



Why choice lists increase risk taking

David J. Freeman¹ · Guy Mayraz² 

Received: 24 September 2017 / Revised: 21 July 2018 / Accepted: 25 July 2018 /
Published online: 4 August 2018
© Economic Science Association 2018

Abstract

Choice lists with random incentives are widely used for preference elicitation. It is commonly assumed that subjects choose the same option in each question as they would have if it were the only question, but recent findings challenge this assumption. We conduct a large sample experiment varying incentives and presentation independently, and examine choices both near and away from certainty. We consistently find more risk taking when a choice between a safe prize and a risky lottery is embedded in a choice list than when it is presented on its own. This difference remains when we inform subjects of the paid choice in advance, implying that isolation fails not because of the random incentives scheme, but simply because the choice appears in a list together with others. We conjecture that subjects are uncertain about their preferences, reduce this uncertainty through considering the choices that confront them, and make cautious decisions in the interim. Other conditions and non-choice data support this interpretation. Our results open up the possibility that preferences inferred from choice lists offer a better indication of informed preferences than preferences inferred from single choices.

Keywords Choice lists · Random incentive scheme · Discovered preferences · Presentation effect

JEL Classification C91 · D03 · D81

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10683-018-9586-z>) contains supplementary material, which is available to authorized users.

✉ Guy Mayraz
g.mayraz@unimelb.edu.au

¹ Department of Economics, Simon Fraser University, Burnaby, BC V5A 1S6, Canada

² Department of Economics, University of Melbourne, Parkville, VIC 3010, Australia

1 Introduction

A single binary choice reveals only a single preference, so experiments typically include not one, but many choice problems. Unfortunately, this complicates the link between preferences and choices. Experimentalists therefore seek designs in which subjects choose the same option in each choice problem as they would have in an experiment that consisted solely of that particular choice problem—a feature known as “isolation”. A standard approach is to present choices in a list, such as the one in Fig. 1, and to incentivise them using the Random Incentive Scheme (RIS). Subjects make their selection in all the choices in the list, and one of their choices is then randomly selected for payment. Choices are interpreted on the assumption that subjects isolate.

Unfortunately, several recent studies have found significant violations of isolation (Cox et al. 2014, 2015; Harrison and Swarthout 2014; Freeman et al. (in press); Brown and Healy 2018). But while we now know that isolation can fail, we do not know why. The leading hypothesis puts the blame on the randomness that is inherent to the RIS, which creates a compound lottery over the options that subjects choose. Subjects who opt for a ‘certain’ \$1 over some chance of \$1.40, only receive the \$1 if that choice happens to be selected for payment. This makes no difference for subjects whose preferences satisfy the Independence Axiom (Holt 1986; Karni and Safra 1987), but violations of Independence are well-documented. A preference for certainty in violation of the Independence Axiom thus provides a possible explanation of isolation failure.

This argument, however, assumes that subjects think through the implications of the incentive scheme, and that they optimise accordingly. But there is evidence that people engage in “narrow bracketing”—making each decision separately from others (Tversky and Kahneman 1981; Read et al. 1999; Rabin and Weizsäcker 2009).¹ If subjects engage in “narrow bracketing” they are unlikely to think through the implications of the compound lottery that RIS creates.

Alternatively, it may be the difference in presentation between choice lists and single choices that causes isolation to fail. Previous work suggests that subjects tend to switch around the middle of a choice list (Andersen et al. 2006; Beauchamp et al. 2015), and that they treat the fixed side of a choice list as a reference point (Castillo and Eil 2014; Sprenger 2015). These or other presentation effects offer an alternative explanation of isolation failure.

We seek to disentangle the role played by incentives on the one hand and presentation on the other. We thus conduct a between-subjects experiment with conditions that vary both. Following previous research (Freeman et al. in press), we focus on choices between a certain monetary prize and the chance of a larger one. One example is choosing between \$1.00 with 100% chance and \$1.40 with 85% chance—a

¹ For example, in problem 3 in Tversky and Kahneman (1981), 84% of subjects choose a certain gain of \$240 over a 25% chance of \$1000 in their first decision, and 87% of subjects choose a 75% chance of losing \$1000 over a certain loss of \$750 in their second decision. Not a single subject makes the combined modal choice when it is offered as part of a single decision problem.

#	Option A	Option B
1.	<input type="radio"/> \$1.00 with 100% chance	<input type="radio"/> \$1.40 with 100% chance
2.	<input type="radio"/> \$1.00 with 100% chance	<input type="radio"/> \$1.40 with 95% chance
3.	<input type="radio"/> \$1.00 with 100% chance	<input type="radio"/> \$1.40 with 90% chance
4.	<input type="radio"/> \$1.00 with 100% chance	<input type="radio"/> \$1.40 with 85% chance
5.	<input type="radio"/> \$1.00 with 100% chance	<input type="radio"/> \$1.40 with 80% chance
6.	<input type="radio"/> \$1.00 with 100% chance	<input type="radio"/> \$1.40 with 75% chance
7.	<input type="radio"/> \$1.00 with 100% chance	<input type="radio"/> \$1.40 with 70% chance

Fig. 1 A choice list in descending order. This particular list was used in two conditions: *R-list Descending* and *K-list Descending*. These two conditions differed only in incentives

choice we refer to as Q85. Three conditions, known as *R*-lists, consist of choice lists with RIS incentives. Figure 1 shows one; the other two include the same choices, but in different order. Each single choice (*SC*) condition includes only one of the choices from the list. Finally, in conditions we call *K*-lists, subjects complete the same questions as in *R*-list conditions, but they *know* in advance that they would be paid on their choice in Q85. Each *K*-list condition is thus identical in presentation to the corresponding *R*-list condition, but has the same incentives as the *SC* condition in which Q85 is the only question.

We find a statistically significant and quantitatively large violation of isolation. In Q85, the proportion of risky choices is 41% in *R*-list and only 23% in *SC*. The figure for *K*-list is 35%—considerably closer to *R*-list than to *SC*. Presentation thus appears to be the primary driver of isolation failure.

Further insight is obtained by breaking down the results for *K*-list subjects by the question that they answered first. Most subjects in both *R*-list and *K*-list conditions start at the top of the list, and the *K*-list subjects who do so are about as risk seeking in their answer to Q85 as *R*-list subjects. Nearly a quarter of *K*-list subjects jump straight to Q85, however, and those subjects are about as risk averse in their answer to Q85 as *SC* subjects. Thus, even the relatively small difference in choices we observe between the *K*-list and *R*-list conditions appears to be the result of the attention drawn to the incentivised question, rather than of the difference in incentives *per se*. Figure 2 illustrates these comparisons.

While we focus primarily on choices between a certain and a risky option, we also include *Allais* variants of *R*-list and *SC*, in which winning chances are multiplied by a common factor of 0.4. Subjects in these two conditions are more likely to choose the riskier option than subjects in the corresponding main conditions, but the difference is much greater in *SC* than in *R*-list. As a result, subjects in the *SC Allais*

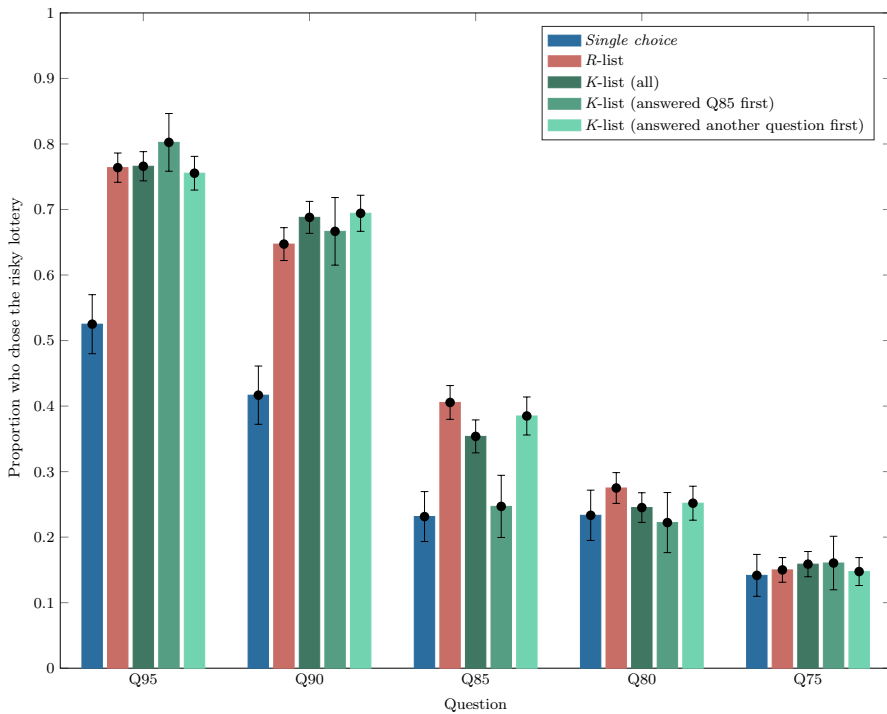


Fig. 2 Risk taking by condition. Bars show the proportion of subjects who chose the risky option. Results for *K*-list are presented both in aggregate and broken down by whether subjects answered Q85 first. Errors bars denote the standard error of the mean

condition are no more risk averse than subjects in the *R*-list *Allais* condition. Judging by this finding, isolation failure is specific to choices between a safe and a risky option.

To explain our results we need a presentation effect that generates violations of isolation between a safe and a risky option, but not between two risky options. These and other findings do not fit any of the theories we considered *ex-ante* (Sect. 3). We thus suggest a new model of presentation-dependence in the spirit of the Discovered Preferences Hypothesis (Plott 1996). Decision makers have well-defined preferences, which they gradually discover by considering the choices that they have to make. Choices are made using a cautious decision criterion that evaluates the desirability of a lottery by the lowest certainty equivalent that is consistent with the preferences that are considered ‘reasonable’ at the time of making the choice. Since list conditions include many related choices, they induce subjects to narrow down their set of ‘reasonable’ preferences, causing the certainty equivalent of risky lotteries to go up. And since the certainty equivalent of riskless prizes is fixed, this process results in an increased tendency towards risk taking in choices between a riskless prize and an uncertain lottery.

If RIS incentives were the cause of isolation failure, it would be right to see choice lists as providing a biased measure of preferences relative to the gold standard of single choice experiments. Our results suggest a more complicated and more interesting picture: there is no true gold standard for measuring preferences, but perhaps the preferences inferred from choice lists come closer to what we may think of as informed preferences.

Related Literature Ever since the discovery by Lichtenstein and Slovic (1971, without incentives) and Grether and Plott (1979, with RIS incentives) of preference reversals between matching tasks incentivised using the Becker-DeGroot-Marschack and pairwise choice tasks, economists have been broadly aware of the potential for presentation (Tversky et al. 1990) and RIS incentives (Holt 1986; Karni and Safra 1987) to affect choices in experiments. There are many existing tests of the isolation hypothesis where subjects make a small number of pairwise choices. Earlier such studies have been interpreted as support for the isolation hypothesis when subjects make pairwise choices among simple lotteries (Starmer and Sugden 1991; Beattie and Loomes 1997; Cubitt et al. 1998) and against the hypothesis that subjects integrate their portfolio (Starmer and Sugden 1991; Hey 2005).²

More recent studies provide evidence against isolation. These studies cover both the standard case where all choices in the relevant RIS condition are displayed on one screen (Cox et al. 2014, 2015, PORpi condition), and a variation in which they are displayed on separate screens with no prior information about forthcoming choices (Harrison and Swarthout 2014; Cox et al. 2015, PORnp condition). Cox et al. (2015) also establish violations of isolation for a variety of non-RIS incentive schemes, including versions of the “pay all questions” scheme. These recent papers interpret their results as due to incentives created by the RIS and related incentive schemes.

As regards widely-used choice lists, Freeman et al. (in press) show that in choices between a safe and a risky option, subjects who face an ordered list of choices incentivised by the RIS exhibit significantly less risk aversion than subjects who make only a single choice. They find less pronounced effects away from certainty. They too interpret their findings as driven by the RIS incentives. Other studies document an effect of varying the content and position of other choices in a list on inferred preferences, and interpret their findings as driven by presentation (Andersen et al. 2006; Lévy-Garboua et al. 2012; Sprenger 2015; Castillo and Eil 2014; Beauchamp et al. 2015), but do not compare choices made in lists to the behaviour of subjects who make a single choice on its own. Our study builds on previous work by separately controlling for presenting many choices on one screen, the RIS incentives, and the ordered list structure (via our scrambled condition). Another novelty in our study is our use of click data—recording not

² These papers (and ours) use a between-subjects design to infer an inconsistency in how a subject would choose in a given question across different experimental conditions that vary incentives and presentation. Other papers that use a within-subjects design find inconsistency in how the same subject chooses when responding to the same question on multiple occasions (Hey and Orme 1994; Loomes and Sugden 1998; Agranov and Ortoleva 2017).

only the choices subjects make in each line of the list, but also when they make them. This additional information sheds light on the process by which subjects make their decisions.

Our use of the *K*-list condition to study the effect of list presentation while controlling for incentives is analogous to conditions used in previous research that does not study choice lists. In particular, the ordinal payoff scheme (Tversky et al. 1990; Cubitt et al. 2004) has been used to study differences between matching vs. choice tasks while controlling for incentives. Experiments by both Starmer and Sugden (1991) and Cubitt et al. (1998) also have all subjects make all possible choices, and thus their tests of isolation are most comparable to our comparison between *K*-list and *R*-list conditions. The comparisons between the OT, impure OT, and PORpi conditions in Cox et al. (2015) are analogous to the comparisons between our *SC*, *K*-list, and *R*-list conditions. Concurrent work by Brown and Healy (2018) does study choice lists, and also uses *SC*, *K*-list, and *R*-list conditions, but does not vary question order nor consider a scrambled list, and does not consider choices with a certain option. Prior work by Lévy-Garboua et al. (2012) compares *R*-list conditions (based around the Holt and Laury (2002) design) with varying question orders (ascending, descending, and random), to conditions where each list is presented as a sequence of pairwise choices; however, they do not consider choices with a certain option.

Our results contribute more broadly to the experimental literature on choice under risk. Early experimental work documented substantial violations of the Independence Axiom and some of its generalisations (Colin 1989; Battalio et al. 1990; Wu 1994). Our results suggest that the certainty effect, and perhaps other violations of Independence, are substantially stronger in single choices than in choice lists.

2 The experiment

Subjects were randomly assigned to one of thirteen condition: six choice list conditions, five single choice conditions, and two *Allais* conditions (Table 1). Online Appendix A shows the instructions and task screen for the different conditions.

Subjects in the six main list conditions completed the same seven choices. The option on the left was always \$1.00 with 100% chance, and the option on the right was \$1.40 with chances ranging from 70 to 100% in steps of 5%. We refer to these questions as Q70, Q75, etc. One of the seven questions was played out for real. In *R*-list conditions this question was chosen randomly at the end of the experiment, whereas in *K*-list conditions it was fixed in advance to be Q85, and subjects were informed of this *before* starting the task. Both *R*-list and *K*-list conditions came in three question orders variants: *Descending*, *Ascending*, and *Scrambled* (Table 2). Figure 1 shows the list in the *Descending* question order.

There were five main single choice (*SC*) conditions: one for each of Q75, Q80, Q85, Q90, and Q95. Subjects in each of these five conditions made only a single pairwise choice that was then played out for real.

The *R*-list conditions represent the choice lists commonly used in experiments. A comparison of the proportion of risky choices between *R*-list and *SC* conditions

Table 1 Experimental conditions

Condition	Description
<i>R-list Descending</i>	Choice list with the random incentive scheme. A single question is randomly chosen for payment after subjects complete the task. The three <i>R-list</i> conditions differ in the question order (Table 2)
<i>R-list Ascending</i>	
<i>R-list Scrambled</i>	
<i>K-list Descending</i>	<i>K-list</i> conditions are identical to <i>R-list</i> conditions in presentation, but differ in incentives. The question chosen for payment is always Q85, and subjects know this before tackling the questions in the list
<i>K-list Ascending</i>	
<i>K-list Scrambled</i>	
<i>SC Q95</i>	<i>SC</i> stands for <i>Single Choice</i> . The task in these conditions consists of a single choice corresponding to one of the questions in the list conditions. There is a separate <i>SC</i> condition for each of the list questions except the two extreme ones (Q100 and Q70)
<i>SC Q90</i>	
<i>SC Q85</i>	
<i>SC Q80</i>	
<i>SC Q75</i>	
<i>R-list Allais</i>	Like <i>R-list Descending</i> , except winning probabilities on both sides are multiplied by 0.4
<i>SC Allais</i>	Like <i>SC Q85</i> , except winning probabilities on both sides are multiplied by 0.4

There were between 118 and 121 subjects in each of these 13 conditions (1560 in total)

Table 2 Choice list order variants (*R-list* and *K-list*)

#	Descending	Ascending	Scrambled
1.	Q100	Q70	Q80
2.	Q95	Q75	Q95
3.	Q90	Q80	Q70
4.	Q85	Q85	Q85
5.	Q80	Q90	Q100
6.	Q75	Q95	Q75
7.	Q70	Q100	Q90

The name of the question is the chance of winning the risky \$1.40 prize—for example, 85% in Q85. The safe option in all the questions is \$1.00 with 100% chance. Figure 1 illustrates the descending order. Note that Q85 occupies the middle position in all three orders

offers a test of the isolation hypothesis. *K-list* conditions have the same list structure as *R-list* conditions, and the same incentives as the Q85 *SC* condition. Adding *K-list* to the comparison makes it possible to identify the separate role of presentation and incentives. Comparing risk taking in Q85 between *SC* and *K-list* conditions tests for a presentation effect while controlling for incentives. Comparing risk-taking in Q85 between *K-list* and *R-list* conditions tests for an incentive effect while controlling for presentation.

The two *Allais* conditions were variants of *R-list Descending* and *SC Q85* with the winning probabilities for all options multiplied by 0.4. The fixed option in *R-list Allais* was \$1.00 with 40% chance, and the varying option was \$1.40 with evenly spaced probabilities from 28 to 40%. The question in the *SC Allais* condition was Q34: \$1.00 with 40% chance or \$1.40 with 34% chance. These two conditions

enable us to test (1) whether our subjects have a preference for certainty in violation of the Independence Axiom, and (2) whether the differences in risk taking between *R*-list and *SC* continue to hold in choices in which both options are risky.

The experiment was programmed in Javascript, and accessed via a browser. Each subject faced six screens: (1) instructions, (2) comprehension quiz, (3) task, (4) optional survey, (5) results, and (6) optional feedback. Online Appendix A provides screenshots, and explains how each screen differs between conditions. We recorded the timing of all choices and button presses, making it possible to determine the order (including possible repetitions) in which subjects completed the task, and the time they spent between decisions.

We recruited 1560 US-based subjects using the Amazon Mechanical Turk (MTurk) online labour market during US daytime hours. Payment included a \$1 participation fee and any amount won in the task. The median subject took 6 minutes to complete the experiment, earning an hourly wage of between \$10 and \$24. The participation fee was given as a ‘HIT payment’ in MTurk terminology, and the additional payment (if any) was paid as a ‘HIT bonus’. In order to minimise sampling noise, we only allowed access to US based workers with a minimum of 1000 completed tasks and a 97% approval rate, and no retaking was allowed.³

Using MTurk enabled us to recruit many more subjects than we could have with a more traditional student sample. Our statistical comparisons are binary, so a large number of subjects was required to obtain the statistical power to reject an incorrect null hypothesis. MTurk subjects are, however, more heterogeneous than student subjects, and they complete the experiment in more heterogeneous environments. These factors increase noise, and partially offset the benefits of the large sample.⁴ Subject attrition in this short and relatively well remunerated experiment was not a significant issue: only 0.7% of the subjects who started the task failed the finish.⁵

3 Hypotheses

This section describes our hypotheses. We expect to find violations of isolation in our data (Sect. 3.1), but are agnostic about the reason.⁶ Section 3.2 develops the testable implications of the two broad categories of explanation of isolation failure:

³ Similar restrictions were used in Berinsky et al. (2012) and Freeman et al. (in press) Mturk workers based in the US have to provide Amazon with their taxpayer identification information, which is then verified. This effectively ensures that each worker has only one account.

⁴ See Horton et al. (2011), Mason and Suri (2011) and Paolacci et al. (2010) for discussion of the advantages and challenges of using Amazon Mechanical Turk in science experiments, and Difallah et al. (2018) for a discussion of the MTurk subject pool.

⁵ There were 11 dropouts (4 *R*-list subjects, 3 *K*-list subjects, and 4 *SC* subjects).

⁶ In line with Freeman et al. (in press), DF thought that isolation fails because of a preference for certainty, and that we would therefore find support for Hypotheses 2, 5 and 6. GM was sceptical about subjects’ ability to think through the implications of RIS, and instead thought that it is the difference in presentation that matters, including the order of questions in the list. GM expected to find support for Hypotheses 3, 4 and 10.

incentives and presentation. Section 3.3 examines a series of specific explanations that fall under one of these two categories.

3.1 Isolation failure

Recent studies found violations of isolation in choices between a certain money amount and a risky lottery. The choices in our experiment fit this pattern, and in a pilot study comparing risk taking in Q85 we found significantly more risk taking in the choice list.⁷ We thus expect isolation to fail in our data—at least in Q85:

Hypothesis 1 (*Isolation failure in Q85*) The propensity for risk taking in Q85 differs between *SC* and at least one of the three *R*-list conditions.

In the reminder of this section we assume that this is indeed the case, and focus on identifying the reason.

3.2 Incentives or presentation?

R-list and *SC* conditions differ on two dimensions: incentives and presentation. In *R*-list conditions, subjects make a sequence of choices, and a randomly selected choice is paid. In *SC* conditions, they make only one choice, and are paid according to that choice. *K*-list conditions occupy a middle ground between the main *R*-list and *SC* conditions. A *K*-list condition with a given question order is identical in presentation to the corresponding *R*-list condition, but the incentives in Q85 are the same as in *SC* Q85 condition.⁸ The incentives and presentation explanations of isolation failure can therefore be separated by their implications for *K*-list conditions.

Suppose that isolation in Q85 fails because of the difference in incentives between *R*-list and *SC*. Since the incentives in *K*-list conditions are the same as in the *SC* Q85 condition, there should also be a difference in risk taking between otherwise identical *R*-list and *K*-list conditions:

Hypothesis 2 (*Incentives effect*) The propensity for risk taking in Q85 differs between *R*-list and *K*-list conditions that share the same question order.

Supposed, instead, that isolation fails because of the difference in presentation between *R*-list and *SC*. Since a *K*-list condition with a given question order is identical in presentation to the corresponding *R*-list condition, there should also be a difference in risk taking between *SC* and one or more of these *K*-list conditions:

⁷ Pilot sessions only included variations on *R*-list *Descending* and *SC*. 18 of 32 list subjects (56.25%) chose the risky option in Q85 as compared with 7 out of 30 single choice subjects (23.33%).

⁸ Q85 is the only incentivized question in *K*-list conditions. The other questions must be answered, but subjects have no monetary incentive to reveal their preferences.

Hypothesis 3 (*Presentation effect*) The propensity for risk taking in Q85 differs between SC and one or more of the K-list conditions.

If risk taking is affected by the difference in presentation between choice lists and single choices, it is natural to ask whether it is also affected by how the questions in the choice list are ordered:⁹

Hypothesis 4 (*Question order effect*) The propensity for risk taking in one or more questions differs between the three main R-lists conditions.

3.3 Theories of isolation failure

We now turn to more specific explanations of isolation failure, starting with theories that focus on the incentive scheme. The best known of these is based on the assumption that subjects have a preference for certainty in violation of the Independence Axiom. We can test this assumption in our subject population by comparing the level of risk taking between SC Q85 and SC Allais. The question in SC Allais is Q34—obtained from Q85 by multiplying both winning probabilities by a common factor of 0.4. If subjects have a preference for certainty, the propensity for risk taking in Q34 should be higher:

Hypothesis 5 (*Preference for certainty*) The propensity for choosing the riskier option in SC Allais is higher than in SC Q85.

A preference for certainty offers a possible explanation of isolation failure in choice lists with RIS if subjects correctly account for the risk created by their other choices (Freeman et al. in press). Since subjects in R-list conditions do not know which question would be selected for payment, they can only avoid risk entirely by choosing the certain option in all the questions. Assuming a preference for certainty, subjects in any of the main R-list conditions should therefore be more likely to choose the risky option in Q85 than either subjects in K-list conditions or subjects in SC Q85. Since a presentation effect may exist alongside an incentive effect, we control for presentation by comparing choices in R-lists and K-lists that share the same question order:¹⁰

Hypothesis 6 (*Isolation failure due to a preference for certainty*) The propensity for risk taking in Q85 is higher in each R-list condition than in the K-list condition that shares the same question order.

⁹ We limit Hypothesis 4 to R-lists, because in K-lists only one question is incentivised.

¹⁰ See Freeman et al. (in press, Sect. 4) for a more formal version of this argument, building on Karni and Safra (1987).

A more basic issue with the RIS is that each question has much less of an impact on the subject's payoff than the corresponding question in a *SC* condition. If reduced incentives increase the influence of noise, the propensity for risk taking should be biased towards 50% in *R*-list conditions, and even more so in the non-incentivised choices in *K*-list conditions:¹¹

Hypothesis 7 (*Isolation failure due to lower financial incentives in RIS*) In Q85, the propensity for risk taking in *R*-list should be between the propensity for risk taking in *SC* and 0.5, and the propensity for risk taking in *K*-list should be the same as that of *SC*. In all other questions, the propensity for risk taking in *R*-list and *K*-list conditions should be between the propensity for risk taking in *SC* and 0.5.

We now turn to theories of isolation failure that focus on the choice list presentation. The best known theory in this group is *middle bias*, which posits that subjects facing an ordered list are biased towards switching near its middle Andersen et al. (2006); Beauchamp et al. (2015). Since Q85 is in the middle of all our main list conditions, middle bias has no implications for choices in Q85 itself.¹² Subjects affected by middle bias are more likely to choose the risky option in the questions in which the risky option is better than in Q85, and are more likely to choose the safe option in the questions in which it is worse:

Hypothesis 8 (*Isolation failure due to middle-bias*) The propensity for risk taking in the ordered variants of *R*-list relative *SC* is higher in Q90 and Q95 and lower in Q80 and Q75.

The next presentation theory we consider is *static reference-dependence*. Two recent papers suggest that subjects treat the fixed-side of a choice list as a reference point, and are biased towards it (Castillo and Eil 2014; Sprenger 2015).¹³ The fixed side in our list conditions is the safe option. Static reference-dependence therefore implies *less* risk taking than in *SC* conditions:

Hypothesis 9 (*Isolation failure due to static reference-dependence*) In all the questions, the propensity for risk taking in *R*-list conditions is lower than than in *SC* conditions.

A related idea is that making several choices on the same side of the list creates a dynamic status-quo, biasing subsequent choices in the same direction.¹⁴ Such a

¹¹ For discussions of models of noisy maximisation for fitting experimental data, see Wilcox (2008), Hey et al. (2009) and Berg et al. (2010).

¹² Since other presentation effects do have testable implications for Q85, this makes it possible for us to test for their presence independently of our test for middle bias.

¹³ According to Castillo and Eil (2014), subjects see the fixed-side of the list as the status-quo option, and are affected by status-quo bias. According to Sprenger (2015), subjects see the fixed-side of the list as an endowment, and are loss averse about switching away from it.

¹⁴ Loomes and Pogrebná (2014) find evidence of such an effect.

bias would result in more risk taking in the *R*-list *Descending* condition, in which subjects are likely to choose the risky option in the first several questions, than in the *R*-list *Ascending* condition, in which they are likely to do the opposite:

Hypothesis 10 (*Isolation failure due to dynamic reference-dependence*) In all the questions, the propensity for risk taking is lower in *R*-list *Ascending* than in *R*-list *Descending*.

A final idea comes from the the psychology literature on choice and matching tasks. Tversky et al. (1990) suggests that subjects in choice problems overweigh the prominent attribute (*attribute prominence*), while in matching problems they overweigh the dimension being matched (*scale compatibility*). Subjects in our experiment are only required to solve binary choice problems, so the alleged asymmetry between choice and matching tasks is not obviously relevant. Nevertheless, it is conceivable that subjects decide on their choices in the list by first matching the fixed side of the list with its probability equivalent on the other side.¹⁵ Combined with an asymmetry between choice and matching tasks, such a choice heuristic offers a possible explanation of isolation failure. The direction of the effect is not obvious, and may depend both on the attribute that is prominent in choice tasks and on the relative strength of the attribute prominence and scale compatibility effects.¹⁶ At minimum, however, any consistent asymmetry between choice and matching tasks should bias choices in the *Allais* conditions in the same direction as in the main conditions:

Hypothesis 11 (*Isolation failure due to an asymmetry between choice and matching tasks*) The propensity for risk taking in Q34 is higher (lower) in *R*-list than in *SC* if and only if the propensity for risk taking in Q85 is higher (lower) in *R*-list than in *SC*.

4 Results

Table 3 reports the proportion of risky choices for each combination of question and condition. Figure 2 provides a graphical overview of the same data, pooling different order variants together, and focusing on the main *SC*, *R*-list, and *K*-list conditions. Tables 4, 5 and 6 report the results of statistical tests for the significance of differences in risk taking across conditions. We conducted two types of tests: (1) likelihood ratio tests that compare the proportion of risk taking between conditions and (2) probit regressions for whether a particular subject chooses the risky option with

¹⁵ For example, a subject from whom the probability equivalent of \$1 with 100% chance is \$1.40 with 82% chooses the risky option in Q100, Q95, Q90 and Q85, and the safe option in Q80, Q75 and Q70.

¹⁶ According to Tversky et al. (1988), ‘there is no obvious reason that probability is more prominent than money or vice versa.’

individual controls from the post-experiment questionnaire.¹⁷ All the results include subjects with monotonicity violations.¹⁸

4.1 Isolation failure

Contrary to the isolation hypothesis, subjects in all three *R*-list conditions were more likely to choose the risky option than subjects in *SC* conditions (Table 3). The difference in Q75 and Q80 is small, but in Q85, Q90, and Q95 it is substantial. In Q85, for example, only 23% of *SC* subjects chose the risky option, whereas in *R*-list conditions the proportion of risky choices varied from 39 to 44% with an overall average of 41%.

Table 4 reports the *p* values in likelihood ratio tests comparing the proportion of risk taking between *SC* and each of the three *R*-list conditions. The null hypothesis of no difference is rejected for all question orders in Q85, Q90, and Q95. In likelihood ratio tests combining the three question orders, the *p* value is 0.001 for Q85 and even smaller for Q90 and Q95. In a probit regression for risk taking in Q85 the *p* value on *R*-list is 0.003 (Column 1 of Table 6).

Observation 1 *In Q85, Q90, and Q95, subjects in R-list conditions were substantially more likely to choose the risky option than SC subjects. Hypothesis 1 (isolation failure in Q85) is strongly supported.*

4.2 Presentation or incentives

In Q85, the proportion of risk taking by *K*-list subjects was 35% on average—less than the 41% average for *R*-list conditions, but much higher than the 23% figure for *SC*; in Q90 and Q95 it was actually higher than in *R*-list conditions. Since *K*-lists share presentation with *R*-lists and incentives with *SC* conditions, these results are much more consistent with a presentation effect (Hypothesis 3) than with an incentive effect (Hypothesis 2).

Our formal tests focus on Q85 (the only question incentivised in *K*-list conditions), and combine the three *K*-list order variants in order to maximise statistical power. The 12% difference in risk taking between the *K*-list and *SC* is statistically significant both in a likelihood ratio test ($p = 0.013$; Table 4) and in a probit regression with individual controls ($p = 0.021$; Column 1 of Table 6). The 6% difference in risk taking between *K*-lists and *R*-lists is not statistically significant: $p = 0.15$ in a

¹⁷ The most statistically significant individual control was the self-reported tolerance for risks. Whether it improves the accuracy of our estimates depends on the extent to which it is a useful measure of a risk taking personality. Encouragingly, it did not differ meaningfully across conditions ($p = 0.21$), even when actual risk taking was significantly different.

¹⁸ As in other experiments that use choice lists, a small proportion of subjects in list conditions violated monotonicity (Online Appendix B). For example, a subject may choose the risky option in Q85 and the safe one in Q90. Some experimenters consider such subjects to be ‘noise subjects’, and drop them from their analysis. Since we can only identify such subjects in list conditions, it was not possible to drop them without introducing bias in our comparisons between list and single choice conditions.

likelihood ratio test (Table 5)¹⁹ and $p = 0.357$ in a probit regression with individual controls (Column 2 of Table 6).

Observation 2 *Subjects in K-list conditions were substantially more likely to choose the risky option in Q85 than SC subjects, and only insignificantly less so than R-list subjects. These results support Hypothesis 3 (presentation effect), but not Hypothesis 2 (incentive effect).*

4.3 The impact of question order

According to Table 3, question order does not seem to make an appreciable difference to risk taking in either R-list or K-list conditions. Differences in the proportion of risk taking are small, and are not consistent from one question to the next. Formal likelihood tests support this impression: no comparison is statistically significant after a Bonferonni correction for multiple hypotheses testing:²⁰

Observation 3 *The order of questions in a choice list has no statistically measurable impact on risk taking. Hypothesis 4 is rejected.*

4.4 Preference for certainty

The last two lines of Table 3 report the proportion of risky choices in the two *Allais* conditions. The substantial difference in risk taking between R-list and SC in Q85 disappears and even slightly reverses: 55% of SC subjects choose the risky option in Q34 vs. 51% of R-list subjects. A comparison with risk taking in Q85 reveals a strong certainty effect in SC (55 vs. 23%, $p = 5 \cdot 10^{-9}$) and a substantially weaker certainty effect in R-lists (51 vs. 41%, $p = 0.003$).²¹

Observation 4 *Consistent with Hypothesis 5 (certainty effect), risk taking in SC conditions is much higher in Q34 than in Q85. The difference in R-list is less—so much so that in Q34 R-list subjects were no more likely to choose the risky option than SC subjects.*

¹⁹ The p value in the test comparing the K-list *Scrambled* condition with the corresponding R-list condition has a marginally smaller p value of 0.13, but the conventional statistical significance threshold should be adjusted down to allow for the separate testing of the three order variants.

²⁰ In an ordered probit regression with the total number of risky choices as the dependent variable, the lowest p value is $p = 0.397$ (between R-list *Ascending* and R-list *Scrambled*). In a question specific comparison, the smallest p value is $p = 0.03$ (in Q70 between K-list *Descending* and K-list *Scrambled*).

²¹ In a contemporaneous experiment, Brown and Healy (2018) compare the impact of different mechanisms on a choice between two risky lotteries (a 50% chance of \$10 and a 50% chance of \$5 vs. a 70% chance of \$15). Like us, they find no difference in risk taking between single choices and lists with RIS incentives. They do, however, find a statistically significant ($p = 0.041$) difference between their analogues of R-lists and K-lists.

4.5 Insights from click data

Our click data enabled us to examine the order in which subjects tackle the different questions in the list, whether subjects revised their choices, and the time they spent between one question and the next. Figure 3 shows a histogram of the first question answered by *R*-list and *K*-list subjects. Fully 23% of *K*-list subjects started with the middle question—Q85—while not a single *R*-list subject did the same. Presumably, these *K*-list subjects zeroed in on the question that they knew would determine their payment, and only later paid attention to the other questions in the list.

Figure 2 and Table 3 break down the results for *K*-list subjects by whether they answered Q85 first. Those who did were substantially less likely to choose the risky option: only 25% of them did so, compared with 39% of other *K*-list subjects. According to the likelihood ratio tests in Tables 4 and 5, the group who answered Q85 first is statistically close in their choices in Q85 to *SC* subjects ($p = 0.800$) and far from *R*-list subjects ($p = 0.008$), while the remaining *K*-list subjects are close to *R*-list subjects ($p = 0.597$) and far from *SC* subjects ($p = 0.003$). The probit regressions in Q85 in Table 6 tell a similar story. With *SC* as the omitted category (Column 3), the coefficient on the group of *K*-list subjects who answered Q85 first is only 0.01, whereas the coefficient on other *K*-list subjects is 0.43—exactly the same as on *R*-list. With *R*-list as the omitted category (Column 4), the coefficient

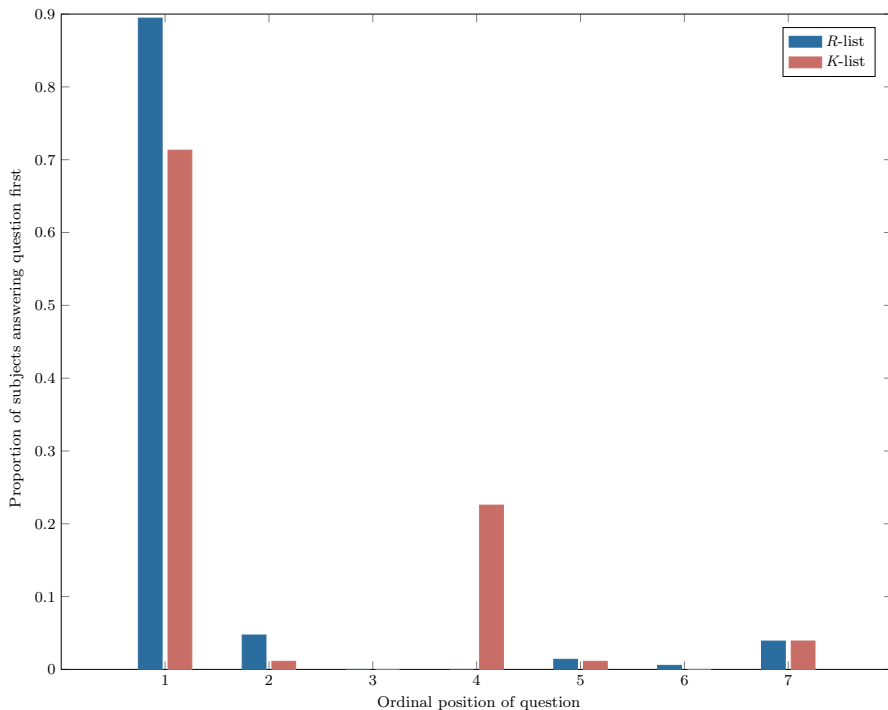


Fig. 3 The first question tackled by *R*-list and *K*-list subjects. The 4th question was Q85. Other questions varied between question orders (Table 2)

Table 3 The proportion of subjects choosing the risky option

	Question						
	Q100	Q95	Q90	Q85	Q80	Q75	Q70
<i>Single choice</i>		0.53	0.42	0.23	0.23	0.14	
<i>R-list</i>	0.94	0.76	0.65	0.41	0.28	0.16	0.11
<i>Descending</i>	0.97	0.72	0.60	0.39	0.28	0.16	0.10
<i>Ascending</i>	0.95	0.80	0.61	0.39	0.31	0.15	0.08
<i>Scrambled</i>	0.90	0.77	0.73	0.44	0.24	0.17	0.15
<i>K-list</i>	0.92	0.77	0.69	0.35	0.25	0.15	0.10
<i>Descending</i>	0.96	0.74	0.65	0.35	0.26	0.18	0.14
<i>Ascending</i>	0.90	0.78	0.67	0.37	0.26	0.15	0.11
<i>Scrambled</i>	0.90	0.78	0.74	0.34	0.22	0.12	0.06
<i>Allais conditions</i>							
<i>SC Allais</i>				0.55			
<i>R-list Allais</i>	0.98	0.78	0.70	0.51	0.34	0.21	0.12

Rows correspond to conditions and columns to questions. The *Single choice* row represents the five separate *SC* conditions. Results for *R-list* and *K-list* conditions include a breakdown by question order. In the Allais conditions the chance of winning the risky prize is scaled by a factor of 0.4 relative to the other conditions

Table 4 Risk taking in *SC* versus list conditions: *p* values in LR tests

	Q95	Q90	Q85	Q80	Q75
<i>R-list</i>	0.000	0.000	0.001	0.370	0.661
Descending order	0.002	0.004	0.008	0.397	0.738
Ascending order	0.000	0.003	0.007	0.191	0.855
Scrambled order	0.000	0.000	0.001	0.971	0.573
<i>K-list</i>	0.000	0.000	0.013	0.794	0.815
Descending order	0.000	0.000	0.047	0.680	0.397
Ascending order	0.000	0.000	0.017	0.600	0.813
Scrambled order	0.000	0.000	0.058	0.757	0.564
<i>K-list</i>	0.000	0.000	0.013	0.794	0.815
Answered Q85 first	0.000	0.001	0.800	0.854	0.713
Answered another question first	0.000	0.000	0.003	0.695	0.880

Each cell reports the *p* value in a likelihood-ratio test comparing the proportion of risky choices between the *SC* condition in the column and the corresponding question in the list condition in the row. *p* values below 0.05 are highlighted in boldface. Results for *R-list* and *K-list* conditions are separated by question order, and for *K-list* conditions also by whether subjects answered Q85 first

on the group who answered Q85 first is almost the same as the one on *SC*, and the coefficient on other *K-list* subjects is zero. While we cannot be certain that subjects who answered Q85 first were not inherently more risk averse than other subjects, there was no difference in risk taking between these two groups in any of the other

Table 5 Risk taking in *R*-list versus *K*-list in Q85: *p* values in LR tests

<i>R</i> -list	<i>K</i> -list	<i>p</i> value
All	All	0.152
Descending order	Descending order	0.505
Ascending order	Ascending order	0.766
Scrambled order	Scrambled order	0.131
All	Answered Q85 first	0.008
All	Answered another question first	0.597

Each cell reports the *p* value in a likelihood-ratio test comparing the proportion of risky choices in Q85 between *R*-list and *K*-list conditions. Separate tests are reported by question order, and for *K*-list conditions also by whether subjects answered Q85 first. *p* values below 0.05 are highlighted in boldface

Table 6 Probit regression for risk taking in Q85

	(1)	(2)	(3)	(4)
<i>SC</i>		− 0.43*** (0.15)		− 0.43*** (0.15)
<i>R</i> -list	0.43*** (0.15)		0.43*** (0.15)	
<i>K</i> -list	0.34** (0.15)	− 0.09 (0.10)		
Answered Q85 first			0.01 (0.20)	− 0.42** (0.17)
Answered another question first			0.43*** (0.15)	− 0.00 (0.11)
Risk tolerance	1.16*** (0.19)	1.16*** (0.19)	1.19*** (0.19)	1.19*** (0.19)
Male	0.23** (0.10)	0.23** (0.10)	0.23** (0.10)	0.23** (0.10)
Observations	817	817	817	817

Separate probit regressions in each column. Dependent variable: whether the subject chose the risky lottery in Q85. The omitted category is *SC* in Columns (1) and (3) and *R*-list in Columns (2) and (4). Risk tolerance is a self-reported real number between 0 and 1. The regressions also control for age, education and household income, but the coefficients are small and far from statistical significance. Standard errors in parentheses. Statistical significance indicators: **p* < 0.10; ***p* < 0.05, ****p* < 0.01

questions in the list, not was there a difference in self-reported tolerance for risk (part of the post-experiment questionnaire).

Click data also allowed us to examine cases in which a subject revised an initial choice before submitting her choices for the list. Seven of the *K*-list subjects who focused initially on Q85 switched their initial choice in Q85 from safe to risky, and

only one made the opposite switch ($p = 0.01$ in a one-sided test); among other *K*-list subjects there were just as many switches from risky to safe as from safe to risky.

Observation 5 *Risk-taking in Q85 depends on whether subjects answer Q85 first. K-list subjects who started with Q85 were about as likely to choose the safe option as SC subjects, though some later revised their choice to risky. The remaining K-list subjects were about as likely to choose the risky option as R-list subjects.*

4.6 Theories of isolation failure

Our data does not support the existence of an incentive effect (Hypothesis 2). Despite the difference in incentives, we found only a small and statistically insignificant difference in risk taking in Q85 between *R*-lists and *K*-lists. Moreover, (1) this small difference is entirely explained by the 23% of *K*-list subjects who focused on Q85 first, and (2) in other questions all *K*-list subjects were as likely to choose the risky option as *R*-list subjects. Since the difference in incentives cannot explain isolation failure in our data, nor can the specific isolation theories in Hypotheses 6 and 7.

Our data offers strong support for a presentation effect (Hypothesis 3), but none of the particular presentation effect theories we consider in Sect. 3 can explain our findings. Middle-bias (Hypothesis 8) cannot explain the excess risk taking in choice lists in Q85, and makes the counter-factual prediction of less risk taking in choice lists in Q75 and Q80. Static reference–dependence (Hypothesis 9) gets the direction wrong, predicting *less* risk taking in choice lists (including in Q85). Since we find no question order effect (Hypothesis 4), we *ipso facto* reject the specific question order effect implied by Dynamic reference–dependence (Hypothesis 10). Finally, while in Q85 the propensity for risk taking in *R*-list was 18% higher than in *SC*, in Q34 it was 4% *lower*. This runs counter to Hypothesis 11 (isolation failure due to an asymmetry between choice and matching tasks), and suggests that the mechanism responsible for isolation failure in our data is specific to choices between a safe and a risky option.

5 Preference narrowing with cautious decisions

In this section we suggest a novel model that is consistent with the entirety of our findings. While the model can account for our findings better than the presentation effect theories we considered in Sect. 3.3, it was constructed *ex-post*, and should be tested in a separate experiment. After presenting the model, we discuss a couple of non-obvious implications that can be tested in such a future experiment.

The model consists of the following three assumptions: (1) decision makers are initially uncertain about their own preferences, and consider multiple preference

relations as ‘reasonable’;²² (2) uncertainty about preferences is resolved by evaluating each lottery by its lowest certainty equivalent (as in Cerreia-Vioglio et al. 2015), as²³ (3) the process of considering choices causes the set of ‘reasonable’ preferences to shrink.

Subjects in list conditions narrow down the set of ‘reasonable’ preferences as they consider the choices in the list. Since lotteries are evaluated by their lowest certainty equivalent, preference narrowing causes the value of risky lotteries to go up, making them gradually more attractive relative to a riskless option. Consequently, preference narrowing can cause a change of preferences from the safe to the risky option, but not vice versa. Since the choices in our main conditions are between a safe and a risky option, this explains why we obtain more risk taking in *R*-list and *K*-list conditions than in *SC* conditions.

In choices between two risky lotteries, both options will generally become more attractive over time in the sense of having higher certainty equivalents, so the preference between them can change in either direction. Consider the choice between a 40% chance of \$1.00 and a 34% chance of \$1.40, and suppose a decision maker initially considers two ‘reasonable’ preference relations: one values the two lotteries at \$0.30 and \$0.40 respectively, and the other reverses these values. Since the lowest certainty equivalent of both lotteries is \$0.30, the decision maker would be initially indifferent between them. If preference narrowing eliminates the first of these two preference relations, she would come to have a strict preference for the safer lottery; if preference narrowing eliminates the second, she would come to have a strict preference for the risky one. Therefore, preference narrowing does not make a directional prediction in such choices, and is consistent with the absence of difference in risk taking between the *Allais R*-list and *SC* conditions. Online Appendix C provides a worked out example that includes both types of choices.

Formally, let Δ denote the set of simple lotteries with positive prizes. We write (x, p) to denote a lottery yielding the prize $\$x$ with probability p (and otherwise nothing). Let \mathcal{W} denote a set of utility functions over Δ , with each $U \in \mathcal{W}$ representing a preference relation that is complete, transitive, continuous, and strictly monotonic over prizes. \mathcal{W} is the set of utility functions representing preferences that the decision maker initially considers reasonable. When making a pairwise choice between lotteries in Δ , the decision maker evaluates a lottery (x, p) according to its *cautious certainty equivalent* $c_{\mathcal{W}}(x, p)$:

$$c_{\mathcal{W}}(x, p) = \min_{U \in \mathcal{W}} \{z : U(z, 1) = U(x, p)\} \quad (1)$$

²² We view this first assumption as closely related to what Butler and Loomes (2007) refer to as ‘preference imprecision’. Ok (2002) discusses the link between preference incompleteness and multiple possible preference relations.

²³ Such a tendency is captured in the cautious expected utility model of Cerreia-Vioglio et al. (2015), which assumes that each preference relation under consideration satisfies expected utility. The multiple weighting model of Dean and Ortoleva (2016) is similar in spirit, but assumes that each preference relation under consideration satisfies rank-dependent utility with a common utility-for-money function. Either model could be applied to obtain the results here.

Finally, let $\mathcal{W}(C)$ denote the set of utility functions that the decision maker considers reasonable after making a set of choices C . The preference narrowing hypothesis posits that (1) $\mathcal{W}(C) \subseteq \mathcal{W}$, and (2), for all C_1 and C_2 , $\mathcal{W}(C_1) \cap \mathcal{W}(C_2) \neq \emptyset$. Making one or more choices reduces the set of ‘reasonable’ preferences from the initial \mathcal{W} to some $\mathcal{W}' \subseteq \mathcal{W}$. It follows immediately from Equation 1 that $c_{\mathcal{W}'}(x, p) \geq c_{\mathcal{W}}(x, p)$ for any simple lottery (x, p) , with equality if $p = 1$. Making choices therefore increases the attractiveness of risky lotteries relative to safe prizes. Consider a choice between a safe and risky prize that appears in both *SC* and list conditions. Every subject who would pick the risky option in *SC*, would also pick the risky option in a list condition, but not vice versa. Given random assignment into conditions, the proportion of risky choices in the list conditions should be higher than in single choice conditions when a certain option is available (Online Appendix C).

As subjects decide on the choice in more and more questions, the preference narrowing effectively converges to a single ‘true’ preference relation.²⁴ This key feature fits in with the Discovered Preference Hypothesis (Plott 1996), and contrasts with psychological models of constructed preferences, which have no analogue of a true preference relation (Lichtenstein and Slovic 2006).

The preferences that subjects end up with depend on the set of choices that they consider, but not on the order in which they consider these choices. Subjects in the three main *R*-list conditions are presented with the same set of questions in three different orders. The fact that question order does not seem to affect choices (Observation 3) is consistent with the preference narrowing model, and further suggests that subjects consider all the choices in the list before committing to any of them.

5.1 Testable implications

The preference narrowing model explains the increased rate of risk taking in choice lists as the result of subjects learning about their preferences. The increased propensity for risk taking in these choices should therefore persist for at least some time afterwards. This can be tested in a within-subjects experiment that consists of three separate screens: (1) Q85 on its own as a single choice, (2) a choice list such as *R*-list *Descending*, and (3) a repeat of Q85 as a single choice. Subjects will be paid on one randomly selected question from the entire experiment.

The key to this design is the second screen, which offers subjects an opportunity to learn about their preferences. The prediction of the preference narrowing model is that the propensity for risk taking should increase from the first screen to the third. Importantly, a preference for consistency would result in subjects making the same choice in Q85 in all three screens, and would thus work against the prediction of the preference narrowing model.²⁵

²⁴ Since the set of preference relations is finite, the preference narrowing process must converge. Convergence to a set consisting of more than one ‘reasonable’ preference relation is possible, but the cautious decision criterion will map it to one effective preference relation.

²⁵ If we were not concerned with subjects having a preference for consistency, we could use a simpler two screen design, consisting of a choice list followed by a single choice. In this simpler design, however, the prediction of the preference narrowing model cannot be separated from that of a preference for consistency.

A second testable prediction arises from our model's use of cautious certainty equivalents. If this model is correct, the tendency for increased risk taking in the choice list should hold in all choices between a certain prize and a risky lottery, including choices in which the risky lottery offers the small chance of a much larger prize. In such choices, however, subjects are often risk seeking (the *possibility effect*), so a tendency towards risk neutrality in the list would have the opposite implication. Testing this prediction would require finding a choice question—call it Q85—that illustrates the possibility effect. Suppose that when embedded in a choice list, roughly half the subjects choose the risky option in this question. The hypothesis is that the propensity for risk taking would be less when the choice is presented on its own in a SC Q85' condition.

6 Conclusion

We compared choice lists with RIS incentives with single choices, and found considerably more risk taking in lists. Examination of choice lists with single choice incentives revealed that isolation failed not because of the RIS incentives, but because of the difference in presentation. These findings are consistent with the experimentalist intuition of Tversky and Kahneman (1981): subjects do not integrate their choices into one grand decision. Nevertheless, presentation effects cause subjects to alter their preferences as they consider the options in the list. We thus have neither integration nor isolation, but something in between the two.

Our results join a diverse group of findings that undermine the idea of stable preferences. Most important, perhaps, are preference reversals (Lichtenstein and Slovic 1971; Grether and Plott 1979), the decoy effect (Huber 1982), and the anchoring effect (Ariely et al. 2003). Our preference narrowing model, however, retains a notion of true preferences that are approached through experience. Preferences are malleable, but they are not arbitrary.

Single choices are traditionally considered the gold standard for revealing preferences, and behavioural differences with choice lists are taken to imply that choice lists are biased. This can only be so in a world of stable preferences, but in a world of malleable preferences the situation is considerably more complicated. Preference narrowing offers a guide: preferences are malleable, but people have true preferences, which they approach through experience. Single choices offer only a minimal opportunity for preference discovery, and the preferences we observe are biased towards risk aversion. Choice lists offer greater opportunity for preference discovery, and choices come closer to true preferences.

It may seem, therefore, that we should completely reverse the traditional view, and see single choices as biased. This may be the right conclusion if we are interested in true preferences, but it may or may not be appropriate if we are interested in predicting choices. Experienced decision makers have likely converged to their true preferences, and we can safely use choice lists to learn about their preferences. Inexperienced decisions makers are different. They have not had time to learn about their true preferences, and single choices may well offer a better indication of the choices they are likely to make.

Since choice lists result in different preferences, experiments can lead to different conclusions depending on the mechanism used to elicit preferences. In choices between a safe and risky option, experimenters using single choices would estimate greater risk aversion and a more powerful certainty effect. For any given sample size, experiments using single choices are more likely to reject expected utility than comparable experiments using choice lists.

Acknowledgements We thank the editor, referees, and seminar and conference audiences for comments that improved the paper. Funding for the experiment was provided by the Canadian Institute for Advanced Research (CIFAR).

References

- Agranov, M., & Ortoleva, P. (2017). Stochastic choice and preferences for randomization. *Journal of Political Economy*, 125(1), 40–68.
- Andersen, S., Harrison, G., Lau, M., & Rutström, E. (2006). Elicitation using multiple price list formats. *Experimental Economics*, 9(4), 383–405.
- Ariely, D., Loewenstein, G., & Prelec, D. (2003). Coherent arbitrariness: Stable demand curves without stable preferences. *The Quarterly Journal of Economics*, 118(1), 73–105.
- Battalio, R., Kagel, J., & Jiranyakul, K. (1990). Testing between alternative models of choice under uncertainty: Some initial results. *Journal of Risk and Uncertainty*, 3(1), 25–50.
- Beattie, J., & Loomes, G. (1997). The impact of incentives upon risky choice experiments. *Journal of Risk and Uncertainty*, 14(2), 155–168.
- Beauchamp, J. P., Benjamin, D. J., Chabris, C. F., & Laibson, D. I. (2015). Controlling for the compromise effect debiases estimates of risk preference parameters. Technical report, National Bureau of Economic Research.
- Berg, J. E., Dickhaut, J. W., & Rietz, T. A. (2010). Preference reversals: The impact of truth-revealing monetary incentives. *Games and Economic Behavior*, 68(2), 443–468.
- Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's mechanical Turk. *Political Analysis*, 20(3), 351–368.
- Brown, A. L., & Healy, P. J. (2018). Separated decisions. *European Economic Review*, 101, 20–34.
- Butler, D., & Loomes, G. (2007). Imprecision as an account of the preference reversal phenomenon. *American Economic Review*, 97(1), 277–297.
- Colin, C. F. (1989). An experimental test of several generalized utility theories. *Journal of Risk and Uncertainty*, 2(1), 61–104.
- Castillo, M., & Eil, D. (2014). Tariffing the multiple price list: Imperceptive preferences and the reversing of the common ratio effect. Working paper.
- Cerreia-Vioglio, S., Dillenberger, D., & Ortoleva, P. (2015). Cautious expected utility and the certainty effect. *Econometrica*, 83(2), 693–728.
- Cox, J. C., Sadiraj, V., & Schmidt, U. (2014). Asymmetrically dominated choice problems, the isolation hypothesis and random incentive mechanisms. *PloS ONE*, 9(3), e90742.
- Cox, J. C., Sadiraj, V., & Schmidt, U. (2015). Paradoxes and mechanisms for choice under risk. *Experimental Economics*, 18(2), 215–250.
- Cubitt, R. P., Munro, A., & Starmer, C. (2004). Testing explanations of preference reversal. *Economic Journal*, 114(497), 709–726.
- Cubitt, R. P., Starmer, C., & Sugden, R. (1998). On the validity of the random lottery incentive system. *Experimental Economics*, 1(2), 115–131.
- Dean, M., & Ortoleva, P. (2016). Allais, Ellsberg, and preferences for hedging. *Theoretical Economics*, 12(1), 317–424.
- Difallah, D., Filatova, E., & Ipeirotis, P. (2018). Demographics and dynamics of mechanical Turk workers. In *Proceedings of the eleventh ACM international conference on web search and data mining*. ACM, pp. 135–143.

- Freeman, D., Halevy, Y., & Kneeland, T. (in press). Eliciting risk preferences using choice lists. *Quantitative Economics*. <https://www.econometricsociety.org/publications/econometrica/about/journal-news/2018/05/01/new-papers-posted-quantitative-economics>.
- Grether, D., & Plott, C. (1979). Economic theory of choice and the preference reversal phenomenon. *American Economic Review*, 69(4), 623–638.
- Harrison, G., & Swarthout, J. (2014). Experimental payment protocols and the bipolar behaviorist. *Theory and Decision*, 77(3), 423–438.
- Hey, J. D., & Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica*, 62, 1291–1326.
- Hey, J. D., & Lee, J. (2005). Do subjects separate (or are they sophisticated)? *Experimental Economics*, 8(3), 233–265.
- Hey, J. D., Morone, Andrea, & Schmidt, U. (2009). Noise and bias in eliciting preferences. *Journal of Risk and Uncertainty*, 39(3), 213–235.
- Holt, C. (1986). Preference reversals and the independence axiom. *American Economic Review*, 76(3), 508–515.
- Holt, C., & Laury, S. (2002). Risk aversion and incentive effects. *American Economic Review*, 92(5), 1644–1655.
- Horton, J., Rand, D., & Zeckhauser, R. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*, 14(3), 399–425.
- Huber, J., Payne, J. W., & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research*, 9(1), 90–98.
- Karni, E., & Safra, Z. (1987). “Preference reversal” and the observability of preferences by experimental methods. *Econometrica*, 55(3), 675–685.
- Lévy-Garboua, L., Maafi, H., Masclet, D., & Terracol, A. (2012). Risk aversion and framing effects. *Experimental Economics*, 15(1), 128–144.
- Lichtenstein, S., & Slovic, P. (1971). Reversal of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology*, 89(1), 46–55.
- Lichtenstein, S., & Slovic, P. (2006). *The construction of preference*. Cambridge: Cambridge University Press.
- Loomes, G., & Pogrebná, G. (2014). Measuring individual risk attitudes when preferences are imprecise. *Economic Journal*, 124(576), 569–593.
- Loomes, G., & Sugden, R. (1998). Testing different stochastic specifications of risky choice. *Economica*, 65(260), 581–598.
- Mason, W., & Suri, S. (2011). Conducting behavioral research on amazon’s mechanical turk. *Behavior Research Methods*, 44(1), 1–23.
- Ok, E. A. (2002). Utility representation of an incomplete preference relation. *Journal of Economic Theory*, 104(2), 429–449.
- Paolacci, G., Chandler, J., & Ipeirotis, P. (2010). Running experiments on amazon mechanical turk. *Judgment and Decision Making*, 5(5), 411–419.
- Plott, C. (1996). Rational individual behaviour in markets and social choice processes. In K. Arrow, E. Colombatto, M. Perlman, & C. Schmidt (Eds.), *The rational foundations of economic behaviour* (pp. 225–250). Basingstoke: Macmillan Press.
- Rabin, M., & Weizsäcker, G. (2009). Narrow bracketing and dominated choices. *American Economic Review*, 99(4), 1508–1543.
- Read, D., Loewenstein, G., & Rabin, M. (1999). Choice bracketing. *Journal of Risk and Uncertainty*, 19(1–3), 171–197.
- Sprenger, C. (2015). An endowment effect for risk: Experimental tests of stochastic reference points. *Journal of Political Economy*, 123(6), 1456–1499.
- Starmer, C., & Sugden, R. (1991). Does the random-lottery incentive system elicit true preferences? An experimental investigation. *American Economic Review*, 81(4), 971–978.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481), 453–458.
- Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review*, 95(3), 371.
- Tversky, A., Slovic, P., & Kahneman, D. (1990). The causes of preference reversal. *American Economic Review*, 80(1), 204–217.

- 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.), *Risk aversion in experiments* (pp. 197–292). Bingley: Emerald Group Publishing Limited.
- Wu, G. (1994). An empirical test of ordinal independence. *Journal of Risk and Uncertainty*, 9(1), 39–60.