

Revisiting the diagnosis of intertemporal preference reversals

Li, Zhihua; Loomes, Graham

DOI:

[10.1007/s11166-022-09369-w](https://doi.org/10.1007/s11166-022-09369-w)

License:

Creative Commons: Attribution (CC BY)

Document Version

Publisher's PDF, also known as Version of record

Citation for published version (Harvard):

Li, Z & Loomes, G 2022, 'Revisiting the diagnosis of intertemporal preference reversals', *Journal of Risk and Uncertainty*, vol. 64, no. 1, pp. 19-41. <https://doi.org/10.1007/s11166-022-09369-w>

[Link to publication on Research at Birmingham portal](#)

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.



Revisiting the diagnosis of intertemporal preference reversals

Zhihua Li¹ · Graham Loomes²

Accepted: 10 January 2022
© The Author(s) 2022

Abstract

Intertemporal preference reversals occur when individuals choose future option A over future option B in a direct choice between the two but place a higher ‘immediate cash’ value on B than on A. Tversky et al. (1990) reported strong evidence of such reversals, which they attributed mainly to valuation biases rather than intransitivity. We find similar levels of reversals, even after adjusting for considerable degrees of variability and imprecision in people’s responses. However, we disagree with Tversky et al.’s conclusions about the causes of the majority of these reversals. We find substantial levels of intransitivity in respondents’ binary choices as well as differential overvaluation of both options relative to the values inferred from their choices.

Keywords Intertemporal choice · Preference reversals · Decision experiment

JEL Classification C91 · D15 · D91

1 Introduction

The preference reversal (PR) phenomenon is probably best known in the context of risky choice, where it refers to the evidence that individuals often choose an option with lower risk and smaller returns (labelled a *P-bet*) over an alternative option with greater risk and larger returns (labelled a *\$-bet*), while also placing a higher certainty equivalent value on the *\$-bet* than on the *P-bet* when evaluating the two options separately. The opposite anomaly – choosing the *\$-bet* over the *P-bet* in the choice task but valuing the *P-bet* more highly – is relatively rarely

✉ Graham Loomes
g.loomes@warwick.ac.uk
Zhihua Li
Z.Li.8@bham.ac.uk

¹ Department of Economics, Birmingham Business School, University of Birmingham, Birmingham B15 2TT, UK

² Warwick Business School, University of Warwick, Coventry CV4 7TL, UK

observed. This phenomenon was reported half a century ago by Lichtenstein and Slovic (1971) and Lindman (1971) and has been replicated many times since (see Seidl, 2002, for a review).

A number of different explanations for the phenomenon have been proposed. For example, such a pattern of response might be accommodated by some general deterministic theory that allows systematic intransitivity (Loomes & Sugden, 1983). Alternatively, it has been suggested that preferences, far from being deterministic, are often rather imprecise, and that such imprecision may produce the observed asymmetries (Butler & Loomes, 2007; MacCrimmon & Smith, 1986). Another possible implication of the imperfectly formed nature of people's preferences is that responses to different tasks (e.g. choice as distinct from valuation) are 'constructed', often using somewhat diverse cognitive processes subject to different influences or biases (see Tversky et al., 1988). From this latter perspective, valuations may focus more on payoffs, thereby favouring the higher-payoff \$-bet, while choices may place more weight on the probabilities of receiving at least some positive return, thereby favouring the P-bet relatively more strongly.

Tversky et al. (1990) – hereafter, TSK – tried to disentangle some of the possible causes of PR. Their Study 1 used a number of combinations of \$-bets, P-bets and sure amounts, on the basis of which they concluded that nearly two-thirds of the observed reversals were due to overvaluation of the \$-bet, with some smaller yet substantial contribution from undervaluation of the P-bet, but with only a minor part of the phenomenon attributable to intransitivity (this latter accounting for no more than 10% of reversals, according to their diagnosis).

To investigate the generality of their 'overvaluation-undervaluation' diagnosis, TSK conducted a second experiment, looking at time preferences rather than risky choices. Their Study 2 design revolved around five amounts to be received (hypothetically) at various times in the future, ranging from \$3550 in 10 years' time to \$1525 in six months, as well as two levels of immediate cash amounts (\$1250, \$1350). They combined these options to produce four triples, each involving a larger sum to be paid after a longer delay (which we shall refer to as the LargerLater or LL option, the intertemporal analogue to the \$-bet), a somewhat smaller sum to be received after a shorter delay (henceforth the Smaller-Sooner or SS option, analogous to the P-bet) and a lower present cash amount C (analogous to a certainty). They found an even higher rate of reversals, which in this context meant placing a higher present cash value on the LL option than on the SS option but picking the SS option in a straight choice between the two. The diagnosis of causes was much the same as in their risky choice Study 1: that is, while about 15% of reversals were compatible with intransitivity, approximately 55% were attributed to overvaluing LL, with the remaining 30% being due either to undervaluing SS or to some combination of overvaluing LL and undervaluing SS relative to choice.

The fact that broadly similar results were found in the domains of risk and time in two large studies using a novel design was highly influential. TSK's diagnostic procedure has been widely endorsed (see Seidl, 2002 for example), the paper has

been cited more than 1,200 times and their conclusions about the causes of PR have become the accepted wisdom.

However, we suggest that there may be reasons to be cautious about those conclusions. The high rates of undervaluation of the P-bet/SS options appear to be somewhat at odds with the ‘anchoring and adjustment’ and ‘contingent weighting’ models offered by Lichtenstein and Slovic (1971) and Tversky et al. (1988). These models would entail something more like the results of a risky choice study by Loomes and Pogrebna (2017), who found that *both* types of bets were overvalued relative to choice, although the *degree* of overvaluation of \$-bets was very much greater than the *degree* of overvaluation of P-bets, a disparity which was sufficient to produce many reversals.

On closer inspection, TSK’s diagnostic procedure appears to be problematic on two counts, each of which will be examined in more detail in Sect. 5. First, TSK’s diagnostic system discarded just under half of the reversals in the risky choice study and almost two-thirds of the reversals in the intertemporal choice study, so that their conclusions were based on a non-random subset of observations. Second, because of the way it is constructed, the system has an inbuilt bias such that it *simply cannot detect* either undervaluation of the \$-bet/LL options or overvaluation of the P-bet/SS options.

The present study revisits the accuracy of TSK’s attributions of cause, using an experiment which adapts and extends a design developed by Loomes and Pogrebna (2017). We suggest that our study improves on TSK’s classic study in four important ways: (i) it uses *all* of the observed reversals, not just a subset; (ii) it can measure both undervaluation and overvaluation for both options; (iii) it makes allowance for the noise and imprecision in people’s responses – and indeed, it sheds new light on the degree of such imprecision in the area of intertemporal choice, where there are very few such data at the moment; and (iv) it explores the use of an additional instrument – the choice list – that has become popular among experimenters to supplement (or even replace) standard binary choice and direct valuation methods (Cheung, 2015; Laury et al., 2012).

In the next section, we describe the experimental design and its rationale in more detail. In Sect. 3, we present the results, with particular reference to points (i) to (iii) in the previous paragraph as they relate to direct valuation and binary choice, comparable with the original TSK study. In Section 4, we summarise the main additional insights provided by the choice list instrument. Whether we use binary choices or choice lists, our data suggest a substantially different attribution of causality than that proposed by TSK. Section 5 considers why this might be the case. Section 6 concludes with a discussion not only relating to the specifics of this study but also reflecting on some broader implications. The preference reversal phenomenon is more than an experimental curiosity. The systematic disparity between valuation and choice has implications for theory, for the interpretation of experimental and survey data, and for applications to public policy in areas such as health, safety and environmental goods – issues to which we return in the final section.

2 Experimental design

We conducted two experiments. The first was a ‘range-finder’, helping us to identify the spread of parameters and the number of repetitions necessary to cover the responses of the great majority of participants.¹ The second experiment took account of the knowledge gained from the first. However, the basic features of the design were the much same for both experiments and the results were broadly similar, so we focus attention in this paper on the second, larger and more finely-tuned, study. The results of the range-finder were available to reviewers and can be obtained from the authors upon request.

Although some explanations of PR are compatible with deterministic models, others involve some degree of indeterminacy or imprecision in people’s preferences. In order to obtain some indication of the variability in people’s preferences and make appropriate allowance for it in our analysis, we took a single pair of future options (as opposed to the four pairings used by TSK) and repeated each of our questions about that pair four times, with repetitions widely separated and interspersed among a number of risky choice and investment decisions which were intended to serve as ‘distractor’ tasks aimed at providing greater variety and more independence between each of the intertemporal decisions.

The pair we chose was the one in which the SS option offered \$1600 in 1.5 years’ time, while the LL option offered \$2500 in 5 years’ time. We chose that pair because the division of choices was the least extreme and because the rate of preference reversal was closest to the average (57% chose SS in TSK’s study and the preference reversal rate was 49%). Because we were recruiting from a UK sample, we changed the currency sign to £. Keeping the numerical magnitudes of the amounts and their timing in line with the TSK study precluded the use of real incentives.²

As is standard in a preference reversal experiment, there were three tasks, presented in varied order: (i) to make a straight choice between SS and LL; (ii) to state the present cash value (PV) of the SS option (i.e. how much money, received today, the respondent would regard as exactly as good as receiving \$1600 in 1.5 years’ time); and (iii) to state the PV of the LL option. We now explain in more detail how each type of task was implemented.

¹ At the time we conducted this first range-finder experiment – in August 2016 – we were unaware of any other experiments that had asked respondents to make repeated intertemporal choices and so we had no prior stock of evidence to inform our design in this respect. Our first exploratory study presented each choice twice and covered a limited range of values. On the basis of the results, we considered that just two presentations gave an insufficiently rich picture of the degree of variability in people’s preferences and we realised that we needed to extend the range of values. We have since become aware of a paper (Blavatskyy & Maafi, 2018) that reports a rather different set of intertemporal choices, each presented twice and thereby also giving some indications of variability, the degree of which is liable to differ from one individual to another and from one question to another. However, those questions were sufficiently dissimilar from ours that, even if we had known about that study, it could not have informed us adequately about the parameters of our design.

² Researchers face a dilemma between using real incentives involving smaller payoffs over shorter periods of time – say, £16 in 6 weeks’ time vs £25 in 5 months – and using hypothetical questions with larger amounts and longer delays. Because we wished to achieve a high level of comparability with the TSK study, we opted for the latter.

A. Getting £1600 in 1.5 years <input checked="" type="radio"/> Definitely A <input type="radio"/> Probably A	B. Getting £2500 in 5 years <input type="radio"/> Probably B <input checked="" type="radio"/> Definitely B
--	--

Fig. 1 Example of a Binary Choice (BC) Question. Respondents were asked to click on the radio button that indicated which option they would choose and how confident they felt that this was the better option for them

2.1 Binary Choice (BC)

On four different occasions, each respondent was asked to make a straight choice between SS and LL. However, in our study, they were not simply asked to choose but were asked to give an indication of the *degree of confidence* they felt about their choice: that is, whether they ‘Definitely’ preferred an option or ‘Probably’ preferred it (the distinction between these two was explained in the Instructions, which are reproduced in Figs. A1, A2, and A3 in the Appendix). Figure 1 provides a screenshot of the question. In order to answer it, a respondent clicked on one of the four radio buttons. He/she could change the answer by clicking on a different button. When satisfied with the answer, the respondent confirmed and was presented with the next task.

Exactly the same format was used to elicit choices between each of the future options and a number of present values which were multiples of £100 from £400 to £1400 inclusive. Our selection of these lower and upper amounts was informed by the range-finding experiment which indicated that repeated presentations of amounts outside this range would have too little discriminatory power to justify the additional demands on respondents’ time and effort.

So, in the course of the experiment a respondent was asked to choose between SS and each of eleven different present cash amounts, with each of those BCs presented on four separated occasions during a session. The respondent was also asked to choose between LL and those same eleven present amounts; and again, each choice was presented four times.

2.2 Direct Valuation (DV)

A second type of task was a direct valuation (DV) question which used a slider such as the one shown in Fig. 2.

By clicking on and sliding the button on the bar, respondents were able to indicate, within a hundred pound range, what sum of money today they would consider to be just as good as getting the future option – in this example, SS. Moving the button changed the numbers in the three statements as shown in Fig. 2. The two statements below the bar try to frame the decision in a way that is analogous to choice. If the respondent felt dissatisfied with any of the statements, he/she could adjust the position of the button until all three statements were acceptable – at which point, he/she confirmed the decision. We took the mid-point of the confirmed range as the estimate of the PV.

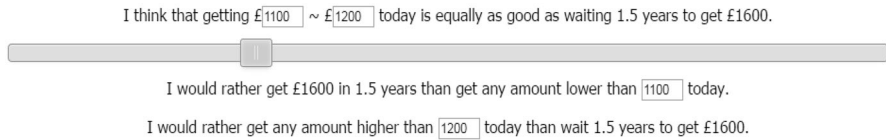


Fig. 2 Example of a Direct Valuation (DV) Question. Respondents were asked to slide the button to the position where the amounts appearing in the boxes best expressed their preferences

This method is most like the way in which values are usually elicited, although it tries to emphasise equivalence and makes the analogous choice implications explicit, whereas some direct formats couch the exercise in terms of the maximum amount respondents would be willing to pay to acquire an option or else in terms of the minimum amount they would be prepared to accept in exchange for selling the option. For example, Freeman et al. (2016) used two types of direct valuation procedure (one linked to the Becker et al. (1964) incentive mechanism, the other involving a second-price auction) which asked respondents to state the lowest sum they would prefer to receive tomorrow instead of receiving €20 at a later date. Likewise, TSK asked participants to state, hypothetically, “the smallest immediate cash payment for which they would be willing to exchange the delayed payment” (p.231). Since we, like TSK, were not using an incentive-compatible mechanism, we tried to avoid possible additional framing effects associated with willingness-to-pay/accept formulations and used a more neutral equivalence format which, if anything, would be likely to reduce any tendency for the willingness-to-accept framing to encourage overvaluation.

To obtain some measure of the variability of individuals’ responses, the DV question for SS was asked on four occasions spread throughout the experiment and separated by numerous intervening tasks. Likewise, the DV question for LL was asked on four occasions, separated from each other and from the corresponding SS question. This gives us the option of analysing an individual’s PVs in terms of the means or medians of the four responses for each option, while the ranges and standard deviations provide some measure of within-person variability.

2.3 Choice Lists (CL)

Our third elicitation method took the form of a choice list (CL) such as the one shown in Fig. 3. Each choice list for SS (and likewise for LL) consisted of an ordered set of the eleven BCs described in Sect. 2.1. As with those BCs, respondents were asked not only to choose an option on each row but also to indicate how confident they felt about each choice. The procedure required one radio button to be clicked on each row, subject to the constraint that responses could not move between columns in a way consistent with a poorer option being more likely to be chosen. So a respondent working steadily down the list in Fig. 3 was not allowed by the program to move right-to-left in any transition from one row to the row below it. An individual could respond to rows in whatever order they wished and could change previous responses, so long as the above constraint was respected.

Row no.	Decision		Option A	Option B	Decision	
1	<input type="radio"/> Definitely A	<input type="radio"/> Probably A	£1400 today	£1600 in 1.5 years	<input type="radio"/> Probably B	<input type="radio"/> Definitely B
2	<input type="radio"/> Definitely A	<input type="radio"/> Probably A	£1300 today		<input type="radio"/> Probably B	<input type="radio"/> Definitely B
3	<input type="radio"/> Definitely A	<input type="radio"/> Probably A	£1200 today		<input type="radio"/> Probably B	<input type="radio"/> Definitely B
4	<input type="radio"/> Definitely A	<input type="radio"/> Probably A	£1100 today		<input type="radio"/> Probably B	<input type="radio"/> Definitely B
5	<input type="radio"/> Definitely A	<input type="radio"/> Probably A	£1000 today		<input type="radio"/> Probably B	<input type="radio"/> Definitely B
6	<input type="radio"/> Definitely A	<input type="radio"/> Probably A	£900 today		<input type="radio"/> Probably B	<input type="radio"/> Definitely B
7	<input type="radio"/> Definitely A	<input type="radio"/> Probably A	£800 today		<input type="radio"/> Probably B	<input type="radio"/> Definitely B
8	<input type="radio"/> Definitely A	<input type="radio"/> Probably A	£700 today		<input type="radio"/> Probably B	<input type="radio"/> Definitely B
9	<input type="radio"/> Definitely A	<input type="radio"/> Probably A	£600 today		<input type="radio"/> Probably B	<input type="radio"/> Definitely B
10	<input type="radio"/> Definitely A	<input type="radio"/> Probably A	£500 today		<input type="radio"/> Probably B	<input type="radio"/> Definitely B
11	<input type="radio"/> Definitely A	<input type="radio"/> Probably A	£400 today		<input type="radio"/> Probably B	<input type="radio"/> Definitely B

Fig. 3 Example of a Choice List (CL) Question. With Option B fixed and Option A varying from row to row, respondents were asked to click on one radio button in each row to indicate which option they would choose in that row and how confident they felt that this was the better option for them

For comparability with the other two methods, each CL was presented on four separated occasions during the session.

2.4 Inferring PVs from binary choices and choice lists

For each of SS and LL, we have 44 BCs ranging over present amounts from £400 to £1400 inclusive. With these data, our method for inferring an estimate of an individual's PV for an option is analogous to the one used in Loomes and Pogrebná (2017) and is based on counting the number of occasions that the present sum is chosen.

To illustrate, consider the case of an individual with precise deterministic preferences whose PV for the LL option is £1150. Such an individual would choose the present amounts whenever those amounts are £1400, £1300 or £1200, but would not choose any present amount of £1100 or less. Thus, choosing the present amount on a total of 12 occasions out of 44 would correspond to a PV of £1150.

However, an individual whose preferences exhibit the kind of variability consistent with imprecise or 'noisy' preferences is liable, for at least some amounts, to choose the present sum on some occasions and the future option on others. Suppose, for example, that such an individual chooses '£1400 today' on all four occasions when that sum is presented and never chooses the present amount when it is £900 or less; but for present amounts of £1300, £1200, £1100 and £1000, he/she chooses the present amount three times, twice, twice and once out of the four repetitions at each level. In this example, our best estimate of the point of *stochastic indifference* (i.e. when the chances of choosing either alternative are 0.5) would be a *stochastic present value* (SPV) of £1150, reflecting the fact that the present amount has been chosen on twelve occasions in total over the full range of values presented. Since every four observations of the present amount being chosen over the future option

reduces the SPV by £100, we can estimate each individual's SPV for an option as $SPV = 1450 - 25A$ where $0 < A < 44$ is the total number of times the present amount is chosen in the 44 BCs involving that option.

Given a degree of stochasticity in people's responses and the limited number of repetitions of each choice, individual patterns will not always be as neatly gradated as in our example and a formulation based on a simple count will necessarily be subject to sampling error. Still, we can expect that, on average, the inverse relationship between SPV and A will apply in the range $0 < A < 44$. In cases where $A = 0$, we infer that the SPV is *at least* £1450; and in cases where $A = 44$ we infer that the SPV is *no greater than* £350. However, our range-finding experiment suggested – correctly, as it turned out – that the great majority of respondents would exhibit $0 < A < 44$.

In addition, as an index of the variability of an individual's preferences, we can take the number of levels of present amount where the individual switched answers at least once between the future option and the present amount. In the case of the deterministic individual in our example above, this measure takes a score of 0, since that person either chose the present sum in all four repetitions at each level where the sum was £1200 or more, or else chose the future option in all four repetitions when the present sum was £1100 or less. For the individual in our example who exhibited some variability, the measure takes a score of 4 since there were four levels of present amount (£1300, £1200, £1100 and £1000) where each alternative was chosen at least once in the course of the four repetitions.³ The size of this score gives a broad indication of the extent of what we shall call the *imprecision interval* – that is, the range over which variability of preference is exhibited.

We can compute an individual's CL-based SPV via the same formula used with the BCs. Since each row is one of those BCs, the four lists can be regarded as constituting a total of 44 BCs and from these we can count the number of times that a present amount was chosen and use this total as the A in that formula. Moreover, repeating the CL task four times allows a measure of the imprecision interval analogous to the one derived from the BCs: by comparing across an individual's four CLs, we can count the number of rows where the present amount was chosen on at least one occasion while the future option was chosen on at least one other occasion.⁴ Thus comparisons can be made between the BC-based and CL-based responses to see whether the two methods elicit broadly similar data or whether there are any systematic differences between them. These comparisons will be discussed in Sect. 4.

Meanwhile, the next section presents the results most directly pertinent to TSK's study: namely, the data from the binary choice and direct valuation questions. The first part of that section gives an overview of the aggregate data. The second part reports the evidence of variability/imprecision at the individual level. The third part

³ Of course, such a measure may not be appropriate in cases where one or other option is always chosen at every value in the £1400–£400 range: the fact that the measure gives a score of 0 in these cases does not necessarily indicate deterministic preferences, since there may be unobserved variability around an SPV that lies outside our range.

⁴ For reasons analogous to those discussed in footnote 3, this measure is not appropriate if there are cases where the individual made no within-list switch from one option to the other.

of the section examines the existence and robustness of preference reversals. Three questions are addressed in that subsection. First, do we observe the PR phenomenon reported by TSK when binary choices between SS and LL are compared with central tendency measures of responses to the DV questions? Second, do the reversals persist if valuations are derived as SPVs from the BCs? Third, do our data support or modify or contradict the conclusions about the causes of PR proposed by TSK?

3 Results: Binary choice and direct valuation

3.1 Overview

122 individuals participated in the experiment, which was run online via Prolific Academic on February 11, 2017. The average time taken was 32 min and each participant received a £4 flat payment. There were 6 individuals whose data were incomplete, so that the initial analysis is based on 116 complete sets of responses.

The direct choice between SS and LL was repeated four times per individual, giving a total of 464 responses to this question. Of these, 167 stated a definite preference (henceforth Def) for SS, 176 gave a probable preference (Prob) for SS, 90 probably preferred LL and 31 registered a definite preference for LL. Overall, 74% of responses favoured SS. The majority of both choices were recorded as Prob rather than Def, with Prob responses constituting 57% of the total of 464.

Figure 4 displays the aggregate distribution of responses to the BC and DV questions. In the top half of Fig. 4, each column shows the distribution of the 464 responses to the binary choice between the future option in question and the particular present cash amount shown on the horizontal axis, with the SS data on the left and the LL data on the right. The black blocks show the number of responses indicating a Def preference for the future payment; striped darker grey blocks indicate a Prob preference for the future option; dotted light grey blocks denote Prob preference for the present amount; and the pale plain grey blocks depict Def preference for the present amount. The dashed line divides the number of observations in half in order to make it easier to identify median responses.

There was no Def-Prob distinction for the DV questions, so in the bottom half of Fig. 4 we show the choices inferred from the statements recorded in those questions: for example, an individual who gave the response shown in Fig. 2 was recorded as having a PV of £1150 and therefore was deemed to choose the SS option over all present amounts from £400 to £1100 inclusive while choosing £1200 or more over SS.

First, compare the top left and top right panels, respectively showing the distributions of BC responses when the SS and LL options are offered against the various cash amounts. SS is chosen more often than LL when each is compared with cash amounts at the lower end of the range, but the opposite is the case when they are each compared with present amounts of £1000 or more, with choices involving LL displaying wider Prob intervals over this range. However, at a present sum of £900, the two more or less coincide, with almost equal splits between the future option and that cash amount, suggesting a sample median SPV for both SS and LL of

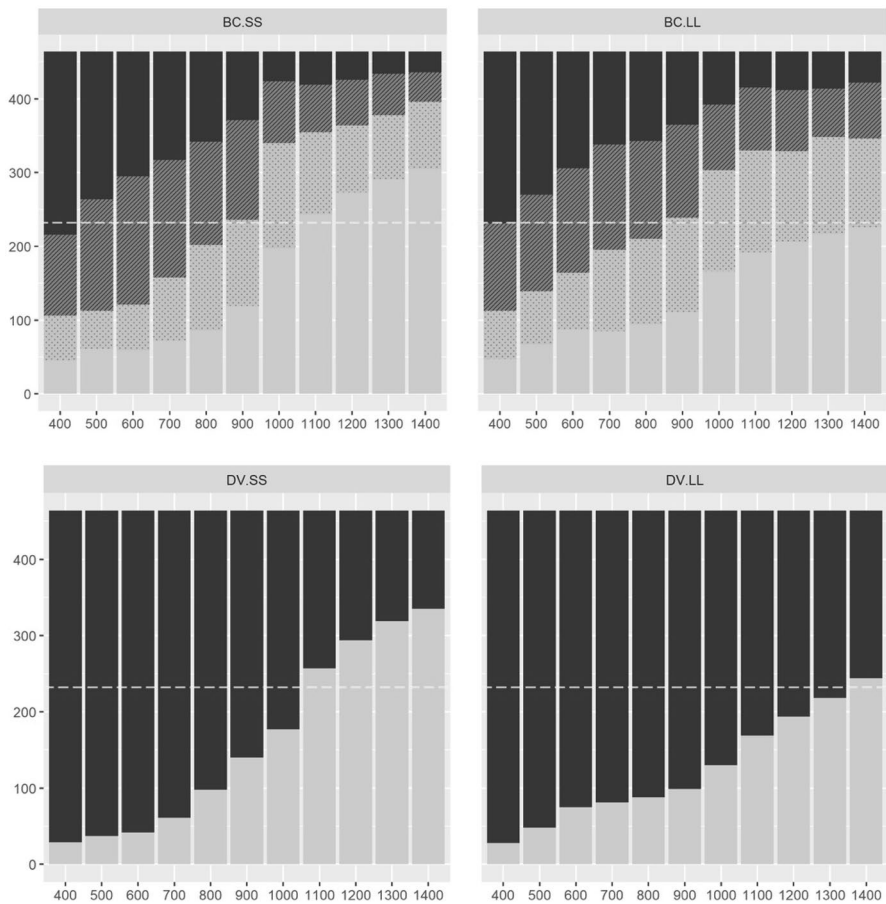


Fig. 4 Aggregate Responses to Binary Choice (BC) and Direct Valuation (DV) Questions. Numbers of responses (max 464) are calibrated on the vertical axes. Horizontal axes show the different levels of immediate cash amounts. The columns in the BC.SS panel show, for each cash amount, the frequencies of Binary Choice responses where the SmallerSooner option was Definitely preferred (black) or Probably preferred (darker grey) or where the immediate amount was Probably preferred (lighter grey) or Definitely preferred (pale grey). The BC.LL panel shows the corresponding distributions for the Binary Choices between the LargerLater option and the various cash amounts. The DV.SS and DV.LL panels show the preferences inferred from the Direct Valuation responses for the SmallerSooner and Larger-Later options respectively: here there was no Definitely/Probably distinction, so the black part of each column shows inferred preference for the future option while the pale grey part shows inferred preference for the cash sum. The horizontal dashed line marks the median response

something close to £900. This is at odds with the data for straight choices between SS and LL, where 74% of responses prefer SS over LL. In due course we shall see how this manifests at the individual respondent level, but these aggregate data are consistent with the standard preference reversal asymmetry whereby the SS option is favoured more frequently in choice than in valuation (with the valuations in this case being the SPVs inferred from the BC questions).

We now turn to the DV panels. A striking feature is that, for both SS and LL, it is as if virtually all of the dotted light grey blocks – that is, the BC-based responses Prob choosing the present amounts – have been reassigned to favour the future options when direct valuations are elicited. One interpretation is that when people feel unsure about their preferences, the DV method systematically influences them to overvalue *both* of the future options relative to their BC responses.

3.2 Individual variability/imprecision

The Prob intervals reported in the BC panels of Fig. 4 reflect participants' subjective judgments of their (lack of) confidence in their decisions; and, as Fig. 4 shows, the proportions of such responses are substantial, often equalling or outnumbering the Def responses, especially in BC questions involving present amounts towards the middle of the range.

To what extent is this reflected by the observed variability in the decisions made? In Sect. 2.4 we proposed an index of the degree of imprecision based on counting the number of levels of present amount where an individual chose differently on different presentations of the same binary choices. Figure 5 shows that for the BC-based distributions, it is only a minority – 19 or 20 out of 116 – whose choices could be deemed consistent with deterministic preferences, while more than half of the sample exhibit variability at three or more levels of present amount.

Turning to individuals' direct valuations, Fig. 6 shows all 116 median DV responses for SS (upper panel) and LL (lower panel), organised in descending magnitude from left to right and with the range of each individual's four DV responses indicated by the brackets.

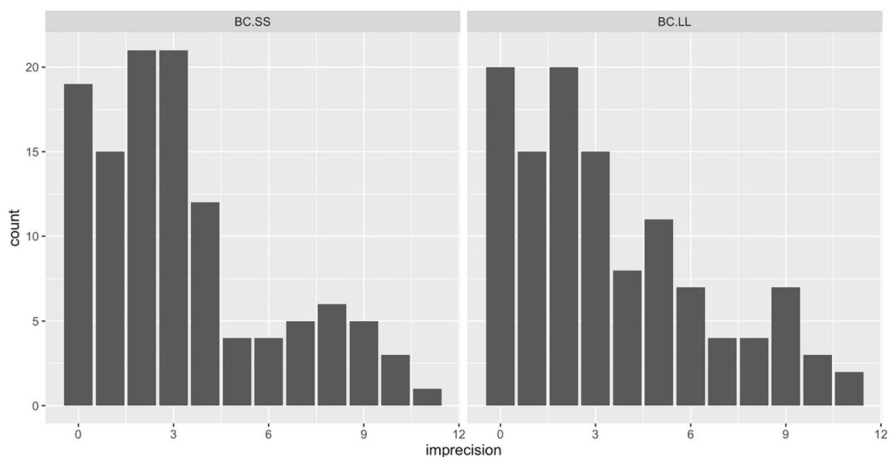


Fig. 5 Histograms of Individuals' Variability of Choice. The vertical axis shows the numbers of individuals. The horizontal axis shows the degree of variability/imprecision as reflected by the number of levels of present amount where an individual chose differently in different repetitions of the same pairs. The BC.SS panel reports the data for Binary Choices between the various immediate cash amounts and the SmallerSooner option and the BC.LL panel does likewise for the LargerLater option

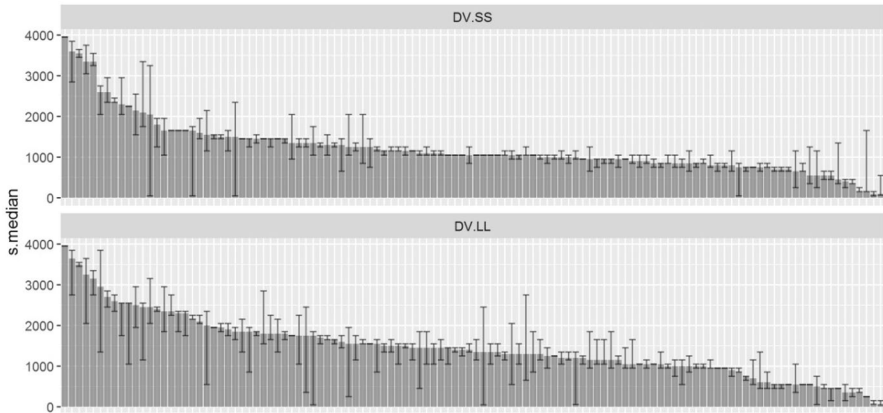


Fig. 6 Individual Median Direct Valuation Responses (with Individual Ranges). The vertical axis calibrates individual subjects' median cash equivalents for the SmallerSooner (upper panel) and LargerLater (lower panel) options. Individuals' medians are sorted from highest to lowest, with individuals' maximum and minimum responses depicted by the brackets

In total, 20 participants display a zero range in their SS responses and 14 display a zero range in their LL responses, with 4 individuals in both categories. Thus with DV, as with BC, we see that only a small minority give responses consistent with deterministic preferences. Since the slider was not constrained to the same range as the set of present amounts used in the BC and CL tasks, some of the ranges are comparatively large, although most are modest.⁵ As with the imprecision indices for the BC questions, the brackets show greater variability for LL than for SS.

Overall, it is clear that most individuals exhibit variability in their responses to both BC and DV tasks, suggesting that any analyses of decision behaviour are likely to be more robust if they make allowance for such imprecision. So in the next subsection we base our analysis of the prevalence and causes of PR upon the *central tendencies* of repeated choices and valuations, as reflected by individuals' SPVs and median DVs.⁶

3.3 Preference reversals

For the purposes of the analysis in this subsection, we focus upon the 102 respondents for whom we are able to compute estimates of their SPVs from both

⁵ There are a few individuals whose DV responses exceed the money amounts offered in the future options. This might be interpreted as exhibiting negative discounting; or as a violation of dominance if negative discounting is not allowed; or simply some misunderstanding of the question. However, removing these individuals does not change our key conclusions.

⁶ We use the median rather than the mean because it is less susceptible to individual outliers that may be the result of error or misunderstanding and is more comparable with the 'counting' method used to estimate SPVs on the basis of BC responses. However, using individual means of DV responses does not give a significantly different picture overall.

Table 1 Summary Statistics for Each Option by Each Elicitation Method (£). The BC-SS column shows the sample ($n=102$) mean, median and standard deviation of the SPVs derived from the Binary Choices between the SmallerSooner option and the various cash amounts. The BC-LL column shows the corresponding data for the LargerLater option. Each individual's DV is the median of his/her four Direct Valuation responses, so the columns DV-SS and DV-LL report the sample means, medians and standard deviations of those measures for the SmallerSooner and LargerLater options

	BC-SS	DV-SS	BC-LL	DV-LL
Sample Mean	878.9	1244.6	891.9	1500.0
Sample Median	900.0	1050.0	875.0	1425.0
Standard Deviation	230.1	600.9	302.0	670.2

their BC and their CL responses: that is, we exclude individuals for whom $A=0$ or for whom $A=44$ for both future options, since in these cases we cannot infer an ordering.

Table 1 provides summary statistics, reporting the sample means, medians and standard deviations for each option by both BC and DV. These data reinforce, at the level of individual values, what earlier figures showed in terms of aggregate responses. For BC, the mean and median values for LL are not significantly different from the mean and median values for SS, so if we were relying on these figures to reflect preferences, we should conclude that, on average, SS and LL were equally preferred, even though 70 of the 102 respondents chose SS over LL at least three times out of four in a straight choice between the two.

For SS, the sample DV mean is 41.6% higher than the BC mean and the DV median is 16.7% higher than the BC median. For LL, the DV mean is 68.2% higher than its BC counterpart and the median DV is 62.9% higher. The differences between BC and DV SPVs are highly significant for both SS ($p<0.001$) and LL ($p<0.001$). Whereas the BC-based measures showed no significant difference between the means for SS and LL, the DV method gave significantly higher values for LL than SS ($p<0.001$), with 74 of the 102 respondents recording DV-LL strictly greater than DV-SS.

Table 2 shows the various combinations of choices and SPV differences. Each individual made four straight choices between SS and LL, and the rows show the different mixtures of those choices. The first two columns show, for each elicitation method, the cases where the SPV for SS is strictly greater than the SPV for LL; the next two columns show equal SPVs; then two columns show cases where the SPV was higher for LL than for SS. The final column provides row totals, which are the same for both methods, although distributed differently.

Cases corresponding with the standard form of PR – that is, where SS was chosen strictly more often than LL in straight choices between the two but where the value for LL was strictly higher than the value for SS – are located in the four cells in the upper-right area of the table. Cases exhibiting the opposite anomaly are in the four lower-left cells.

The asymmetry between the two forms of reversal is clear. For the method closest to standard value elicitation, DV, the ratio of reversals is 47:1, which is very

Table 2 Cross-tabulation of Choices and Value Differences for Each Elicitation Method. Columns show the numbers of individuals whose value of the SmallerSooner (SS) option, Val_{SS} , is either greater than, or equal to, or else less than their value of the LargerLater (LL) option, Val_{LL} , when derived either according to the Direct Valuation (DV) method or else the Binary Choice (BC) based method. The rows indicate how many times (out of 4) those individuals chose the SS option in a direct choice with the LL option

Number of choices of SS	$Val_{SS} > Val_{LL}$		$Val_{SS} = Val_{LL}$		$Val_{SS} < Val_{LL}$		Row totals for each method
	DV	BC	DV	BC	DV	BC	
4	12	26	3	1	38	26	53
3	5	5	3	0	9	12	17
2	2	4	1	2	9	6	12
1	1	2	0	0	10	9	11
0	0	0	1	0	8	9	9
Column totals	20	37	8	3	74	62	102

close to the rate reported by TSK. Basing the SPVs on BCs reduces the asymmetry somewhat – to 38:2 – but the phenomenon is far from eliminated and constitutes a rate of violation of weak stochastic transitivity⁷ that is much higher than the rate of intransitivity detected by TSK.

To check how robust these results are to the imprecision of people's responses, we define $Val_{dif} = Val_{SS} - Val_{LL}$ and consider how things change if we require Val_{dif} to be *strictly greater* than £25 in either direction (£25 being the size of the increment in the formula for computing SPVs from BC responses). Table 3 shows the results. This allowance for imprecision leaves the DV-based reversals unaffected but reduces the numbers of reversals for the BC treatments. Nevertheless, the asymmetry is still 29:1.

So we confirm TSK's finding that systematic preference reversals occur to a substantial degree in intertemporal decisions. Answering the first question posed at the end of Sect. 2, we observe the phenomenon at about the same level as TSK when majority choices between SS and LL are compared with median DV responses. Answering the second question, the reversals persist to only a *slightly* smaller extent if valuations are derived as SPVs from choice-based methods involving binary choices. The asymmetry remains substantial even if we treat SPV differences of £25 or less in either direction as signifying indifference.

In answer to the third question, our data appear to offer a different picture from TSK's in terms of causality. Recall that TSK's diagnostic method suggested that the majority (55%) of reversals were attributable solely to overvaluing LL, with another 30% caused

⁷ For any three options $\{X, Y, Z\}$, weak stochastic transitivity (WST) requires that if $\Pr(X > Y) \geq 0.5$ and also $\Pr(Y > Z) \geq 0.5$, then $\Pr(X > Z) \geq 0.5$. In cases where $Val_{LL} > Val_{SS}$, we can find a present value V^* that lies between the two, which entails $\Pr(LL > V^*) \geq 0.5$ and $\Pr(V^* > SS) \geq 0.5$. WST then requires $\Pr(LL > SS) \geq 0.5$. However, the top two rows of the table show cases where SS was chosen strictly more often than LL in straight choices between the two. Hence the observations in the upper-right cells for BC do not conform with WST.

Table 3 Cross-tabulation of Choices and Value Differences with £25 Thresholds. Columns show the numbers of individuals for whom the difference between the value of the SmallerSooner (SS) option, Val_{SS} , and the value of the LargerLater (LL) option, Val_{LL} , is either strictly greater than £25, or in the range from £25 to -£25 inclusive, or else strictly less than -£25, when those values are derived according to the Direct Valuation (DV) or the Binary Choice (BC) based method. The rows indicate how many times (out of 4) those individuals chose the SS option in a direct choice with the LL option

Number of choices of SS	$Val_{dif} > £25$		$£25 \geq Val_{dif} \geq -£25$		$Val_{dif} < -£25$		Row totals for each method
	DV	BC	DV	BC	DV	BC	
4	12	25	3	8	38	20	53
3	5	5	3	3	9	9	17
2	2	3	1	3	9	6	12
1	1	1	0	1	10	9	11
0	0	0	1	2	8	7	9
Column totals	20	34	8	17	74	51	102

either by undervaluing SS or else to both overvaluing LL and undervaluing SS, while only 15% were due to intransitivity. By contrast, we appear to find considerably higher rates of intransitivity, even under the more demanding criteria of Table 3.

Moreover, on the basis of the aggregate data displayed in Fig. 4, it appears that DV not only tends to overvalue LL but also tends to overvalue SS relative to BC. When we examine the relationship between DV and BC on a within-person basis, this tendency is strongly confirmed: for SS, 84 of the 102 individuals have a strictly higher DV-based value; for LL, that is the case for 94 of the 102 individuals. So, even when we allow for stochasticity by basing the analysis on medians and SPVs, we find clear evidence that the DV method overvalues *both* SS and LL relative to the binary choice benchmark. However, it does so to a markedly greater degree for LL than for SS: from Table 1 we can see that the sample mean differences are £365.70 for SS and £608.10 for LL. Using the medians from that table, the differences are £150 for SS and £550 for LL.

So our data appear to attribute the causes of preference reversals rather differently than the diagnostic procedure used by TSK. In Sect. 5, we will analyse why there is such a difference in attributions. Before doing so, however, we summarise the main results from the choice list component in the design in the following section.

4 Choice list results

As noted in the introductory section, choice lists – also referred to as multiple price lists – have gained some traction among experimenters interested in a compact way of trying to identify individuals' indifference points. We were aware of a review comparing BC with CL in the context of several incentivised studies of risky decisions (Loomes et al., 2019) which suggested that the two instruments showed high degrees of positive correlation while at the same time exhibiting a particular systematic discrepancy that appears to be very much in line with the range-frequency effect identified by Parducci (1965).

To be specific, Loomes et al. (2019) found that when the various binary choices between a particular risky lottery and a range of sure amounts were presented in an ordered list, the sure amounts at the top of the list tended to be chosen more often than when the options were presented in separate binary choices, whereas the sure amounts at the bottom of the lists were liable to be chosen less often than in the corresponding stand-alone binary choices. In the middle of the ranges these effects were often absent and there was little difference between CL and BC responses.

We find much the same relationship between the CL and BC distributions for our intertemporal choices as for those risky choice studies. In particular, at higher levels of present cash sums, those amounts are chosen more often in the CL task than in the BC task. In the case of the SS option, for example, the cash sum of £1400 is chosen 432 times in CL compared with 396 times in BC. At lower cash sums, the opposite is the case: against SS, £400 is only chosen 45 times in CL compared with 106 times in BC. The data for LL show a broadly similar pattern: £1400 is chosen 407 times in CL as compared with 346 times in BC, whereas £400 is chosen 57 times in CL but 113 times in BC.

However, over the middle range of amounts (£800–£1000) the differences in the choice splits between CL and BC for each option are small and at a present sum of £900 the two more or less coincide. The sample mean SPVs derived from the CL data are not significantly different from the BC-based sample means: £885.30 compared with £878.90 for SS; and £911.00 compared with £891.90 for LL.

Figure 7 shows, for both SS and LL, plots of the 102 individual pairings of CL-based and BC-based SPVs, together with linear regression lines fitted to both charts, indicative of the high degree of positive correlation between CL

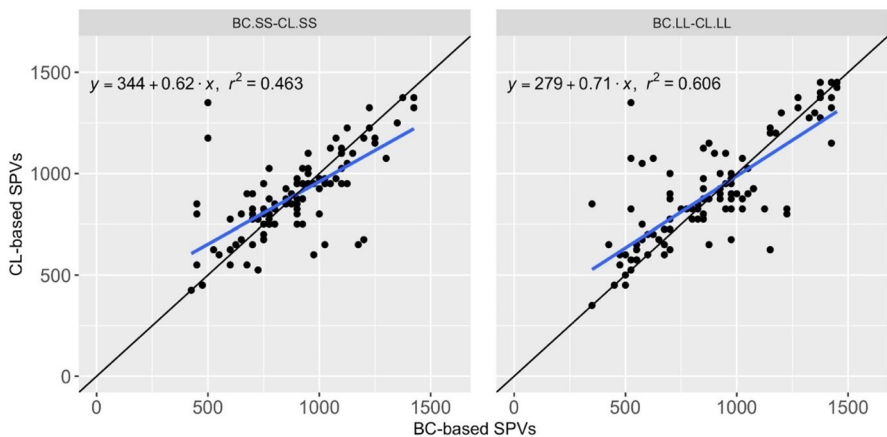


Fig. 7 Individual Choice List (CL) based and Binary Choice (BC) based Stochastic Present Values (SPVs). Each point plots an individual's CL-based SPV against his/her BC-based SPV. The left-hand panel shows those plots for the SmallerSooner (SS) option and the right-hand panel does so for the LargerLater (LL) option. For both SS and LL, the fitted regression line has a positive intercept and a slope significantly less than the 45° line

and BC. Consistent with the Parducci range-frequency effect, the future options appear more highly valued by the CL method when the present cash amounts occupy the lowest ranks in the list, but have relatively smaller SPVs than their BC counterparts in the vicinity of the top-ranked cash amounts. In both cases, the intercepts of the fitted lines are significantly greater than 0 and the slopes are significantly less than 1 ($p < 0.01$ in all four tests).

Given those results, it is not surprising to find that the preference reversal patterns are no less strong for CL as for BC. Whereas the BC-based asymmetry in Table 2 is 38:2, the CL counterpart is 40:2. Range-frequency effects may be at play when choice lists are used instead of binary choices, but the PR phenomenon is robust for both CL and BC. Moreover, the attribution of causes is much the same: DV overvalues both future options compared with CL, but does so to a greater extent for LL than for SS; and violations of weak stochastic transitivity are no less substantial.

5 TSK's diagnostic method

To understand the reasons for the differences in the attributions of cause between TSK's study and ours, we need to review the key features of the TSK design – and, in particular, we need to examine their diagnostic system more closely.

In TSK's time preference study, participants were asked to state an immediate cash value for each of five future options and to make binary choices between four pairings of those options and between each option and some predetermined present amount which TSK denoted by X . Each of these questions was asked just once. The stated present cash value equivalents were denoted by C_S for the option in any pair which paid out sooner, while the stated present value for the option which paid out later was denoted by C_L .

The four pairings together with their respective X s constituted four triples (see TSK's Table 4). TSK's diagnostic test was conducted on the pooled data given in their Appendix 2, which we reproduce as Fig. 8. The only difference between TSK's Appendix 2 and our Fig. 8 is that we have shaded the twelve cells that constitute standard preference reversals. These are the cases where the columns numbered 1, 3 and 5 (for which $C_L > C_S$) intersect with the rows numbered 1, 3, 6 and 8 (for which S is preferred to L in a straight choice between the two). In total, there are 296 such reversals.

TSK's system of attributing the causes of PR operated as follows. They focused exclusively on reversals where their predetermined value of X happened to fall between the participant's stated C_L and C_S – that is, column 1. Thus their diagnosis was based on a somewhat arbitrary⁸ subset of 104 reversals (just 35% of the total number of reversals).

⁸ It was arbitrary in the sense that the subset was determined by the value of X that the experimenters had preselected: had they chosen some other value of X , a different subset of reversals would have served as the basis for the procedure.

	1	2	3	4	5	6	Total
	$C_L > X > C_S$	$C_S > X > C_L$	$C_L > C_S > X$	$C_S > C_L > X$	$X > C_L > C_S$	$X > C_S > C_L$	
1. $X > S, S > L, X > L$	57	14	63	5	32	65	236
2. $S > X, L > S, L > X$	19	0	59	1	4	1	84
3. $X > S, S > L, L > X$	16	1	10	1	4	1	33
4. $S > X, L > S, X > L$	1	2	7	0	4	4	18
5. $X > S, L > S, X > L$	14	3	9	0	2	3	31
6. $S > X, S > L, L > X$	12	2	24	3	5	1	47
7. $X > S, L > S, L > X$	9	1	12	1	2	2	27
8. $S > X, S > L, X > L$	19	11	42	8	12	18	110
Total	147	34	226	19	65	95	586

Fig. 8 Reproduction of TSK's Appendix 2, Highlighting All Cases of Preference Reversal. The rows show the eight different combinations of binary choices between sure amount (X), sooner-paying option (S) and later-paying option (L). The columns numbered 1 – 6 show the possible orderings over X and the stated cash equivalent values of the sooner- and later-paying options, those cash equivalents being denoted by C_S and C_L respectively. The shaded cells identify cases where a preference for S over L in the straight choice between the two coincided with placing a higher cash value on L than on S and thereby constituted a preference reversal

Within that subset, they reasoned as follows. $C_L > X > C_S$ was compatible with choices $L > X$ and $X > S$. When taken in conjunction with $S > L$ in the straight choice between the two future options, this constitutes a cycle. This is the combination shown in row 3: so those 16 cases were attributed to intransitivity. In row 1, the choice between X and S was in line with the inequality between X and C_S , but $C_L > X$ ran counter to X being chosen over L, so these 57 reversals were ascribed to L being overvalued relative to choice. In row 6, choice and value were in harmony for L and X, but here there was a disparity whereby S was chosen over X but $X > C_S$: so those 12 observations were diagnosed as undervaluing S relative to choice. Finally, in row 8, choice and valuation were in conflict for both $\{X, L\}$ and $\{S, X\}$, so that those 19 reversals were judged to represent a combination of overvaluing L and undervaluing S.

Notice, however, that this system not only excludes two-thirds of all the reversals by ignoring columns 3 and 5 but also, in the absence of a choice cycle, makes it impossible for anything other than overvaluation of L or undervaluation of S to be identified. The sort of cases which appear to be typical of the data in our experiment – that is, where both LL and SS are overvalued but with LL being overvalued to a greater extent – simply cannot be detected by TSK's method.

To give a numerical illustration, consider an individual who chooses SS over LL most or all of the times, who has a BC-based SPV of £925 for LL and has a BC-based SPV of £875 for SS. Such an individual would be located in one of the upper-right BC cells in Tables 2 and 3.

However, suppose that this individual, like the majority of respondents, overvalues both LL and SS when present values are elicited via DV, giving (let us say) DV-based values of £1250 for LL and £1150 for SS. For such values, the TSK system restricts attention to cases where X lies between them: in this example, that would be when $X = £1200$. Given the individual's BC-based SPVs of £925 and £875, we can expect most if not all of his/her binary choices to favour a present amount of £1200 over both LL and SS. So the majority of BCs when X is £1200 would be $X > LL$ and $X > SS$ which, in conjunction with $SS > LL$, gives the TSK row 1 pattern that they interpret as only overvaluing LL. Thus, in cases where direct value elicitation actually overvalues both options relative to choice but overvalues LL to a greater degree than SS, there are liable to be reversals that the TSK system attributes solely to overvaluing LL, while it misses cases of intransitivity that occur at lower levels of X (such as at a present value of £900, which would generate intransitivity in our example).

To summarise, the diagnostic method proposed by TSK is incomplete and potentially misleading because: (i) it is liable to exclude many – in Fig. 8, a majority – of reversals; (ii) it excludes the possibility of diagnosing *any* overvaluation of SS, even though there are good theoretical reasons to allow for that possibility and there is strong evidence that it actually happens; and (iii) in cases where SS is overvalued, the method excludes lower levels of X where choice cycles may occur and is therefore liable to miss cases of intransitivity such as the one in our example.

6 Concluding remarks: Implications for theory and practice

It has been known for a very long time – since Mosteller and Nogee (1951) at least – that decision making under risk is probabilistic rather than deterministic, and there is now a considerable body of evidence that individuals display variability and imprecision in their choices between, and valuations of, risky prospects. Very much less evidence of that kind has been collected in the domain of intertemporal choice and valuation. By repeating each task four times, we were able to demonstrate that most individuals display considerable variability and imprecision – indeed, rather more than Blavatskyy and Maafi (2018) were able to observe by presenting each choice just twice. In that respect, our findings reinforce the message in Blavatskyy and Maafi (2018) that attention needs to be paid to allowing for ‘noise’ in the data and to trying to find appropriate stochastic specifications when fitting models and estimating parameters. But there is another important implication of imprecision: namely, that the existence of uncertainty in many individuals’ stated preferences may make their responses susceptible to procedural effects of one kind or another. Our study has provided evidence of various such effects, briefly summarised as follows.

We found that individuals’ choice-based SPVs, (whether obtained via BCs or via CLs) favoured the LL option more relative to the SS option than straight choices

between the two. Although the row totals in Table 2 show that 70 (53 + 17) respondents chose SS strictly more often than LL while only 20 (11 + 9) chose LL more often, the column totals show that just 37 had a higher BC-based SPV for SS while 62 had a higher SPV for LL. This represents a much more substantial degree of intransitive choice than the diagnosis offered by TSK and shows that – in the intertemporal context at least – present values of options inferred from repeated choices do not necessarily accord with preferences expressed through direct comparisons between those same options. One possible implication for those who use dichotomous choice contingent valuation methods to generate inputs into health, safety and environmental cost–benefit analyses is that the results of such analyses may not correspond with orderings over competing projects if elicited through direct choice or ranking.

We also found that our DV instrument – even though it was presented to respondents in among BC and CL questions and even though it tried to make the binary preference implications explicit – produced higher SPVs for both SS and LL than did either BCs or CLs. It was as if those who opted for the present sum in BC or CL tasks, but only ‘probably’ preferred it, were liable to require higher values in the DV task. The fact that the degree of overvaluation relative to BCs/CLs was greater for LL than for SS is consistent with the idea that this form of task prompts the response-generation thought process to give greater weight to the money dimension, as suggested by Tversky et al. (1988).

This raises a further issue: when different elicitation tasks produce significantly different patterns of response, how do we judge which (if any) gives us the ‘best’ answers? Throughout this paper we have followed the TSK precedent of referring to ‘overvaluation’ as if straight choice is the preferred benchmark while valuation is more vulnerable to biases such as an ‘anchoring effect’ of the kind discussed by Kahneman et al. (1999, Sect. 6). That may be the case. But before accepting that interpretation too readily, we might consider the implications of our data for another parameter of interest to policymakers in areas such as health and the environment: namely, the appropriate discount factor to apply to costs and benefits whose timing may vary considerably and may be spread over many years.

The very considerable disparities we have found between the different methods of preference elicitation are reflected in very substantial differences between the discount factors we might infer from them. To illustrate with respect to the means reported in Table 1, a DV present value of £1500 for LL translates to an annual discount factor of 0.903 (corresponding to an annualised rate of return – ARR – of 10.8%) while the DV mean of £1244.7 for SS implies an annual discount factor of 0.846 (ARR = 18.2%). By contrast, the BC-based means entail discount factors of 0.814 (ARR = 22.9%) for LL and 0.671 (ARR = 49.1%) for SS.

Viewed from this perspective, it may be that DV produces values and discount rates more in line with much economic activity (although still, arguably, overstating ARR) whereas choice-based methods somehow encourage excessive implicit rates of discounting. On this interpretation, choice-based methods, far from being the

‘gold standard’, may *undervalue* both LL and SS by discounting them too harshly – and may do so to an even greater extent for the SS option. If this is correct, policy-makers should be cautious about placing (too) much weight on estimates of subjective discount rates inferred from choice-based stated preference studies; and perhaps they should be especially cautious about taking estimates from studies using relatively short time frames and applying these in contexts such as health and environmental policy where the timing may involve many years.

To sum up. We found extensive evidence of imprecision and variability in people’s stated intertemporal preferences, suggesting that future theoretical and empirical work in this domain (as in risky choice) should aim to take account of the ‘noise’ in people’s responses. To allow for it in the present study, we repeated all of our choice and valuation tasks four times and built our analysis of the disparities between choices and valuations around within-person measures of central tendency. Our design avoided the omissions and biases inherent in the TSK diagnostic tool of 30 years ago and came to rather different conclusions about the sources of intertemporal preference reversals. Whether direct valuation methods lead to overvaluation or whether choice-based methods lead to undervaluation in this domain is an issue which requires further research; but however that may turn out, our study suggests that the degree of intransitivity is a very much more substantial contributor to intertemporal PR than had previously been supposed on the basis of the TSK study. The disparities in values and in discount factors within and between methods was substantial, suggesting that researchers and policymakers should be wary of generalising too much from any one method applied to small numbers of questions, especially if these are asked only once. There is scope – and need – for more work on the nature and impact of imprecise preferences in intertemporal decision making and for the development of stochastic specifications of models in this domain.

Acknowledgements We are grateful to the Leverhulme Trust for support under their ‘Value’ programme, grant reference RP2012-V-022; Graham Loomes also acknowledges the support of the UK Economic and Social Research Council via its Network for Integrated Behavioural Science, grant reference ES/K002201/1.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix: Instructions with practice examples

Choice between two options

In some of the questions that follow, you will be selecting between two options involving different money amounts received at different times. For example, the choice could be between the two options presented below. As in this example, each choice involves either getting a smaller amount of money sooner, or else waiting longer and getting a larger amount. We simply want to know how YOU personally would balance the extra waiting time against the extra money.

For each pair of options, you can give one of four possible answers, as shown in the example below. If you feel SURE about which option you prefer, please click on 'Definitely' just below that option. If you feel that the two options are quite evenly balanced so that you are not completely sure, please use 'Probably' to indicate which option you think you are more likely to choose.

A. Getting £1000 today

B. Getting £1500 in 2 years

☐ Definitely A

☐ Probably A

☐ Probably B

☐ Definitely B

You can change your answer by clicking on a different circle. Once you have made your final decision you will be asked to click on NEXT to go to the next question.

Fig. A1 Screenshot of instructions for BC questions

Choices in a table

In some other cases, you will see a number of choices listed in one table as below. Each row in the table represents one choice. Option B is the same throughout, while Option A changes from one row to the next. Please click on one circle in each row to show how you feel about each choice in turn.

Row no.	Decision		Option A	Option B	Decision	
1	<input type="radio"/> Definitely A	<input type="radio"/> Probably A	£1600 today	£1500 in 1.5 years	<input type="radio"/> Probably B	<input type="radio"/> Definitely B
2	<input type="radio"/> Definitely A	<input type="radio"/> Probably A	£1500 today		<input type="radio"/> Probably B	<input type="radio"/> Definitely B
3	<input type="radio"/> Definitely A	<input type="radio"/> Probably A	£1400 today		<input type="radio"/> Probably B	<input type="radio"/> Definitely B
4	<input type="radio"/> Definitely A	<input type="radio"/> Probably A	£1300 today		<input type="radio"/> Probably B	<input type="radio"/> Definitely B
5	<input type="radio"/> Definitely A	<input type="radio"/> Probably A	£1200 today		<input type="radio"/> Probably B	<input type="radio"/> Definitely B
6	<input type="radio"/> Definitely A	<input type="radio"/> Probably A	£1100 today		<input type="radio"/> Probably B	<input type="radio"/> Definitely B
7	<input type="radio"/> Definitely A	<input type="radio"/> Probably A	£1000 today		<input type="radio"/> Probably B	<input type="radio"/> Definitely B
8	<input type="radio"/> Definitely A	<input type="radio"/> Probably A	£900 today		<input type="radio"/> Probably B	<input type="radio"/> Definitely B
9	<input type="radio"/> Definitely A	<input type="radio"/> Probably A	£800 today		<input type="radio"/> Probably B	<input type="radio"/> Definitely B

You can change your answer in any row by clicking on a different circle. Answers in other rows may change in order to keep your overall set of responses consistent. Once you are happy with your set of answers, you will be asked to click on NEXT to go to the next question.

Fig. A2 Screenshot of instructions for CL questions

Choice by a slider

In some other questions, you will be asked to consider an amount you would get some time in the future and to say, within a hundred pound range, what amount of money would be just as good as that if you were to get it today. For example, imagine you own a bond that will mature in 2 years time and will then pay you £1,500. If you could exchange that for a sum of money you would get today, what amount today would you consider EQUALLY desirable TO YOU PERSONALLY as £1,500 in 2 years time?

Please use the slider below to indicate, within a hundred pound range, what sum of money today you would consider just as good as getting £1,500 in 2 years.

I think that getting £ ~ £ today is equally as good as waiting 2 years to get £1500.

Just to check that this is how you feel, please look at the two statements below. If you agree with both of them, you have set the amount above at the right level for you. If you disagree with either of the statements below, you should go back to the slider and adjust the amounts until you agree with all of the statements.

I would rather get £1,500 in 2 years than get any amount lower than £ today.

I would rather get any amount higher than £ today than wait 2 years to get £1,500.

As you move the slider, you change the amounts shown in the boxes. Please move the slider until the amounts in the boxes best express how you feel. Once you have settled on a decision, clicking NEXT will take you to the next question.

The types of question described above are all about different amounts of money that might be received at different TIMES. Interspersed with those questions are other questions about different CHANCES of gaining or losing different amounts of money.

Fig. A3 Screenshot of instructions for DV questions

References

- Becker, G. M., DeGroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Systems Research and Behavioral Science*, 9(3), 226–232.
- Blavatskyy, P. R., & Maafi, H. (2018). Estimating representations of time preferences and models of probabilistic intertemporal choice on experimental data. *Journal of Risk and Uncertainty*, 56(3), 259–287.
- Butler, D. J., & Loomes, G. C. (2007). Imprecision as an account of the preference reversal phenomenon. *American Economic Review*, 97(1), 277–297.
- Cheung, S. (2015). Risk preferences are not time preferences: On the elicitation of time preference under conditions of risk: Comment. *American Economic Review*, 105(7), 2242–2260.
- Freeman, D., Manzini, P., Mariotti, M., & Mittone, L. (2016). Procedures for eliciting time preferences. *Journal of Economic Behavior and Organization*, 126, 235–242.
- Kahneman, D., Ritov, I., & Schkade, D. (1999). Economic preferences or attitude expressions? An analysis of dollar responses to public issues. *Journal of Risk and Uncertainty*, 19, 203–235.
- Laury, S., McInnes, M., & Swarthout, J. (2012). Avoiding the curves: Direct elicitation of time preferences. *Journal of Risk and Uncertainty*, 44(3), 181–217.
- Lichtenstein, S., & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology*, 89(1), 46–55.
- Lindman, H. R. (1971). Inconsistent preferences among gambles. *Journal of Experimental Psychology*, 89(2), 390–397.
- Loomes, G., Maafi, H., & Pogrebna, G. (2019). *Multiple price lists and the elicitation of risk attitudes*. Mimeo, available from authors on request.
- Loomes, G., & Pogrebna, G. (2017). Do preference reversals disappear when we allow for probabilistic choice? *Management Science*, 63(1), 166–184.
- Loomes, G., & Sugden, R. (1983). A rationale for preference reversal. *American Economic Review*, 73(3), 428–432.
- MacCrimmon, K., & Smith, M. (1986). Imprecise equivalences: Preference reversals in money and probability. *University of British Columbia working paper*, 1211.
- Mosteller, F., & Nogee, P. (1951). An experimental measurement of utility. *Journal of Political Economy*, 59(5), 371–404.
- Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review*, 72(6), 407–418.
- Seidl, C. (2002). Preference reversal. *Journal of Economic Surveys*, 16(5), 621–655.
- Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review*, 95(3), 371–384.
- Tversky, A., Slovic, P., & Kahneman, D. (1990). The causes of preference reversal. *American Economic Review*, 80(1), 204–217.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.