise preferences and the WTP-WTA disparity. *Journal of Risk and Uncertainty*, *9*, 115–133.
- Dubourg, W. R., Jones-Lee, M. W., & Loomes, G. (1997). Imprecise preferences and survey design in contingent valuation. *Economica*, *64*, 681–702.
- Hanley, N., Kiström, B., & Shogren, J. F. (2009). Coherent arbitrariness: On value uncertainty for environmental goods. *Land Economics*, *85*(1), 41–50.
- Harless, D. W., & Camerer, C. F. (1994). The predictive utility of generalized expected utility theories. *Econometrica*, *62*(6), 1251–1289.
- Hey, J. D. (2005). Why we should not be silent about noise. *Experimental Economics*, *8*, 325–345.
- Hey, J. D., & Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica*, *62*(6), 1291–1326.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263–291.
- Keren, G. (1991). Calibration and probability judgements: Conceptual and methodological issues. *Acta Psychologica*, *77*, 217–273.
- Loomes, G. (2005). Modelling the stochastic component of behaviour in experiments: Some issues for the interpretation of data. *Experimental Economics*, *8*(4), 301–323.
- Loomes, G., & Sugden, R. (1995). Incorporating a stochastic element into decision theories. *European Economic Review*, *39*(3–4), 641–673.
- Loomes, G., & Sugden, R. (1998). Testing different stochastic specifications of risky choice. *Economica*, *65*, 581–98.
- Loomes, G., Moffatt, P., & Sugden, R. (2002). A microeconomic test of alternative stochastic theories of risky choice. *Journal of Risk and Uncertainty*, *24*, 103–130.
- Luce, R. D. (1958). A probabilistic theory of utility. *Econometrica*, *26*(2), 193–224.
- Luce, R. D. (1959). *Individual choice behavior*. New York: Wiley.
- MacCrimmon, K., & Smith, M. (1986). Imprecise equivalences: Preference reversals in money and probability. Working paper 1211, University of British Columbia.
- Morrison, G. C. (1998). Understanding the disparity between WTP and WTA: Endowment effect, substitutability, or imprecise preferences? *Economics Letters*, *59*, 189–194.
- Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review*, *117*, 864–901.
- Quandt, R. E. (1956). A probabilistic theory of consumer behavior. *Quarterly Journal of Economics*, *70*, 507–536.
- Read, D. (2005). Monetary incentives, what are they good for? *Journal of Economic Methodology*, *12*, 265–276.
- Seidl, C. (2000). Preference reversal. *Journal of Economic Surveys*, *6*, 621–655.
- Shafir, E., Simonson, I., & Tversky, A. (1993). Reason-based choice. *Cognition*, *49*, 11–36.
- Slovic, P. (1995). The construction of preferences. *American Psychologist*, *50*, 364–371.
- Starmer, C. (2000). Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature*, *38*, 332–382.
- Tversky, A., Slovic, P., & Kahneman, D. (1990). The causes of preference reversal. *American Economic Review*, *80*, 204–217.

- Wakker, P.P. (2010). *Prospect theory for risk and ambiguity*. Cambridge University Press.
- Wilcox, N. T. (2008). Stochastic models for binary discrete choice under risk: A critical primer and econometric comparison. In J. C. Cox & G. W. Harrison (Eds.), *Research in experimental economics. Volume 12: Risk aversion in experiments*. Greenwich: JAI Press.
- Wu, G., & Gonzalez, R. (1996). Curvature of the probability weighting function. *Management Science*, *42*, 1676–1690.
- Wu, G., & Gonzalez, R. (1998). Common consequence conditions in decision making under risk. *Journal of Risk and Uncertainty*, *16*, 115–139.
- Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, *13*, 75–98.