Inferring Risk Preferences from Choices: The Impact of Context and Incentives

David Freeman

Department of Economics, Simon Fraser University

Guy Mayraz

Department of Economics, University of Melbourne

May 15, 2017

Abstract

It is well known that choice lists with random incentives are not incentive compatible if subjects have non-expected utility preferences. Here we show that subjects are much more likely to choose the risky option in a choice list even if they are incentivised only on their answer to that particular question. This implies that isolation fails for reasons that have nothing to do with incentives. One possibility—supported by non-choice data—is that subjects gradually discover their preferences as they go through the list. Far from being the gold standard for preference revelation, single choices may instead reveal an initial response that is not reflective of the choices of an experienced decision maker.

JEL Classification:
Keywords: choice lists, discovered preferences, random incentive scheme

1 Introduction

People reveal their preferences through their choices. If a person chooses an 85% chance of $1.40 over a certain $1, we conclude that she prefers the risky bet, and predict that she would do the same in the future. But when several choices are made together, inferences become considerably more complicated. One issue is the difference in incentives. As outcomes depend jointly on all the choices, the person may rationally combine them all into a single joint decision. Depending on her preferences, the resulting choices may differ from the ones she would make if she were making separate decisions in each case. The other issue is the difference in context. If her preferences are affected by context, she does not need to optimise jointly over all the choice problems for individual choices to be affected by others. Both these issues complicate the link between choices, preferences, and subsequent choices in different settings.

These issues are particularly salient in experiments on risk preferences. Many such experiments have subjects make choices in one or several choice lists, such as the one in Figure 1. Most commonly, one of the choices is randomly chosen for payment, in what is known as the Random Incentive Scheme (RIS). Even if subjects make a single joint decision, RIS does not distort their choices (it is *incentive compatible*) if their preferences satisfy the Independence Axiom (Holt, 1986; Karni & Safra, 1987).[1] In addition, it has been proposed that subjects make each choice in isolation from others. For example, Tversky and Kahneman (1981) report an experiment in which a large majority of subjects made two choices that, taken together, combine into a first-order stochastically dominated lottery. Evidently, subjects made each of the two choices as if the other did not exist.[2]

According to the Isolation Hypothesis, experiments that use choice lists with RIS incentives reveal the same preferences in each problem as would an experiment that included only that one problem. Early tests were taken

[1]Kahneman and Tversky (1979) suggest that people often ignore the first common stage in a lottery—an idea generalised and formalised by Segal's Compound Independence Axiom (Segal, 1990). For such people, RIS would be incentive compatible even if their preferences violate Independence over first-stage lotteries (what Segal calls Mixture Independence).

[2]In problem 3 in Tversky and Kahneman (1981), 84% of subjects choose a sure 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 sure 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. The original problems were hypothetical, but the authors also report a successful replication with real stakes.

to be reassuring, though some came close to statistical significance (Starmer & Sugden, 1991; Beattie & Loomes, 1997; Cubitt et al., 1998). Interest in testing isolation has recently rekindled, however, and statistically significant violations have been reported (Cox et al., 2014, 2015; Harrison & Swarthout, 2014; Freeman et al., 2016).

Since Independence is frequently violated in experiments, RIS incentives offer a compelling explanation of isolation failure—but only if subjects integrate their choices into one decision. Integrating choices is, however, cognitively demanding, and runs counter to the findings of Tversky and Kahneman (1981). In this paper we investigate empirically the reasons for isolation failure. Is it the RIS incentives or a context effect? And if context matters, what is it about the context that makes a difference? What are the implications for the inferences that we should draw from the choices we observe? And finally, are there any broader implications for our understanding of choice under risk?

We conducted a between-subjects experiment with conditions that vary both context and incentives (Table 1). Following previous research, we focused mostly on choices between a certain and a risky lottery. For example, choosing between \$1.00 with 100% chance and \$1.40 with 85% chance—a choice we refer to as Q85. In single choice (*SC*) conditions, we asked just one question; in list conditions with RIS incentives (*R*-lists), we asked seven questions (Figure 1); and finally, in conditions we call *K*-lists, subjects had to complete the same seven questions, but were told in advance that their payoff would be determined by their choice in Q85. Their incentives were therefore the same as in the *SC* condition in which Q85 was the only question.

We found an economically and statistically significant violation of isolation. In Q85, for example, 41% of *R*-list subjects chose the risky option, compared with only 23% of *SC* subjects ($p < 0.0005$). The figure for *K*-list subjects was 35%: significantly above *SC* ($p = 0.013$ in a two-sided test), but not significantly below *R*-list ($p = 0.152$). Context effects must therefore be largely responsible for the failure of isolation in our data.

But while incentives, did not have a statistically significant impact on the choices, they certainly influenced the question that subjects started with: not a single *R*-list subject started with Q85, but 23% of *K*-list subjects did (Figure 3). This difference in attention was associated with a large difference in choices: only 25% of these subjects chose the risky option in Q85—very close to the figure for *SC* ($p = 0.800$)—while the figure for other *K*-list subjects was 39%—only slightly below *R*-list subjects ($p = 0.597$). Both groups were indistinguishable from *R*-list subjects in their answers to the other six questions, which were only incentivised in *R*-list conditions (Figure 2).

We could not find any other difference between these two groups of *K*-list subjects, and thus believe that the difference in their choices was the result of starting with Q85. It appears that these subjects were initially focused entirely on Q85, and only responded to the complete list context in subsequent questions. If so, even the statistically insignificant difference in choices between *K*-list and *R*-list conditions is a function of the idiosyncratic nature of *K*-list incentives, and has nothing to do with the RIS incentive scheme. We conclude that the *RIS* incentives played no role in the behavioural difference between *R*-lists and *SC*. Isolation failed because of a context effect.

Our data offer some additional clues as to the nature of this context effect. We found a consistent tendency towards risk taking in the main lists conditions, but not in choices between two risky options. We included three order variations of the list conditions, and found no behavioural difference between them. We looked at the time subjects take to complete their choices in list conditions, and found that it decreases sharply from start to end. These features do not fit any of the context effects we considered *ex-ante* (Section 3), nor do they fit such psychological mechanisms as attribute prominence and scale compatibility (Section 5).

We suggest a new model of context-dependent preferences that is in the spirit of the Discovered Preferences Hypothesis (Plott, 1996). Decision makers have a well-defined set of preferences, which they gradually discover during the choice process. As they become more certain about their preferences, they make faster decisions, resulting in the time patterns that we find. 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. As subjects proceed down the list, the set of 'reasonable' preferences narrows down, causing the certainty equivalent of risky lotteries to go up. Since the certainty equivalent of certain lotteries is fixed, this process results in an increased tendency towards risk taking—but only in choices between a certain and an uncertain lottery.

2 The Experiment

Subjects were randomly assigned to one of thirteen conditions: six choice list conditions, five single choice conditions, and two Allais conditions (Table 1).

Subjects in all list conditions completed the same seven pairwise choice questions. 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. Figure 1b illustrates one such list condition. The six conditions varied on two dimensions: incentives and question order. In all list conditions 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, while in *K*-list conditions it was fixed in advance to be Q85. Subjects were informed of this before starting the task.[3] Question orders included best-to-worst, worst-to-best, and scrambled (Table 2).

Subjects in the five single choice (*SC*) conditions made a single pairwise choice, which was then played out for real. There was one *SC* condition for each of Q75, Q80, Q85, Q90, and Q95—all the questions in the list conditions except for the first and last questions. Figure 1a shows the choice screen for the Q85 *SC* condition.

The *R*-list conditions represent the choice lists commonly used in experiments. *K*-list conditions have the same list structure as *R*-list conditions, but with the same incentives as the Q85 *SC* condition. A comparison of the proportion of risky choices between *R*-list and *SC* conditions offers a test of the isolation hypothesis. A comparison of Q85 choices with *K*-list conditions makes it possible to identify the separate role of context and incentives.

The last two conditions were Allais variants of *R*-list and *SC*. The *R*-list Allais condition was a best-to-worst list with all winning chances scaled down by a common factor of 0.4. The fixed option was thus \$1.00 with 40% chance, and the varying option was \$1.40 with evenly spaced probabilities from 28% to 40%. The choice in the Allais *SC* condition was a scaled down version of Q85: \$1.00 with 40% chance or \$1.40 with 34% chance. These two conditions enabled us to test whether our subjects violate the Independence Axiom (by comparing choices between the Allais *SC* condition and the Q85 *SC* condition), and also whether differences between the main *R*-list and *SC* conditions continue to hold in choices in which both options are risky.

The experiment was programmed in Javascript, and was accessed via a browser. There were six screens: (i) instructions, (ii) understanding quiz, (iii) task, (iv) optional survey, (v) results, and (vi) optional feedback. Appendix A shows screenshots of these screens, and explains how they differed between conditions. Programming the experiment in Javascript enabled us to record

[3] See Appendix A for the instructions and task screens.

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.

A total of 1,560 US-based subjects were recruited 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.[4]

Our statistical comparisons are binary, so a large number of subjects was required to obtain the statistical power to reject an incorrect null hypothesis for even moderate effect sizes.[5] Using MTurk enabled us to recruit many more subjects than we could have using a more traditional student sample, resulting in a correspondingly greater statistical power. MTurk subjects are more heterogeneous than student subjects and they complete the experiment at home rather than in the lab. These factors add noise, but not to a degree that cancels the numerical advantage.[6]

3 Predictions

We divide the space of possible outcomes into four possibilities: (i) counter to expectations, we may find that the Isolation Hypothesis holds in our data, (ii) Isolation may fail due to the difference in incentives, or an *incentive effect*, (iii) Isolation may fail because of a non-incentive reason, or a *context effect* (if there is a difference in choices simply because a choice is made as part of a choice list vs. making the same choice in on its own), and finally (iv) Isolation may fail due to a combination of both effects. The experiment is

[4] Similar restrictions were used in Berinsky et al. (2012) and Freeman et al. (2016).

[5] Past work that has failed to reject the isolation hypothesis (e.g. Starmer and Sugden (1991)) has found economically large but statistically insignificant effects of RIS when subjects make a small number of pairwise choices.

[6] See Horton et al. (2011), Mason and Suri (2011), and Paolacci et al. (2010) and for discussions of the advantages and challenges of running social science experiments using Amazon Mechanical Turk.

designed to separate these possibilities, regardless of the specific reason that incentives and/or context affect choices. The testable predictions of incentives and context effects are summarised in Table 3 and detailed below, including testable predictions of particular context effects suggested by previous work.

According to the Isolation Hypothesis, subjects respond to each incentivised choice in a sequence of choices just as they would have if they had been asked to make only that particular choice. In our experiment, choices in incentivised questions should not depend on whether the question appears as part of a *R*-list, *K*-list, or *SC* condition. A secondary implication is that the time subjects take to answer questions in *R*-list conditions (in which all the questions are incentivised) may differ from question to question, but should be independent of question order.

An *incentive effect* predicts a difference in the proportion of risky choices between conditions that differ in incentives: *R*-list vs. *SC*, and *R*-list vs. *K*-list. An incentive effect cannot explain a difference in the proportion of risky choices in Q85 between *K*-list and *SC*, nor differences between order variants of the same list mechanism.

We see two possible reasons why the use of the RIS could lead to incentive effects. If subjects exhibit a certainty effect, we conjecture that they will be more willing to take risk the risky options in the *R*-list than in either *K*-list or *SC* conditions.[7] A very different issue with RIS is that individual choices have less impact on the subject's payoff than the one choice in an *SC* condition. If reduced incentives increase the influence of noise, there should be a bias in towards a 0.5 frequency of risky choices in *R*-list, and an even stronger bias towards 0.5 in the non-incentivised choices in *K*-list.[8]

We use the term *context effect* to refer to any difference in subjects' propensity to take the risky option that results from the context in which a question is embedded, as opposed to any difference in incentives. Since *R*-lists and *SC* differ in both context and incentives, a difference in the proportion of risky

[7]If subjects reduce the compound lottery formed by their choices and the randomisation device that picks which question will be played out, as in Karni and Safra, 1987, and if their preference obey the Negative Certainty Independence axiom of Cerreia-Vioglio et al., 2015 which captures preferences that exhibit a certainty effect, then a subject who would opts for the \$1 in the *SC* treatment might instead choose the risky option were it embedded in an *R*-list, but would never exhibit the opposite reversal. See Freeman et al. (2016, Section 4) for a detailed argument.

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

choices between *R*-list and *SC* is consistent with both effects. However, only a context effect would generate a difference in Q85 between *K*-list and *SC*. Some (but not all) context effects also predict a difference between different order variants of a list.

One particular form of context effect, known as *middle-bias* (Andersen et al., 2006; Beauchamp et al., 2015), posits that the structure of an ordered list biases subjects towards switching around the middle. Middle-bias would lead subjects to make more risky choices above the middle of the list (in Q90 and Q95) and fewer risky choices below the middle (in Q80 and Q75). Since Q85 is at the exact middle of the list in all three list orders, middle-bias has no implications for behaviour in Q85. Thus, if middle-bias is the only reason for a difference between *R*-list and *SC*, we should observe more risk taking in the list in Q90 and Q95, less in Q80 and Q75, and no difference in Q85.

Another type of context effect could arise due to reference dependence. Sprenger (2015) proposed that subjects treat the fixed side of a list as a reference point. If are loss averse, they should make fewer risky choices in the list than in *SC*. The same prediction follows if subjects treat the fixed-side of the list as the status-quo option (Castillo & Eil, 2014), to which they are biased. We call this idea *static reference-dependence*.

Another possibility is that making several choices on the same side of the list creates a status-quo, biasing subsequent choices in the same direction.[9] We call this *dynamic reference-dependence*. The implications depend on list order. Assuming subjects complete the list from top to bottom, they should be biased towards risk-seeking in the best-to-worst order, and towards risk-aversion in the worst-to-best order.

4 Results

All 1,560 subjects are included in the analysis.[10] The raw results for the 13 conditions are in Table 4. We start the analysis by comparing *SC*, *R*-list, and *K*-list against each other. The different order variations of the two list mechanisms are lumped together for this comparison. Figure 2 displays the proportion of risky choices for different combinations of question and choice mecha-

[9] Loomes and Pogrebna (2014) find evidence of such an effect.

[10] Experiments on risk preferences frequently drop subjects who violate monotonicity as 'noise subjects'. Since monotonicity violations are only possible in list conditions, dropping such subjects would have biased comparisons between list and single choice conditions.

nism, and Table 5 reports the p-values in likelihood-ratio tests comparing each pair of mechanisms in each question.

A comparison of *R*-list with *SC* reveals substantially more risky choices in the list (76% vs. 53% in Q95, 65% vs. 42% in Q90, and 41% vs. 23% in Q85). These three differences are all strongly statistically significant ($p < 0.001$), providing a clear rejection of the isolation hypothesis. The differences in Q80 and Q75 are in the same direction, but are smaller in magnitude and not statistically significant.

Observation 1. In Q85, Q90, and Q95, subjects in *R*-list conditions were substantially more likely to choose the risky option than *SC* subjects.

K-list conditions share structure with *R*-list conditions and incentives with the Q85 *SC* condition. The proportion of risky choices in Q85 in *K*-list was 35%, compared with 23% in *SC* and 41% in *R*-list. The 12% gap between *K*-list and *SC* is statistically significant ($p = 0.013$), but the 6% gap between *R*-list and *K*-list is not ($p = 0.152$). We thus have clear evidence for a context effect, and only a weak indication of an incentive effect.

Observation 2. Focusing on Q85, subjects in *K*-list conditions were also substantially more likely to choose the risky option than *SC* subjects, and only insignificantly less so than *R*-list subjects.

We now turn to examining the impact of list order, comparing the three order variations of *R*-list against each other, and doing the same for *K*-list. Looking at Table 4, we can see that list order makes little or no difference for choices. No pairwise comparisons are statistically significant after a Bonferonni correction for multiple hypotheses testing.

Observation 3. List order had no statistically discernable impact on choices.

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. Both groups tended to start at the top of the list, but there was a dramatic difference in the proportion of subjects who started in the middle: not a single *R*-list subject and fully 23% of *K*-list subjects. Presumably, these *K*-list were drawn to complete Q85 first because they knew it would determine their payment.

Was the initial focus on Q85 associated with a difference in choices? As is evident from Figure 2 and Table 5, K-list subjects who completed Q85 first were substantially less likely to choose the risky option. Only 25% of them did so, compared with 39% of other K-list subjects. The group that started with Q85 is statistically close to *SC* ($p = 0.800$) and far from R-list ($p = 0.008$), while the remaining K-list subjects are close to R-list ($p = 0.597$) and far from *SC* ($p = 0.003$).

Were there any other differences between these two groups of K-list subjects that could explain the difference in their choices? We found no meaningful difference in survey questions, and in particular no difference in self-reported tolerance for risk. The two groups were also just as likely to choose the risky option in all the other questions. These two observations suggest that the group that started with Q85 was less likely to choose the risky option in this question not because of a difference in risk attitudes, but because it was the first question they answered. Interestingly, the proportion of risky choices in questions other than Q85 was also about the same in R-list subjects—despite these questions being unincentivised in K-list conditions.

Observation 4. Risk-taking in Q85 depended greatly on whether subjects started with this question. K-list subjects who started with Q85 behaved indistinguishably from *SC* subjects, while other K-list subjects behaved indistinguishably from R-list subjects. Both groups were indistinguishable from R-list subjects in other questions—questions that were incentivised in R-list and purely hypothetical in K-list.

The likelihood-ratio tests compare the proportion of risky choices across groups, and ignore any differences in subject characteristics. Given the random allocation and relatively large sample size, such differences are unlikely to be consequential. Nevertheless, we wanted to be sure that subject differences do not affect our most important results. We thus conducted probit regressions of individual choices in Q85 against both condition group and subject characteristics[11] (Table 6). The most significant predictor of choices in these regressions was self-reported tolerance for risk,[12] but the R-list vs. *SC* difference had about twice the impact as gender—the second most important

[11] These were elicited in the optional survey following the main task. The vast majority of subjects completed all the questions in the survey, but there are about 3% missing observations.

[12] It is unfortunately not clear to what extent risk tolerance reflects a stable personality characteristic: It was elicited after the task, and answers could have been affected by the choices subjects made in the task.

individual regressor. Comparing regressions with and without individual controls, we find that such controls (i) reduce the coefficient on *R*-list from 0.50 to 0.43,[13] though the difference remains strongly statistically significant; (iii) increase the similarity between the group of *K*-list subjects who answered Q85 first and *SC* subjects (a coefficient of 0.01 only, with a standard error of 0.20); and (iii) increase the similarity between other *K*-list subjects and *R*-list subjects (both have the exact same 0.43 coefficient).

Observation 5. Controlling for individual subject characteristics does not materially change our results.

Figure 4 shows the median time to answering each of the different questions for subjects who completed the list in order and in one go. Times are measured from the start of the task for the first question (including time spent on reading the instructions), and from the previous question for all other questions. The results show a clear downtrend, with subjects speeding up as they go down the list. For example, the median time for answering Q75 is 3.24 seconds in the worst-to-best order, in which it is the 2nd question, but only 0.91 seconds in the best-to-worst order, in which it is the 6th question.[14]

Observation 6. The time subjects take to respond to each question is determined not by its difficulty, but by its position in the list.

The last two lines of Table 4 report the proportion of risky choices in the Allais conditions. Our only comparison is in the Allais version of Q85, and we find no significant difference in risk taking: 55% in *SC* and 51% in *R*-list.[15] Comparing the Allais and regular conditions, we find a certainty effect (violating Independence) across the board, but the difference in risk taking is much stronger in the *SC* conditions. In Q85 the difference in risk taking between the Allais and regular *R-list* conditions is 10% (51% vs. 41%) and in *SC* conditions it is 28% (55% vs. 23%).

[13] This is not a general result: The coefficient increases in some other questions.

[14] We cannot rule out the possibility that these differences are the result of subjects becoming more familiar with the interface as they go down the list, but we don't think this is likely. The experiment uses standard radio buttons (Figure 1), and subjects will have used the exact interface in the general instructions and quiz before starting the main task.

[15] In a contemporaneous experiment, Brown and Healy (2016) 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 our Allais results, 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.

Observation 7. Allais comparisons reveal a strong certainty effect in single choice conditions, and a much weaker one in list conditions. This corresponds to a combination of (i) a large difference in risk taking between *R*-list and *SC* in the vicinity of certainty, and (ii) no difference away from certainty.

Finally, we examined the proportion of subjects who violated montonicity. Subjects in *R*-list can violate monotonicity either by choosing \$1.00 in Q100 (with \$1.40 available with 100% chance) or by making multiple switches. We focus attention on *R*-list conditions, since only Q85 is incentivised in *K*-list. 4% of subjects in ordered *R*-list conditions choose \$1.00 in Q100, and 10% did so in the scrambled order *R*-list. Most of these subjects (69%) chose \$1.00 in all other questions. These rates are higher than the 1% of all decisions between lotteries with a dominance relationship in Loomes et al. (2002), but in line with the 0-10% of subjects who choose a lottery that is transparently first-order stochastically dominated in Agranov and Ortoleva (2017, Table A8). Approximately 3% of subjects in ordered *R*-lists switched back-and-forth, and 14% did so in the Scrambled *R*-list condition. These rates are relatively low. For example, Holt and Laury (2002) find that 13% of subjects switch back-and-forth in the first low-payoff choice list they face.

Observation 8. Monotonicity violations are in line with other studies.

5 Discussion

The isolation hypothesis is rejected in our data, with substantially more risk taking in *R*-list conditions than in *SC* conditions. This pattern holds across several questions with p-values below 0.001.

Since *R*-list and *SC* conditions differ in both context and incentives, both factors can potentially explain this finding. However, *K*-list conditions share incentives with the Q85 *SC* condition, and the difference in risk taking between them can only be explained by the difference in context. Since *K*-list and *R*-list conditions have the same structure, this context effect must also be at least partly responsible for the difference in risk taking between *R*-list and *SC* conditions.

Any incentive effect must manifest itself by a difference in risk taking between *R*-lists and *K*-lists. We did find a small difference, but it was statistically insignificant, and entirely explained by the 23% of *K*-list subjects who tackled Q85 first (presumably because it was the question that would determine their payment). In all other respects these subjects were indistinguishable from other *K*-list subjects. We thus believe that the reason they made different choices in Q85 is not because of any pre-existing difference in risk attitudes, but because their attentional focus on this question effectively put them into the Q85 *SC* condition. As this has no analogue in *R*-list conditions, we conclude that context effects are responsible for the entire difference in risk taking between *R*-list and *SC* conditions. The choices of *K*-list subjects in other questions offer additional support to this conclusion. Though their answers in these questions could not affect their payment, their behaviour was indistinguishable from that of *R*-list subjects.

While our findings imply that isolation fails because of a context effect, none of the context effect theories we considered in Section 3 quite work. Static reference-dependence predicts that subjects will biased towards risk aversion in list conditions—the opposite of what we observe. Dynamic reference dependence predicts a dependence of choices on list order, but we find none. Middle-bias cannot explain the excess risk taking in choice lists in Q85, and makes the counter-factual prediction of less risk taking in Q75 and Q80. Middle-bias may perhaps help explain the bigger difference in risk-taking in Q90 and Q95 than in Q75 and Q80, but only if combined with a separate explanation of the greater propensity to takes risks in choice lists.

The preference-reversal literature offer other possible explanations. Tversky et al. (1990) suggests that subjects in choice tasks focus on the prominent attribute (*attribute prominence*) and that in matching tasks they focus on the dimension being matched (*scale compatibility*). Attribute prominence offers a possible explanation of our data if we take probability to be the prominent attribute,[16] and scale compatibility applies if we equate choice lists with a matching task. But as shown by Cubitt et al. (2004), scale compatibility works in opposite directions when probabilities are being matched. Moreover, attribute prominence and scale compatibility apply equally to our Allais conditions, where we find no difference in risk taking. We thus conclude that attribute prominence and scale compatibility cannot explain our results.

We believe that our non-choice data points to a more promising explanation. Subjects speed up as they go down the list, suggesting that early choices inform subsequent decision-making. The difference in risk taking in Q85 be-

[16]Tversky and Slovic vacillated on this point. According to Tversky et al. (1988), 'there is no obvious reason that probability is more prominent than money or vice versa.'

tween K-list subjects who started with this question and other K-list subjects further suggests that the difference in choices between R-list and SC conditions was mediated by the experience of prior choices. We next present a model of decision making that uses these insights to explain our findings.

6 Preference narrowing with cautious decisions

The model has three features, which combine to explain our findings: (i) decision makers are initially uncertain about their own preferences, and consider multiple preference relations as 'reasonable';[17] (ii) uncertainty about preferences is resolved by evaluating each lottery by its lowest certainty equivalent;[18] (iii) the process of considering choices causes the set of 'reasonable' preferences to shrink.

Subjects in choice lists conditions narrow down the set of 'reasonable' preferences as they proceed down the list. Since lotteries are evaluated by their lowest certainty equivalent, this process causes risk lotteries to become gradually more attractive. This leads to increased risk seeking in choices between safe and risky lotteries, with no comparable implication for choices between two risky lotteries.

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 $\mathscr{W}$ denote a set of utility functions over Δ, with each $U \in \mathscr{W}$ representing a complete, transitive, and continuous preference relation. $\mathscr{W}$ is the of set preference relations 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* $c_{\mathscr{W}}(x, p)$:

$$c_{\mathscr{W}}(x, p) = \min_{U \in \mathscr{W}} [z : U(z, 1) = U(x, p)] \tag{1}$$

[17]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.

[18]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 $\mathscr{W}(C)$ denote the set of preference relations that the decision maker considers reasonable after making a set of choices C. The preference narrowing hypothesis posits that (i) $\mathscr{W}(C) \subseteq \mathscr{W}$, and (ii), for all C_1 and C_2, $\mathscr{W}(C_1) \cap \mathscr{W}(C_2) \neq \emptyset$.

Making one or more choices reduces the set of 'reasonable' preferences from the initial $\mathscr{W}$ to some $\mathscr{W}' \subseteq \mathscr{W}$. It follows immediately from Equation 1 that $c_{\mathscr{W}'}(x,p) \geq c_{\mathscr{W}}(x,p)$ for any simple lottery (x,p), with equality if $p = 1$. Making choices 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 (e.g. Q75-Q95). 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, we conclude that the proportion of risky choices in the list conditions should be (at least weakly) higher than in single choice conditions.

These assumptions imply that the preference narrowing process effectively converges to a single 'true' preference relation $\succsim^*$.[19] The model fits in with the Discovered Preference Hypothesis (Plott, 1996), and contrasts with psychological models of constructed preferences (Lichtenstein & Slovic, 2006), which have no analogue of a true preference relation.

7 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 context effects. These findings are consistent with the experimentalist intuition of Tversky and Kahneman (1981): subjects do not integrate their choices into one grand decision. Nevertheless, context effects cause subjects to alter their preferences as they proceed down 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 & Slovic, 1971; Grether & Plott, 1979), the decoy effect (Huber et al.,

[19] The cautious decision criterion maps each set of 'reasonable' preference relations to an effective preference relation. Thus, while the decision process need not converge to a single preference relation, it does converge to an effective preference relation.

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 subjects 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 would be the right conclusion if we are interested in true preferences, but it may or may not be correct 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 much more likely to reject expected utility than comparable experiments using choice lists.

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. *Quarterly Journal of Economics*, *118*(1), 73–105.

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*. 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. & Healy, P. (2016). *Separated decisions*. Working Paper.

Butler, D. & Loomes, G. (2007). Imprecision as an account of the preference reversal phenomenon. *American Economic Review*, *97*(1), 277–297.

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., 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.

Freeman, D., Halevy, Y., & Kneeland, T. (2016). *Eliciting risk preferences using choice lists*. Working Paper.

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., Morone, A., & 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.

Kahneman, D. & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, *47*(2), 263–291.

Karni, E. & Safra, Z. (1987). "preference reversal" and the observability of preferences by experimental methods. *Econometrica*, *55*(3), 675–685.

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 University Press.

Loomes, G., Moffatt, P. G., & Sugden, R. (2002). A microeconometric test of alternative stochastic theories of risky choice. *Journal of Risk and Uncertainty*, *24*(2), 103–130.

Loomes, G. & Pogrebna, G. (2014). Measuring individual risk attitudes when preferences are imprecise. *Economic Journal*, *124*(576), 569–593.

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). Macmillan Press.

Segal, U. (1990). Two-stage lotteries without the reduction axiom. *Econometrica*, *58*(2), 349–377.

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). Emerald Group Publishing Limited.

Table 1: Experimental conditions

Condition	Description
R-list Best-to-worst Worst-to-best Scrambled	Choice list comprising seven questions with the random incentive scheme: a single question is randomly chosen for payment after subjects complete the task. The three *R*-list conditions differ in the question order (Table 2).
K-list Best-to-worst Worst-to-best Scrambled	These conditions were identical to *R*-list in presentation, but were different in incentives: the question chosen for payment is always question 4 (Q85), and subjects know this before tackling the questions in the list.
Single choice (*SC*) Q95 Q90 Q85 Q80 Q75	The task consists of a single question corresponding to one of the questions in the list conditions. Separate conditions for each of these questions except the two extreme ones (Q100 and Q70).
Allais *R*-list *SC*	Same as (i) *R*-list with best-to-worst order, and (ii) the Q85 *SC*, except all winning probabilities are scaled by a common factor of 0.4.

[a] There were between 118 and 121 subjects in each of these 13 conditions (1,560 in total).

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

#	Best-to-worst	Worst-to-best	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

[a] 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 best-to-worst order. Note that Q85 occupies the middle position in all three orders.

Table 3: Predictions

Proportion of risky choices	Isolation hypothesis	Incentives effect	Context effect
R-list vs. SC	=	≠	≠
K-list vs. SC	=	=	≠
K-list vs. R-list	=	≠	=
Question order	=	=	?

[a] See Section 3 for details.

Table 4: Proportion of subjects choosing the risky option.

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

[a] Rows correspond to conditions and columns to questions. The *Single choice* row represents the five separate *Single choice* conditions. Results for R-*list* and K-*list* conditions are given both in combined form and separately for each list 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 5: LR-tests comparing risk taking between conditions

	Question						
	Q100	Q95	Q90	Q85	Q80	Q75	Q70
				Single choice			
R-list		**0.000**	**0.000**	**0.001**	0.370	0.661	
K-list		**0.000**	**0.000**	**0.013**	0.794	0.815	
Q85 first		**0.000**	**0.001**	0.800	0.854	0.713	
Other		**0.000**	**0.000**	**0.003**	0.695	0.880	
				R-list			
K-list	0.304	0.946	0.246	0.152	0.361	0.769	0.727
Q85 first	0.406	0.455	0.740	**0.008**	0.331	0.962	1.000
Other	0.373	0.803	0.211	0.597	0.510	0.706	0.673

[a] *p*-values in likelihood ratio tests comparing the proportion of risky choices in two conditions. The first set compares list conditions with *Single choice*. The second set compares *K*-list with *R*-list. Results for *K*-list subjects are presented both in aggregate and broken down by whether they answered the 85% question first. *p*-values below 0.05 are in boldface.

Table 6: Regression analysis of the propensity for risk taking in Q85[a]

	(1)	(2)	(3)	(4)
R-list	0.50***	0.50***	0.43**	0.43**
	(0.14)	(0.14)	(0.15)	(0.15)
K-list	0.36**		0.34**	
	(0.14)		(0.15)	
Answered Q85 first		0.05		0.01
		(0.20)		(0.20)
Answered another question first		0.44**		0.43**
		(0.15)		(0.15)
Risk tolerance			1.16***	1.19***
			(0.19)	(0.19)
Male			0.23**	0.23**
			(0.10)	(0.10)
Observations	840	840	817	817

[a] Separate probit regressions in each column. Dependent variable: whether the subject chose the risky lottery in Q85. Standard errors in parentheses. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

[b] Dummies. Omitted category: *Single choice*. K-list separated by question first answered.

[c] Variables elicited in post-experimental survey. Risk tolerance is a self-reported real number between 0 and 1. Age, education and household income were included in the regressions in columns (3) and (4), but the coefficients were very small and statistically insignificant.

a

Option A	Option B
○ $1.00 with 100% chance	○ $1.40 with 85% chance

b

#	Option A	Option B
1.	○ $1.00 with 100% chance	○ $1.40 with 100% chance
2.	○ $1.00 with 100% chance	○ $1.40 with 95% chance
3.	○ $1.00 with 100% chance	○ $1.40 with 90% chance
4.	○ $1.00 with 100% chance	○ $1.40 with 85% chance
5.	○ $1.00 with 100% chance	○ $1.40 with 80% chance
6.	○ $1.00 with 100% chance	○ $1.40 with 75% chance
7.	○ $1.00 with 100% chance	○ $1.40 with 70% chance

Figure 1: Single choices and choice lists. In a *single choice* (a) the subject has to answer a single question, which in this paper is always a choice between two lotteries. In a *choice list* (b) the subject has to answer a series of questions organised in a list. This turns out to make a difference to choices. In this example, subjects are more likely to choose the risky option in question 4 in the list than when the same question is presented as a single choice.

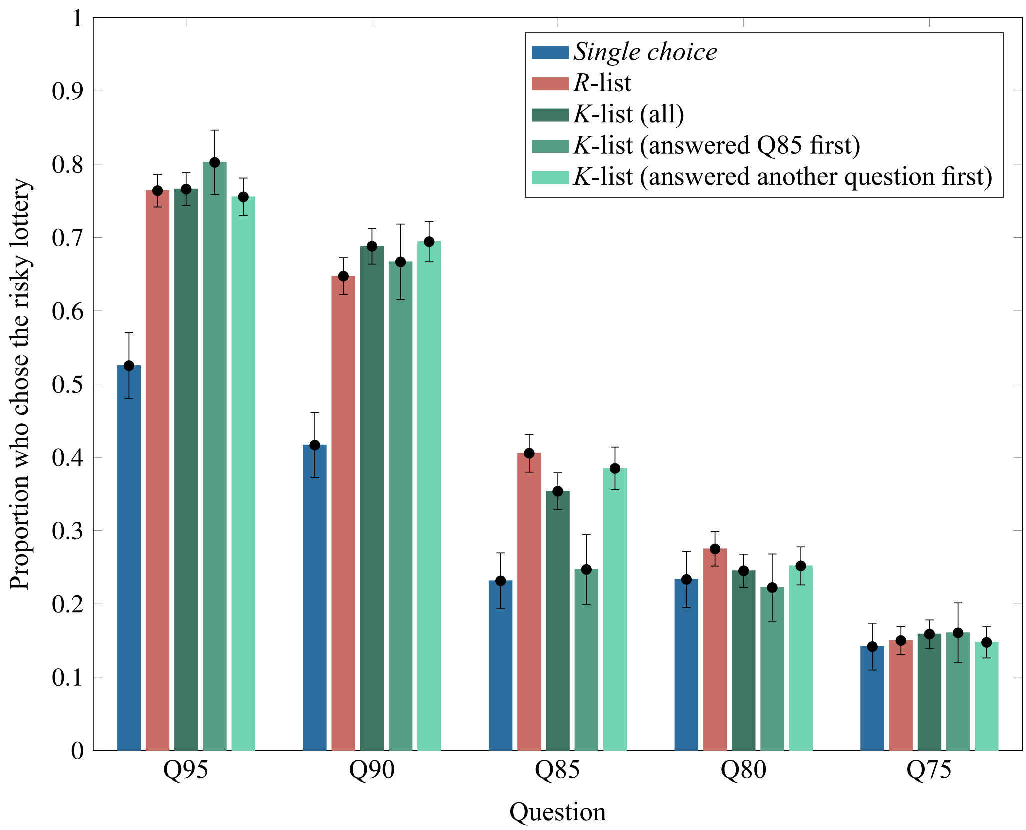


Figure 2: Risk taking by condition. Bars show the proportion of subjects who chose the risky option. Results for K-list are presented both in aggregate and broken down by whether subjects answered the 85% question first. Errors bars denote the standard error of the mean.

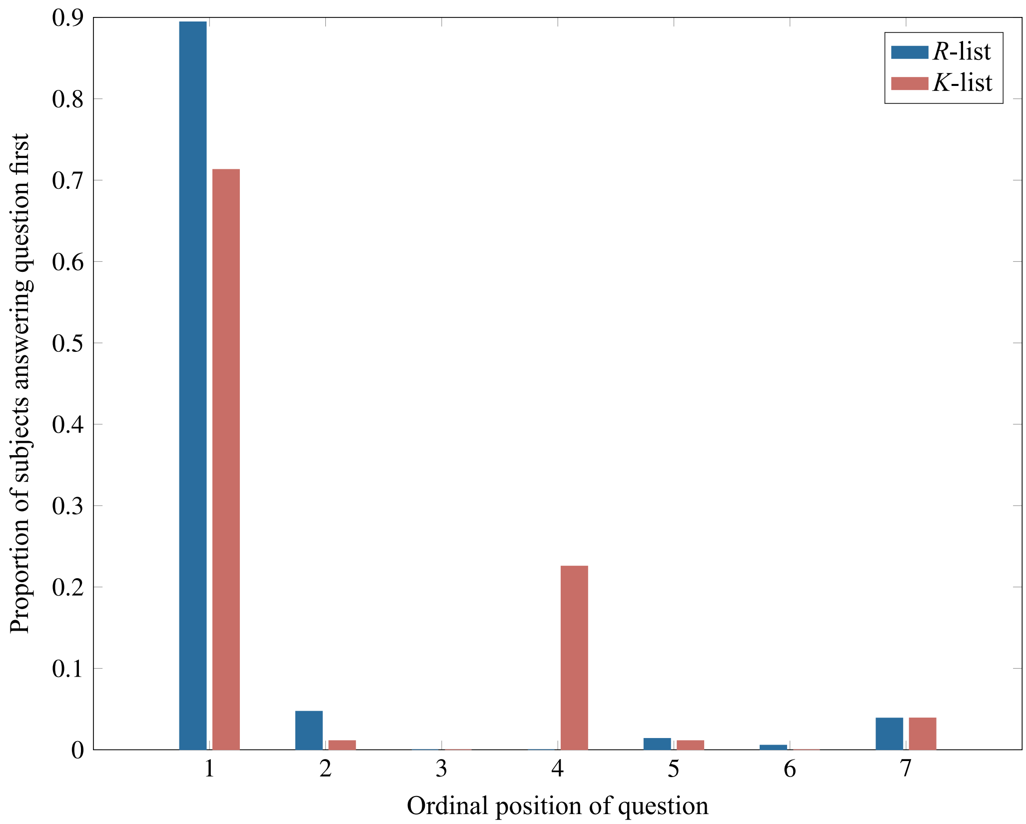


Figure 3: The first question tackled by R-list and K-list subjects.

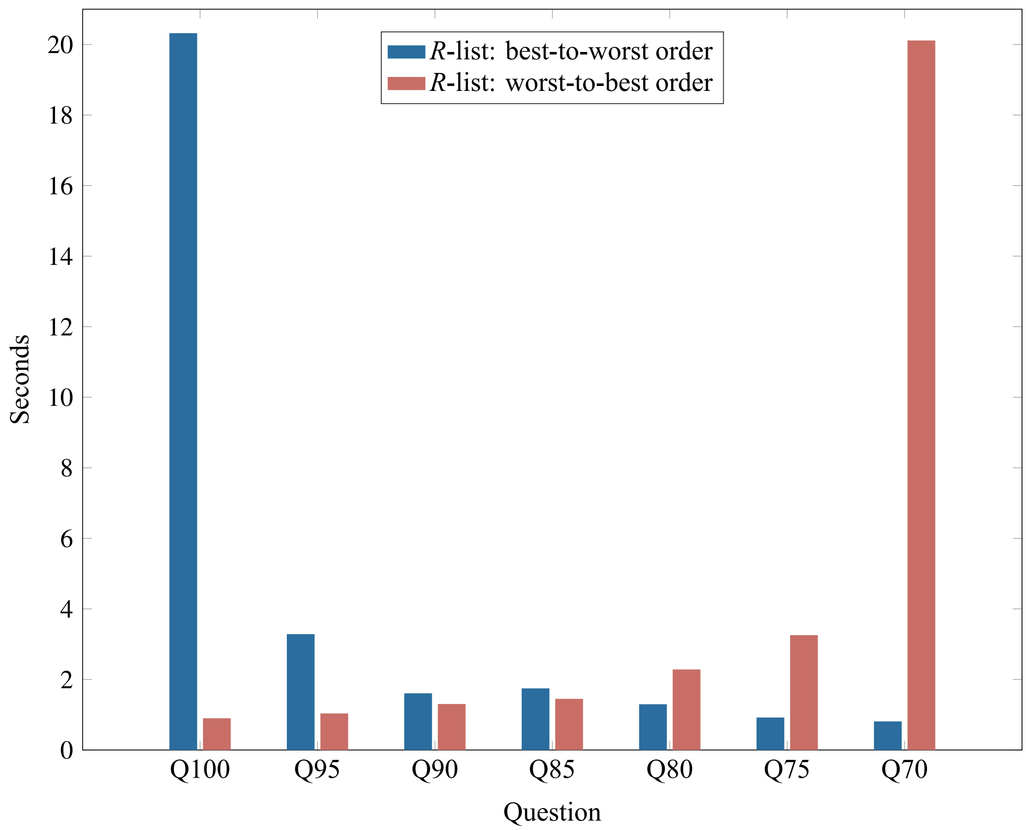
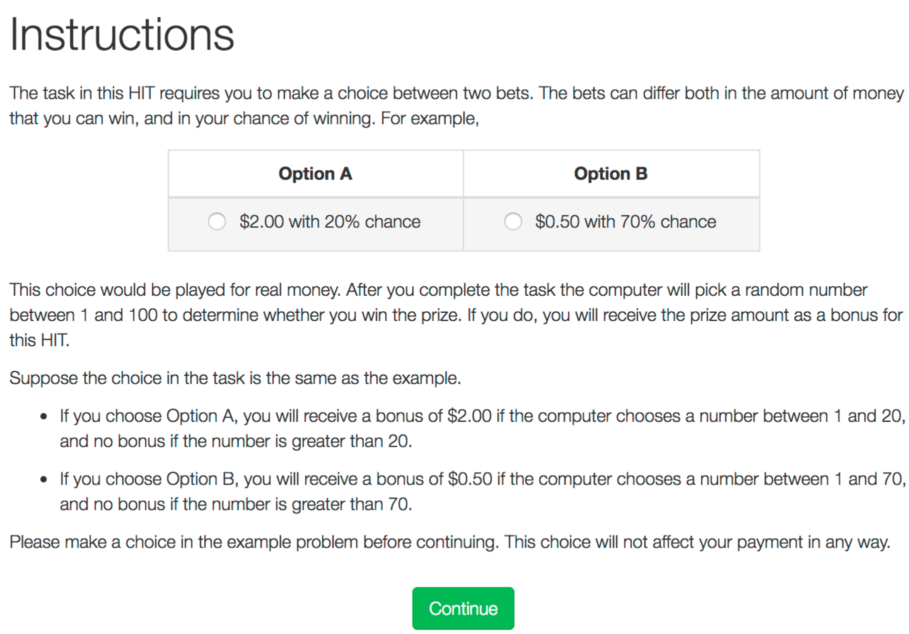


Figure 4: Median time to answer the different questions in the list. Sample limited to subjects who answered the questions once and in order.



Instructions

The task in this HIT requires you to make a choice between two bets. The bets can differ both in the amount of money that you can win, and in your chance of winning. For example,

Option A	Option B
$2.00 with 20% chance	$0.50 with 70% chance

This choice would be played for real money. After you complete the task the computer will pick a random number between 1 and 100 to determine whether you win the prize. If you do, you will receive the prize amount as a bonus for this HIT.

Suppose the choice in the task is the same as the example.

- If you choose Option A, you will receive a bonus of $2.00 if the computer chooses a number between 1 and 20, and no bonus if the number is greater than 20.
- If you choose Option B, you will receive a bonus of $0.50 if the computer chooses a number between 1 and 70, and no bonus if the number is greater than 70.

Please make a choice in the example problem before continuing. This choice will not affect your payment in any way.

Continue

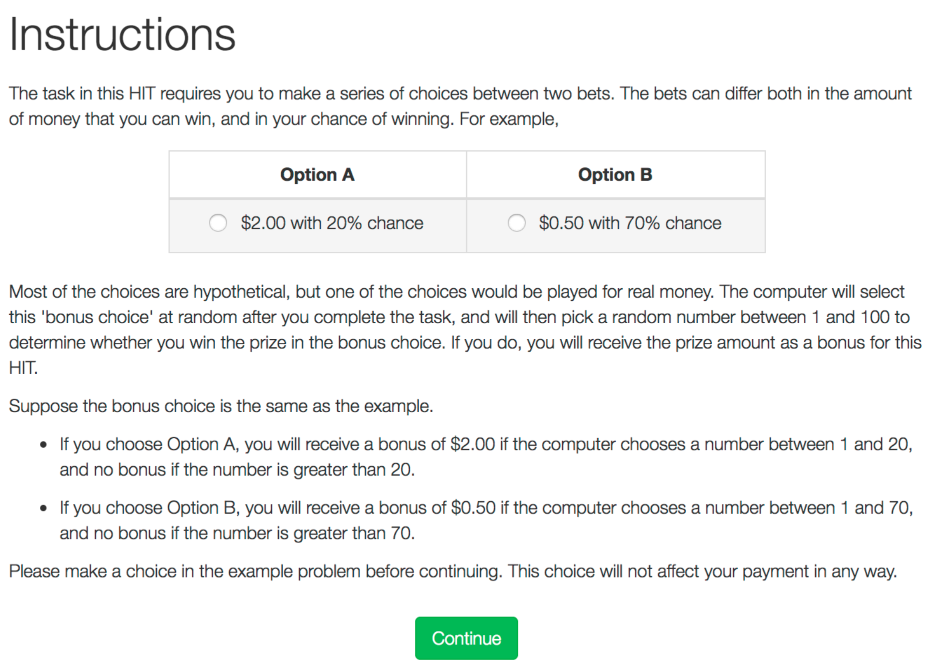
Figure A.1: Instructions in *SC* conditions (including Allais variation).

A Experiment screen shots

This appendix includes screenshots of the experiment, and explains how these differed between conditions.

A.1 General instructions

The first screen in the experiment was the instructions. These differed a little between *SC* (Figure A.1), *R*-list (Figure A.2), and *K*-list conditions (Figure A.3). The difference was (i) in the description of the task: "make a choice between two bets" or "make a series of choices between two bets", and (ii) in the description of the payment: "This choice would be played for real money"; "Most of the choices are hypothetical, but one of the choices woudl be played for real money. The computer will select this 'bonus choice' at random after you complete the task…"; or "Most of the choices…You will be told which of the choices is the 'bonus choices' when you start the task." The instructions explained the payment scheme, and gave subjects an opportunity to try the interface in a choice that had little to do with the actual choice(s) in the task.



Instructions

The task in this HIT requires you to make a series of choices between two bets. The bets can differ both in the amount of money that you can win, and in your chance of winning. For example,

Option A	Option B
◯ $2.00 with 20% chance	◯ $0.50 with 70% chance

Most of the choices are hypothetical, but one of the choices would be played for real money. The computer will select this 'bonus choice' at random after you complete the task, and will then pick a random number between 1 and 100 to determine whether you win the prize in the bonus choice. If you do, you will receive the prize amount as a bonus for this HIT.

Suppose the bonus choice is the same as the example.

- If you choose Option A, you will receive a bonus of $2.00 if the computer chooses a number between 1 and 20, and no bonus if the number is greater than 20.
- If you choose Option B, you will receive a bonus of $0.50 if the computer chooses a number between 1 and 70, and no bonus if the number is greater than 70.

Please make a choice in the example problem before continuing. This choice will not affect your payment in any way.

Continue

Figure A.2: Instructions in *R*-list conditions (including Allais variation).

Instructions

The task in this HIT requires you to make a series of choices between two bets. The bets can differ both in the amount of money that you can win, and in your chance of winning. For example,

Option A	Option B
$2.00 with 20% chance	$0.50 with 70% chance

Most of the choices are hypothetical, but one of the choices would be played for real money. You will be told which of the choices is the 'bonus choice' when you start the task. After you complete the task the computer will choose a random number between 1 and 100 to determine whether you win the prize in the bonus choice. If you do, you will receive the prize amount as a bonus for this HIT.

Suppose the bonus choice is the same as the example.

- If you choose Option A, you will receive a bonus of $2.00 if the computer chooses a number between 1 and 20, and no bonus if the number is greater than 20.
- If you choose Option B, you will receive a bonus of $0.50 if the computer chooses a number between 1 and 70, and no bonus if the number is greater than 70.

Please make a choice in the example problem before continuing. This choice will not affect your payment in any way.

Continue

Figure A.3: Instructions in K-list conditions (including Allais variation). In this screenshot the subject has chosen Option A in the example choice.

A.2 Quiz

The instructions were followed by a quiz, which subjects had to complete correctly before starting the task. The quiz was designed to ensure that subjects understand that their choices have real money implications, and how exactly do different options result in different prizes. Figure A.4 shows the quiz screen in *SC* conditions. In list conditions the phrase "the choice in the task" was replaced with "the bonus choice in the task", but the quiz itself was the same.

A.3 Task

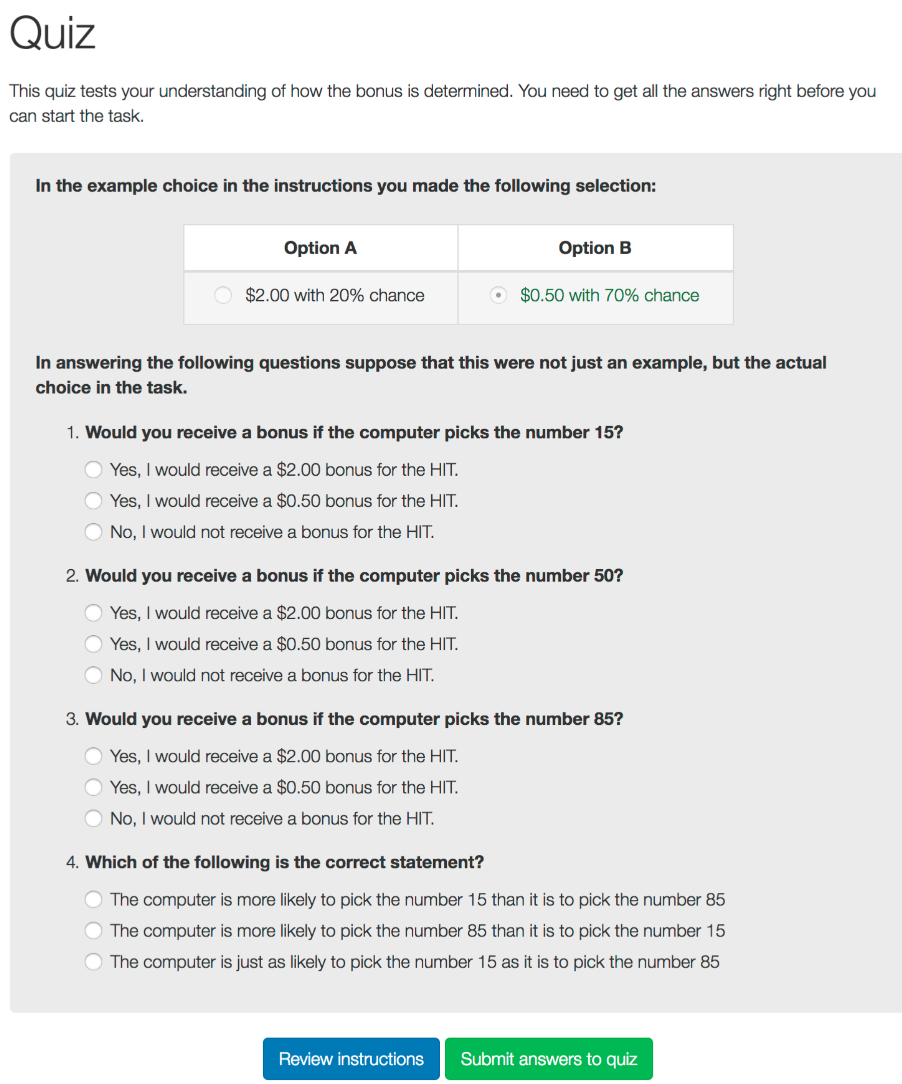
Subjects started the experimental task after completing the quiz. The questions differed between each single choice condition, as did the choice lists between each list conditions. The instructions for the task were, however, identical between all *SC* conditions (Figure A.5), all *R*-list conditions (Figure A.6) and all *K*-list conditions (Figure A.7). Each *K*-list condition had a corresponding *R*-list condition with the exact same choice list, but the instructions were different. The *R*-list instructions reminded subjects that "After you complete the task, the computer will randomly select one of these choices to be the bonus choice", while the *K*-list instructions informed subjects that "The bonus choice is choice #4".

A.4 Optional survey

The task was followed by an optional survey (Figure A.8), which asked subjects some demographic questions (year of birth, age, gender, household income, and the degree to which they face difficulties paying regular expenses), previous experience with similar choices, judgments of similarity (between the prizes, and between the chances), and expected feelings.

A.5 Results screen

The results screen informed subjects of the computer draw, and their consequent financial results.


Quiz

This quiz tests your understanding of how the bonus is determined. You need to get all the answers right before you can start the task.

In the example choice in the instructions you made the following selection:

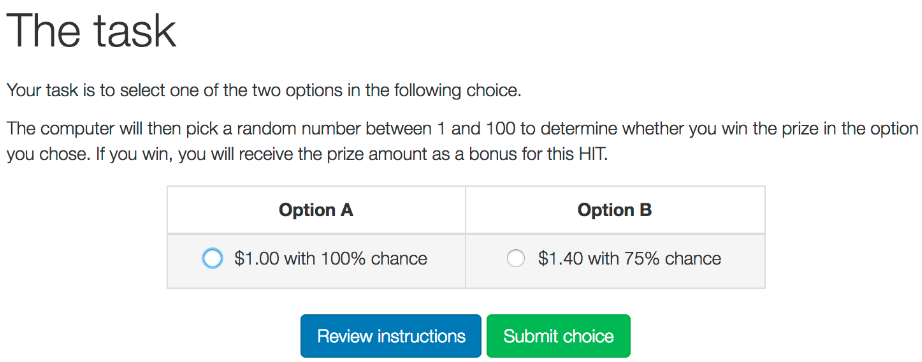
Option A	Option B
$2.00 with 20% chance	$0.50 with 70% chance

In answering the following questions suppose that this were not just an example, but the actual choice in the task.

1. **Would you receive a bonus if the computer picks the number 15?**
   - Yes, I would receive a $2.00 bonus for the HIT.
   - Yes, I would receive a $0.50 bonus for the HIT.
   - No, I would not receive a bonus for the HIT.
2. **Would you receive a bonus if the computer picks the number 50?**
   - Yes, I would receive a $2.00 bonus for the HIT.
   - Yes, I would receive a $0.50 bonus for the HIT.
   - No, I would not receive a bonus for the HIT.
3. **Would you receive a bonus if the computer picks the number 85?**
   - Yes, I would receive a $2.00 bonus for the HIT.
   - Yes, I would receive a $0.50 bonus for the HIT.
   - No, I would not receive a bonus for the HIT.
4. **Which of the following is the correct statement?**
   - The computer is more likely to pick the number 15 than it is to pick the number 85
   - The computer is more likely to pick the number 85 than it is to pick the number 15
   - The computer is just as likely to pick the number 15 as it is to pick the number 85

Review instructions | Submit answers to quiz

Figure A.4: Quiz in single choice conditions (including Allais variation).



The task

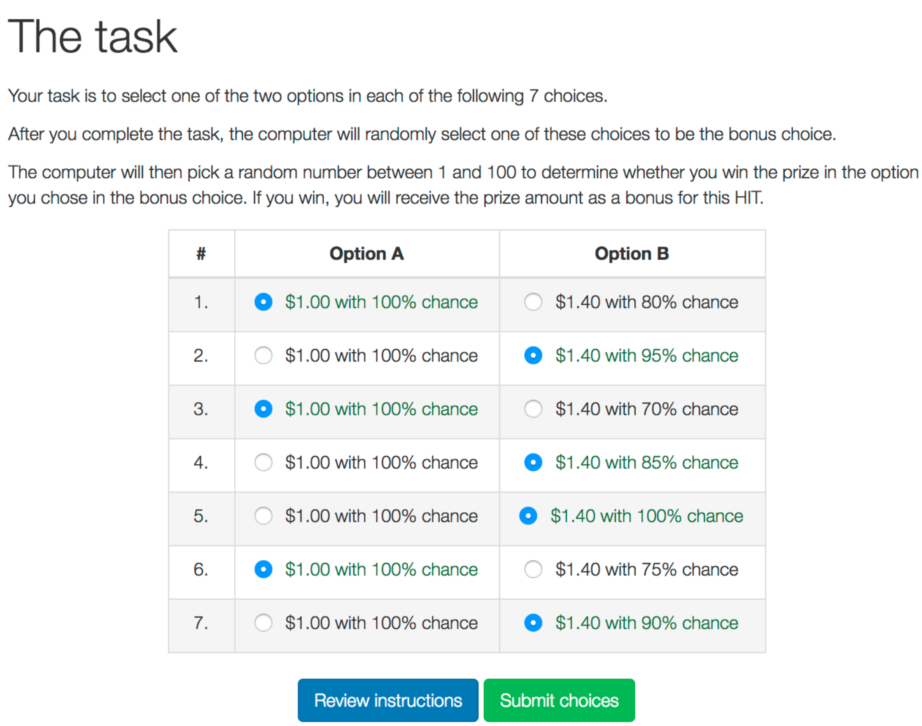
Your task is to select one of the two options in the following choice.

The computer will then pick a random number between 1 and 100 to determine whether you win the prize in the option you chose. If you win, you will receive the prize amount as a bonus for this HIT.

Option A	Option B
$1.00 with 100% chance	$1.40 with 75% chance

Review instructions Submit choice

Figure A.5: The task in the Q75 *SC* condition. Other *SC* conditions included a different question, but the instructions were the same.



The task

Your task is to select one of the two options in each of the following 7 choices.

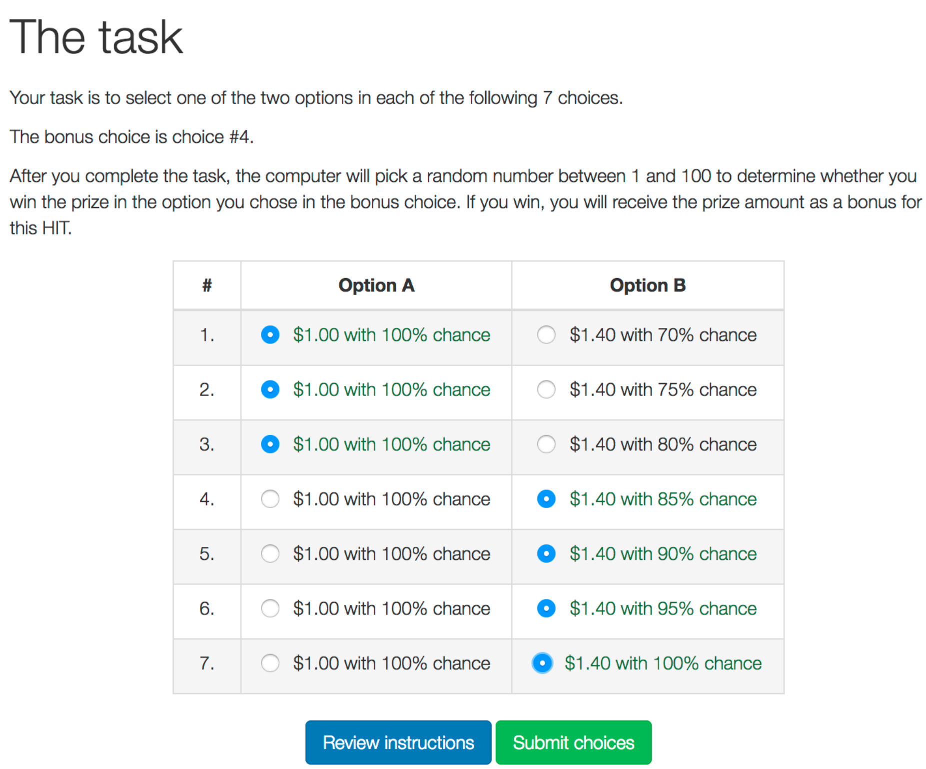
After you complete the task, the computer will randomly select one of these choices to be the bonus choice.

The computer will then pick a random number between 1 and 100 to determine whether you win the prize in the option you chose in the bonus choice. If you win, you will receive the prize amount as a bonus for this HIT.

#	Option A	Option B
1.	$1.00 with 100% chance	$1.40 with 80% chance
2.	$1.00 with 100% chance	$1.40 with 95% chance
3.	$1.00 with 100% chance	$1.40 with 70% chance
4.	$1.00 with 100% chance	$1.40 with 85% chance
5.	$1.00 with 100% chance	$1.40 with 100% chance
6.	$1.00 with 100% chance	$1.40 with 75% chance
7.	$1.00 with 100% chance	$1.40 with 90% chance

Review instructions Submit choices

Figure A.6: The task in the scrambled *R*-list condition. Other *R*-list conditions included a different choice list, but the instructions were the same. The subject in this screen shot chose the risky option in Q85, Q90, Q95, and Q100, and the safe option in Q70, Q75, and Q80.



The task

Your task is to select one of the two options in each of the following 7 choices.

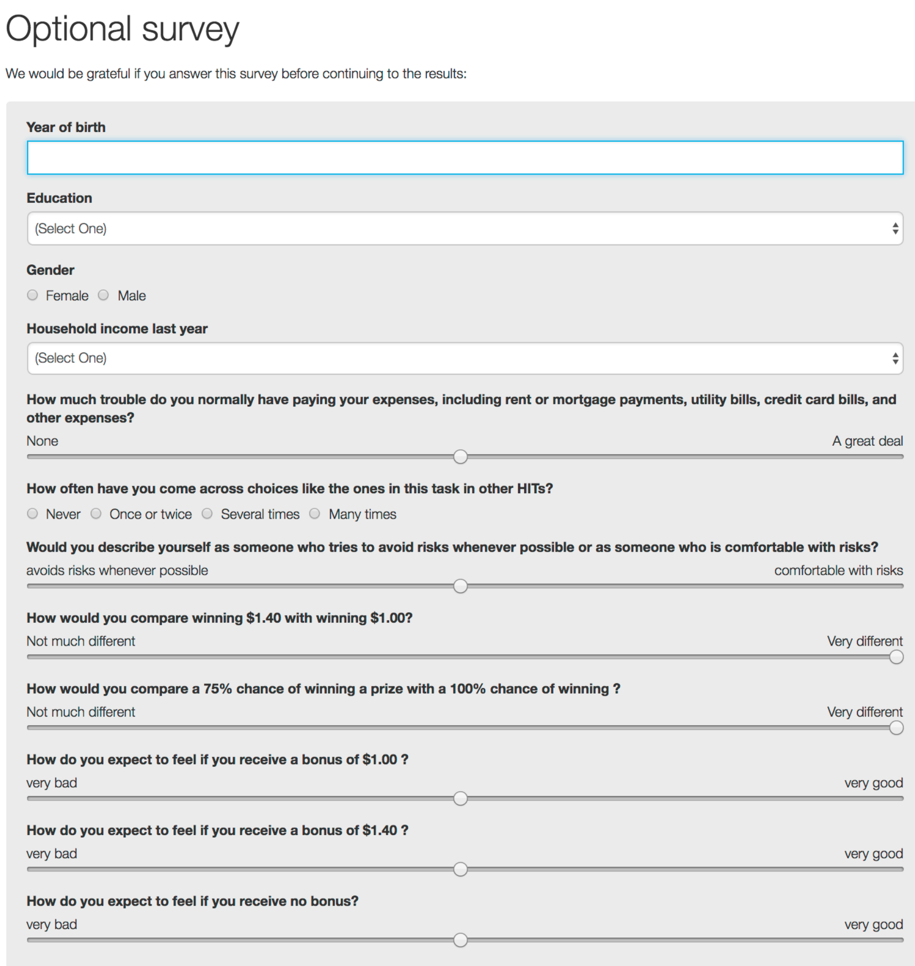
The bonus choice is choice #4.

After you complete the task, the computer will pick a random number between 1 and 100 to determine whether you win the prize in the option you chose in the bonus choice. If you win, you will receive the prize amount as a bonus for this HIT.

#	Option A	Option B
1.	$1.00 with 100% chance	$1.40 with 70% chance
2.	$1.00 with 100% chance	$1.40 with 75% chance
3.	$1.00 with 100% chance	$1.40 with 80% chance
4.	$1.00 with 100% chance	$1.40 with 85% chance
5.	$1.00 with 100% chance	$1.40 with 90% chance
6.	$1.00 with 100% chance	$1.40 with 95% chance
7.	$1.00 with 100% chance	$1.40 with 100% chance

Review instructions Submit choices

Figure A.7: The task in the worst-to-best K-list condition. Other K-list conditions included a different choice list, but the instructions were the same. The subject in this screen shot chose the risky option in Q85, Q90, Q95, and Q100, and the safe option in Q70, Q75, and Q80.



Optional survey

We would be grateful if you answer this survey before continuing to the results:

Year of birth

Education

(Select One)

Gender

Female Male

Household income last year

(Select One)

How much trouble do you normally have paying your expenses, including rent or mortgage payments, utility bills, credit card bills, and other expenses?

None A great deal

How often have you come across choices like the ones in this task in other HITs?

Never Once or twice Several times Many times

Would you describe yourself as someone who tries to avoid risks whenever possible or as someone who is comfortable with risks?

avoids risks whenever possible comfortable with risks

How would you compare winning $1.40 with winning $1.00?

Not much different Very different

How would you compare a 75% chance of winning a prize with a 100% chance of winning ?

Not much different Very different

How do you expect to feel if you receive a bonus of $1.00 ?

very bad very good

How do you expect to feel if you receive a bonus of $1.40 ?

very bad very good

How do you expect to feel if you receive no bonus?

very bad very good

Figure A.8: The post-task survey for a Q75 *SC* subject.

A.6 Feedback

The feedback asked subject to provide feedback on the experiment, and also asked an open question on how they made their choices (we did not analyse the results).

B Preference Narrowing: an Example

The following stylised example illustrates how our preference narrowing model can capture our results. Consider the following three piecewise linear utility-for-money functions. Let $u^0(x) = x$ for all x; let $u^1(x) = x$ for $x \leq 1$ and $u^1(x) = \frac{5}{18}x + \frac{13}{18}$ for $x > 1$; and let $u^2(x) = \frac{87}{68}x$ for $x \leq 2/3$ and $u^2(x) = \frac{15}{34}x + \frac{19}{34}$ for $x > 2/3$. A subject who initially considers the set $\mathscr{W}$ that consists of the three expected utility preferences with utility-for-money functions given above would rank $(\$1.40, .90) \sim (\$1, 1)$ and $(\$1.40, .34) \sim (\$1, .40)$, exhibiting a certainty effect.

Now consider what happens if she reconsiders of her preferences. If she drops only u^1, then she would become less risk averse around certainty and rank $(\$1.40, .85) \sim (\$1, 1)$, but her ranking $(\$1.40, .34) \sim (\$1, .40)$ would remain. If instead she drops only u^2, her preferences around certainty would not change, but she become more risk averse away from certainty and rank $(\$1, .40) \succ (\$1.40, .34)$.[20] If she dropped both u_1 and u^2 from $\mathscr{W}$, then she would be an expected value maximiser.

Thus, a subject who initially considers all rankings in $\mathscr{W}$ when making a single binary choice, but who reconsiders her preferences and drops u^1 after her initial choice in a choice list, would be more likely to choose the risky option in Q85 and Q90 when embedded in a choice list than in Single Choice conditions (as per Observations 1 and 2), yet would exhibit no such difference in our Allais conditions.

[20] This example illustrates that, in our model, preference narrowing need not lead to less risk aversion in choices away from certainty.