Choice lists vs. single choices: same question, same incentives, different answers

David Freeman

Department of Economics, Simon Fraser University

Guy Mayraz

Department of Economics, University of Melbourne

April 29, 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

As researchers, we learn about a person's preferences by observing their choices. The more choices we observe, the more we learn. Experiments that ask subjects to make a large number of choices are thus, potentially, more informative of a subject's preferences. In a common experimental design known as the *choice list*,[1] a sequence of closely-related choice tasks are presented to in a list and one choice task is randomly selected to be paid for real, a payment protocol known as the *random incentive scheme* (RIS). In an argument that relies crucially on context-independent preferences that satisfy the Independence Axiom Holt, 1986; Karni and Safra, 1987, subjects do not know which choice would be paid and thus each choice in the list should reflect their underlying preferences. A behavioural argument Kahneman and Tversky, 1979 posits that subjects tend to make each choice problem they face in isolation from other choices, that is, as though it were the only choice they faced. Either argument leads to the same behavioural implication, referred to as the *isolation hypothesis*, that having a subject respond to many choices incentivised using the RIS does not affect each choice.

The isolation hypothesis can breakdown either because of a*context effect* created by other choices faced in on reported preferences or because of an *incentive effect* due to the RIS or by a combination of both effects. Early tests of the isolation hypothesis found some differences in behaviour depending under the RIS as compared to subjects who made a single incentivised choice between simple lotteries (Starmer & Sugden, 1991; Beattie & Loomes, 1997; Cubitt et al., 1998), however, each observed difference was statistically insignificant, and this literature was taken as support for the use of the RIS. However, several recent prior studies have reported statistically significant violations of isolation (Cox et al., 2014, 2015; Harrison & Swarthout, 2014; Freeman et al., 2016) and attributed their findings to incentive effects. We believe that the structure of choice lists makes the operation of the RIS more transparent to subjects than studies that consist of many unrelated binary choice tasks, yet only one of these papers Freeman et al., 2016 studies a choice list design. In addition, several papers have posited and provided evidence for context effects in choice lists arising from the compromise effect (Andersen et al., 2006; Beauchamp et al., 2015) and reference-dependence (Castillo & Eil, 2014; Sprenger, 2015), however, these studies never compare behaviour in choice lists to subjects who face a single choice. Yet, the most standard revealed preference view would have that a single real choice, by definition, reveals a subject's preference between two options and provides the most nat-

[1] choice list references: old paper, holt and laury, coller and williams, andersen et al., among numerous other papers

ural starting point for documenting any such bias in lists.

Our paper studies whether context or incentive effects lead subjects to choose differently when a choice is embedded in a choice list as compared to when faced on its own. To do so, we study a between-subjects design with a rich set of thirteen experimental conditions that vary both context and incentives (Table 1). In our *R*-list conditions, each subject faced a single choice list with RIS incentives—the one depicted in Figure 1. In each *SC* condition, each subject made a single choice; our five *SC* conditions correspond to the five non-extreme rows of the choice list. In *K*-list conditions, each subject faced a choice lists just like those in *R*-list conditions, but was clearly informed in advance that their middle choice and only that choice would be paid for real. Subjects in *K*-list conditions thus had the same choice list structure as did *R*-list subjects, but since they *k*new for sure that the middle choice would determine their payment, they faced the same incentives as subjects in the *SC* condition corresponding to the middle choice in the list. We also varied whether the questions in a choice list were presented from best-to-worst, worst-to-best, or scrambled in a seemingly random order in both incentive conditions (leading to 2×3 main list conditions). In addition, our two *SC* and *R*-list *Allais* conditions are based on corresponding *SC* and *R*-list conditions with winning probabilities of all lotteries scaled by a factor of 0.4.

In choices involving certainty, we found a economically large and statistically significant difference in choices between *R*-list and *SC*: 41% of *R*-list subjects chose an 85% chance of $1.40 over a certain $1, as compared to 35% of *K*-list subjects and 23% of *SC* subjects. The relative small difference incentivised choices by *K*-list and *R*-list imply that the major difference we observe between choices of *R*-list and *SC* subjects are driven by context effects and not incentive effects. We observe only minor differences in behaviour across different list orders, which suggests that context effects that depend on the order of the list cannot explain the main difference we observe. In addition, we observe no difference in choices between our list and *SC* Allais conditions, indicating that the context effect that we observe is specific to choices around certainty.

The context effect we document, whereby subjects tend to be more likely to choose a risky option over a certain one when the choice is embedded in a list, is in the cannot be explained by existing theories of context effects in choice lists due to reference-dependence (which predicts the opposite) or a compromise effect which biases subjects towards switching in the middle.

Moreover, by exploiting click timing data, we were able to determine that the small gap in risk taking between K-list and R-list subjects is due to a group of K-list subjects who started by answering the question that they knew would determine their payment, and only then proceeded through the remainder of the list. The choices of K-list subjects who completed the list in the usual order are completely indistinguishable from those of *R-list* subjects.[2]

But if RIS incentives cannot explain the difference in risk taking between choice lists and single choices—what can? We consider, and are able to reject, such biases as middle-switching bias (Andersen et al., 2006; Beauchamp et al., 2015) and reference-dependence (Castillo & Eil, 2014; Loomes & Pogrebna, 2014; Sprenger, 2015), and can also reject such psychological mechanisms as the scale compatibility hypothesis (Tversky et al., 1988). Our click timing data shows that subjects generally speed up as they go down the list, suggesting that they gain confidence in their preferences as they go along. This observation fits in with the discovered preferences hypothesis (Plott, 1996), according to which people have a well-defined set of preferences, but they may not know what those preferences are until they discover them—possibly through deliberation, and possibly only through the process of making choices.

In order to account for our data, we construct a formal model of preference narrowing that combines the discovered preferences hypothesis with a notion of cautious decision making that evaluates the desirability of a lottery by the lowest certainty equivalent that is consistent with the current level of uncertainty. In the single choice conditions, subjects have only a limited idea of their preferences. Cautious decision making therefore implies a low certainty equivalent for the risky lottery, making the certain option relatively attractive. In the list conditions, subjects gradually gain a better (more narrow) concept of their preferences, and the cautious criterion assigns a comparatively higher certainty equivalent to each of the risky lotteries. Consequently, subjects in list conditions are more likely to opt for the risky option than are subjects faced with only a single choice.

Preference narrowing undermines the traditional view of the relationship between choice lists and single choices. Single choices are traditionally the standard for revealing preferences, and systematic differences in choices between lists and single choices are traditionally interpreted as indicating a bias in choice lists. The preference narrowing explanation almost reverses this picture. Single choices provide only a minimal opportunity for preference discovery. Being uncertain of their preferences, subjects make cautious choices

[2] $p = 0.597$ as compared with $p = 0.003$ in a comparison with *SC*.

that are biased towards risk aversion relative to their true preferences. Choice lists offer greater opportunity for preference discovery, and choices are consequently less biased. The situation is analogous to the use of repeated trials to approximate the behaviour of experienced market participants. By going through a choice list, subjects gradually discover their true preferences, and come closer to the situation of a person who has acquired significant experience making choice under risk.

Our results join a diverse group of findings that undermine the idea of stable preferences. Perhaps the most important are preference reversals (Lichtenstein & Slovic, 1971; Grether & Plott, 1979) and the anchoring effect (Ariely et al., 2003). But while the anchoring effect suggests that preferences can be affected by transparently irrelevant cues, preference narrowing sets limits to the indeterminacy of preferences—limits that narrow down as subjects gain experience making choices in a particular domain.

2 The Experiment

Subjects were randomly assigned to one of 13 conditions: six choice list conditions, five single choice conditions, and two Allais conditions (Table 1). Subjects in the list conditions were required to complete 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%—henceforth Q70, Q75, etc. The six list conditions differed on two dimensions: incentives and question order. One question was played out for real to determine the payment. In the *R*-list conditions this question was chosen randomly at the end of the experiment, while in the *K*-list conditions subjects knew in advance that it would be the middle question, Q85. The order of the questions could be best-to-worst, worst-to-best, or scrambled (Table 2). Figure 1b shows the choice screen for the best-to-worst order. 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. The *R*-list incentive scheme is often referred to as the Random Incentive Scheme (RIS), though some experimentalists prefer other terms. *K*-list conditions have the same list structure as *R*-list conditions, but the incentives are those of the corresponding single choice—Q85 in all list orders.

The last two conditions were an *R*-list and *SC* Allais conditions. These enable us to test whether our subjects violate the Independence Axiom, and whether any differences between the *R*-list and *SC* conditions are particular to the presence of certainty. 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.

The experiment consisted of six screens: (i) instructions, (ii) understanding quiz, (iii) task, (iv) optional survey, (v) results, and (vi) optional feedback. The experiment was programmed in Javascript, and was accessed via a browser. Appendix A shows screenshots of these screens, and explains how they differed between conditions. We recorded the timing of all choices and button presses. We were thus able 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 online labour market. 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 Amazon Mechanical Turk online labour market enabled us to recruit many more subjects than we could have using a more traditional student sample, resulting in a correspondingly greater statistical power. Our statistical comparisons were binary, so a large number of subjects was required to obtain the necessary statistical power to reject an incorrect null hypothesis for even moderate effect sizes.[3] Amazon Mechanical Turk 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.[4]

In order to ensure a consistent sample of serious subjects, we only allowed

[3]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.

[4]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.

access to US based workers with a minimum of 1000 completed tasks and a 97% approval rate. Invitations were released gradually during US daytime hours.[5] Subjects who accepted the experiment were prevented from retaking it in the future. The participation fee was given as a 'HIT payment' in the Amazon Mechanical Turk terminology, and the additional payment (if any) was paid as a 'HIT bonus'.

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 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 had they been asked to make only that choice. In our experiment, the subjects' incentivised choices should therefore depend only on the choice question, and not on whether the question embedded in a single choice or a *K*- or *R*-list condition. Taken seriously as a hypothesis of the decision-making process, a secondary implication is that the time subjects take to answer any particular question should depend only on the question itself, and not on its position in the list.

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*-

[5] Berinksy et al. (2012) and Freeman et al. (2016) require a 95% approval rate, and the latter study restricts to US-based subjects. Our more stringent criteria are designed to exclude both subjects who tend to complete a large number of HITs without putting in serious effort (by requiring a 97% approval rate) and those subjects who are too new for the preceding requirement to be reliable. Each US account that has completed a minimum number of HITs is associated with a unique Social Security Number and US-based subjects receive monetary payments whereas non-US subjects are paid in Amazon credits.

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.

In the introduction we discussed the RIS incentive compatibility problem. If subjects exhibit the certainty effect, this problem should result in more risk taking in *R*-list than in either *K*-list or *SC*. 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 [REFERENCE], we expect a bias in *R*-list towards a 0.5 frequency of risky choices, and an even greater bias in the non-incentivised choices in *K*-list (no differences between Q85 in *K*-list and in the *SC* conditions).

We use the term *context effect* to refer to any difference in the 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 context as well as incentives, a difference in the proportion of risky choices between *R*-list and *SC* can be generated by either or both effects. However, a context effect would also generate a difference between behaviour in Q85 across the *K*-list to *SC* conditions, but no differences between *R*-list and *K*-list conditions. Finally, some particular context effects also predict a difference between different order variants of a list, but others do not.

One particular form of context effect posits that the structure of an ordered list biases subjects towards switching around the middle—what has become known as *middle-bias* (Andersen et al., 2006; Beauchamp et al., 2015). In our experiment, this 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, middle-bias has no implications for behaviour in Q85. 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*.

It has also been proposed that making several choices on the same side of the list creates a status-quo, biasing subsequent choices in the same direction.[6] 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, including those who violated monotonicity.[7] The raw results for the 13 conditions are in Table 4. We start the analysis by comparing the three choice mechanisms against each other: (i) single choices (*SC*), (ii) choice lists with RIS incentives (*R*-lists), and (ii) choice lists in which subjects know in advance that their payment would depend only on their answer to Q85 (*K*-lists). The different order variations of these two list mechanisms are lumped together for this comparison. Figure 2 displays the proportion of risky choices for different combinations of question and choice mechanism. 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 gaps 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 much smaller and not statistically significant.

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

The K-list has the same structure as the *R*-list and the same incentives as 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 gap between K-list and *SC* is statistically significant ($p = 0.013$), providing clear evidence of a context effect. While we observe a differences between choices in the *R*-list and K-list in Q85, this incentive effect is not statistically significant ($p = 0.152$).

[6] Loomes and Pogrebna, 2014 find evidence of such an effect.

[7] Monotonicity violations are possible in list conditions, but not in single choice conditions. Dropping subjects for monotonicity violations would have risked biasing our results.

Observation 2. Subjects in K-list conditions were also much more likely to choose the risky option than *SC* subjects in Q85, and only insignificantly less likely 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 between decisions. K-list subjects. Figure 3 shows that in list conditions, subjects tended to start at the top of the list, and not a single *R*-list subject started with the middle question (Q85). However, 23% of K-list subjects started with Q85, which (as they were informed) was the only incentivised question they faced.

We then examined whether this initial focus on Q85 affected choices. As is evident from Figure 2, subjects who completed Q85 first were substantially less likely to choose the risky option in this question. 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* ($p = 0.003$). Interestingly, in all the other questions (unincentivised in K-list) both groups made very similar choices to *R*-list subjects. Importantly, there is no difference in the survey measure of general risk tolerance between subjects who started in Q85 and other subjects, and the results hold when we include individual regressors (Observation 5).

Observation 4. Risk-taking in Q85 depended crucially 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[8] (Table 6). The most significant predictor of choices in these regressions was self-reported tolerance for risk,[9] but the *R*-list vs. *SC* difference had about twice the impact as gender—the second most important 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[10]; (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 random differences in individual characteristics does not change our results.

By comparing the timing of choices in the best-to-worst and worst-to-best orders variants of *R*-list, we can test to what extent subjects spend more time on choices in which they are closer to indifference, as would be the case if they answer each choice in isolation. Figure 4 shows the median time to answering each of the different questions in the list, measured from the previous answer. The time for answer the first question is measured from the start of the task, and includes the time for reading the instructions. The sample is limited to subjects who completed the list from top to bottom and did not revise previous answers. 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.[11]

Observation 6. The time subjects take for responding to questions is determined not by how difficult they are, but by their position in the list.

[8]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.

[9]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.

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

[11]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.

The theoretical problem with RIS incentives can only arise if preferences violate the Independence Axiom. We can test for this by comparing the proportion of risky choices between the Q85 *SC* condition and the Allais *SC* condition, in which both chances were multiplied by a factor of 0.4. The results (Table 4) reveal a substantial difference: only 23% of *SC* subjects chose the risky option in Q85, but this number was fully 55% in the corresponding Allais condition ($p < 0.001$). A comparison of the regular and Allais *R*-list conditions makes it possible to examine choices over the full range of questions. The differences are consistent across all 7 questions, though much smaller in size (e.g. 41% vs. 51% in Q85). Notably, the difference between *R*-list and *SC* in the proportion of risky choices disappears in the Allais choices.

Observation 7. Allais comparisons reveal a strong certainty effect in single choice conditions, and a much weaker one in list conditions. This results from a combination of (i) a large difference in risk taking between *R*-list and *SC* in the vicinity of certainty, and (ii) no difference in Allais conditions 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 7% of subjects who choose a lottery that is transparently first-order stochastically dominated in Agranov and Ortoleva (2017).[12] 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.

[12] See Table A8, Part I, FOSD 2

5 Discussion

Our data violate isolation hypothesis with more risk-taking when a choice was embedded in an *R*-list than when it was faced on its own (Observation 1). This violation of isolation arises primarily from context effects; incentive effects are either not present or relatively small in comparison (Observation 2). Since we tend not to observe differences in behaviour across list conditions with different list orders (Observation 3), these context effects in our data must arise primarily due to presenting the subject with many choices. This context effect must be weaker away from certainty, since we do not observe such differences when a studying choices away from certainty (Observation 7).

Yet none of the theories we considered ex-ante can account for these patterns. Static reference-dependence predicts that subjects will be more less prone to choose the safe option when it is embedded in our main list conditions – the opposite of what we observe.[13] Dynamic reference-dependence incorrectly predicts that choices will depend on list order. Middle-bias cannot explain the behavioural differences between choices in Q85 on its own vs. in lists. While middle-bias correctly predicts more risk-taking in Q90 in the list conditions than in the *SC* condition, but it also predicts less risk-taking in Q80 in the list conditions than in the *SC* condition while we find no such difference; thus at best, middle-bias must either be the result of, or operate on top of some other behavioural tendency that leads isolation to be violated. As well, theories based on incentive effects, like those that predict failures of isolation due to the interaction between non-expected utility preferences and the RIS (Holt, 1986; Karni & Safra, 1987), also find no support in our data.

We could perhaps consider context-based explanations from the original preference reversal literature. Tversky et al. (1990, 1) summarise the leading explanations of reversals between 'matching' tasks (like the BDM) and choice tasks. They posit that in matching tasks, subjects tend to put more weight on the dimension being matched (a *scale compatibility* effect), and also that

[13] We note that in studies that have proposed static reference-dependence in lists **castillo2016tariffing**; Sprenger, 2015, each subject faces many choice lists; none of their subjects face a Single Choice type condition. This is also true of studies that have proposed a middle-bias in lists **andersen2006list**; **benjamin2016debiasing** Thus our findings do not directly contradict theirs in any way. These biases may well operate in choice lists, but our data imply that a separate context effect also exists that generates systematic differences in subjects' choices based on whether a question is faced on its own or embedded in a choice list.

subjects tend to put relatively more weight on the more prominent attribute in choice tasks (an *attribute prominence* effect), which is hypothesised to be the winning probability for one-outcome gambles. If we view the choice list as having features of both matching and choice tasks, such an explanation can apply to our data. But as pointed out by Slovic? (can't find the reference right now) and shown by **cubitt2004preference** these two effects work in opposite directions and tend to cancel out when probabilities are being matched – and our choice list is analogous to such a matching task. Given this, we see no reason why only one of these two effects would be markedly stronger in our setting. In addition, each effect applies equally to our Allais conditions, where we observe no difference between choices in our *SC* and *R*-list conditions. Thus, these effects cannot explain our results.[14]

So then what can explain our findings? We believe that our non-choice data gives a hint. The small minority of subjects who did the incentivised Q85 first in the *K*-list conditions behaved indistinguishably from subjects in the *SC* condition who made faced only Q85 (Observation 4). In contrast, choices of the remaining group of *K*-list subjects were indistinguishable for thsoe of *R*-list subjects, none of whom completed Q85 first. These facts suggest that the main difference we observe reflects a carried over influence of prior choices on subsequent ones combined with the strong tendency of subjects to complete the list from top to bottom. In addition, we see that subjects tend to respond more quickly as on questions that appear later in a list (Observation 6). Prior work suggests that subjects take longer to make choices where they are closer to indifference Konovalov and Krajbich, 2016, yet in the worst-to-best list, subject tend to spend more time on the second question (Q75), where the vast majority of subjects choose the safe option, than on later questions (Q85-Q95) where subjects' choice frequencies are more evenly split. From this, we infer that subjects' consideration given to preceding choices carries over and informs subsequent ones. The consideration process occuring before Q85 was answered would have differed between list subjects and *SC* subjects, but a list subject who completed Q85 first would likely have had a consideration process more similar to *SC* subjects than to list subjects who completed the list from top to bottom. Thus we conjecture that this consideration process played a causal role in generating more risk-taking in our list conditions.

Next, we present a formal model that builds on Plott's discovered preference hypothesis and can explain our main findings.

[14] Tversky et al. (1988)

6 Preference narrowing with cautious decisions

Our model of preference narrowing with cautious decisions has three main features, which taken together are sufficient to explain our main findings. First, a decision-maker is initially uncertain about their own preferences, and considers multiple preference relations as 'reasonable',[15] Second, when uncertain about their own preferences, a decision-maker applies a cautious decision criterion by evaluating the desirability of a lottery based on its certainty equivalent as evaluated by the 'reasonable' preference relation that assigns it the lowest certainty equivalent,[16] Third, when considering a choice or related choices, a decision-maker gradually narrows down the set of preferences that they consider 'reasonable' as they make choices. Subjects facing choice lists thus narrow down their preferences as they make choices in the list, or perhaps by considering these choices in advance of actually making them. Either way, when combined with cautious decision-making, this consideration process reduces risk aversion around certainty, leading to more risky choices in the list conditions. We provide a formal treatment below.

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 complete, transitive, and continuous preference relation. $\mathcal{W}$ 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* $c_{\mathcal{W}}(x, p)$:

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

[15] We view this first assumption as closely related to what Butler and Loomes (2007) refer to as "preference imprecision". See Ok (2002) for the link between preference incompleteness and multiple possible preference relations.

[16] 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. We note that since in our model lotteries are compared based on their certainty equivalents, the utility of different lotteries can be computed according to different utility functions and still compared on a common monetary scale.

Finally, let $\mathcal{W}(C)$ 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) $\mathcal{W}(C) \subseteq \mathcal{W}$, and (ii), 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 $p = 1$. Hence, the experience of making choices increases the attractiveness of risky lotteries relative to safe prizes. Consider a particular choice 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, we conclude that the proportion of risky choices in the list conditions should be (at least weakly) higher than in single choice conditions.

The following stylised example illustrates how our preference narrowing model can capture our results. Define the three utility-for-money functions u^0, u^1, u^1 by

$$u^0(x) = x$$

$$u^1(x) = \left\{ \begin{array}{ll} x & \text{if } x \leq 1 \\ \frac{5}{18}x + \frac{13}{18} & \text{if } x > 1 \end{array} \right\}$$

$$u^2(x) = \left\{ \begin{array}{ll} \frac{87}{68}x & \text{if } x \leq \frac{2}{3} \\ \frac{15}{34}x + \frac{19}{34} & \text{if } x > \frac{2}{3} \end{array} \right\}$$

A subject who initially considers the set $\mathcal{W}$ 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 reconsideration 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 $(\$1, .40) \succ (\$1.40, .34)$.[17] If she dropped both u_1 and u^2 from $\mathcal{W}$, then she would be an expected value maximiser. Thus, a subject who initially considers all rankings in $\mathcal{W}$ 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

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

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.

The above assumptions imply that the preference narrowing process effectively converges to a single 'true' preference relation $\succsim^*$.[18] In this sense, we view our model as a refinement of Plott's Plott, 1996 discovered preference hypothesis, which posits that subjects discover their preferences over time and with experience. This key property of our model contrasts with psychological models of constructed preferences (Lichtenstein & Slovic, 2006), which have no analogue of a true preference relation.

Braga and Starmer (2005) show that several well-known anomalies are consistent with the discovered preference hypothesis. They specifically note the study of Barron and Erev (2003)[19] where subjects who repeatedly face Allais type binary choice problems tend towards expected value maximisation over time—consistent with a specification of our model with a $\succsim^*$ that is approximately risk neutral over small stakes.

7 Conclusion

We found that subjects were substantially more likely to choose a risky option when a choice was embedded in a choice list. The primary effect behind this difference was the context created by the choice list and not the RIS incentives. However, this difference did not extend to choices away from certainty.

This failure of isolation has broad implications for the risk preferences that can be inferred from experiments. Our findings indicate that how preferences are elicited do not just affect subjects' inferred risk aversion, but also the extent to which subjects exhibit a certainty effect. That is, experimental procedures have both a quantitatively and also a qualitatively influence on a subject's inferred risk preferences.

In principle, context effects like those we document present a fundametal challenge to economists' model of choice as the maximisation of preferences. Yet our explanation of our findings as mediated by the consideration process – as encapsulated in our model of preference narrowing with cautious decisions

[18] 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.

[19] See also Kuilen and Wakker (2006).

– provides clear guidance for what can and cannot be inferred about underlying preferences from experimental choices. Moreover, the consideration-mediated context effect we posit has clear real world analogues. We posit that if we are interested in studying the choices of experienced descision-makers, then a experiment that uses choice lists may be better suited for that task than a design involving a single choice or a variation thereof, while if instead we are interested in choices of inexperienced decision-makers following a limited amount of consideration, the opposite will be true.

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.

Barron, G. & Erev, I. (2003). Small feedback-based decisions and their limited correspondence to description-based decisions. *Journal of Behavioral Decision Making*, *16*(3), 215–233.

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.

Braga, J. & Starmer, C. (2005). Preference anomalies, preference elicitation and the discovered preference hypothesis. *Environmental and Resource Economics*, *32*(1), 55–89.

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*. presented at the Southern Economic Association Meeting in Atlanta.

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., 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*.

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). The independence axiom and the bipolar behaviorist. *Theory and Decision*, *77*, 423–438.

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.

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*, 675–685.

Konovalov, A. & Krajbich, I. (2016). *Revealed indifference: Using response times to infer preferences*. Working Paper.

Kuilen, G. v. d. & Wakker, P. P. (2006). Learning in the allais paradox. *Journal of Risk and Uncertainty*, *33*(3), 155–164.

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*, 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.
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., 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), 204–217.

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. Three *R*-list conditions differ in the question order (Table 2).
K-list Best-to-worst Worst-to-best Scrambled	Same questions as in *R*-list conditions with one crucial difference: the question chosen for payment is always question 4 (Q85), and subjects know this *before* completing the task.
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 chances 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	$=$	$\neq$	$\neq$
K-list vs. SC	$=$	$=$	$\neq$
K-list vs. R-list	$=$	$\neq$	$=$
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.

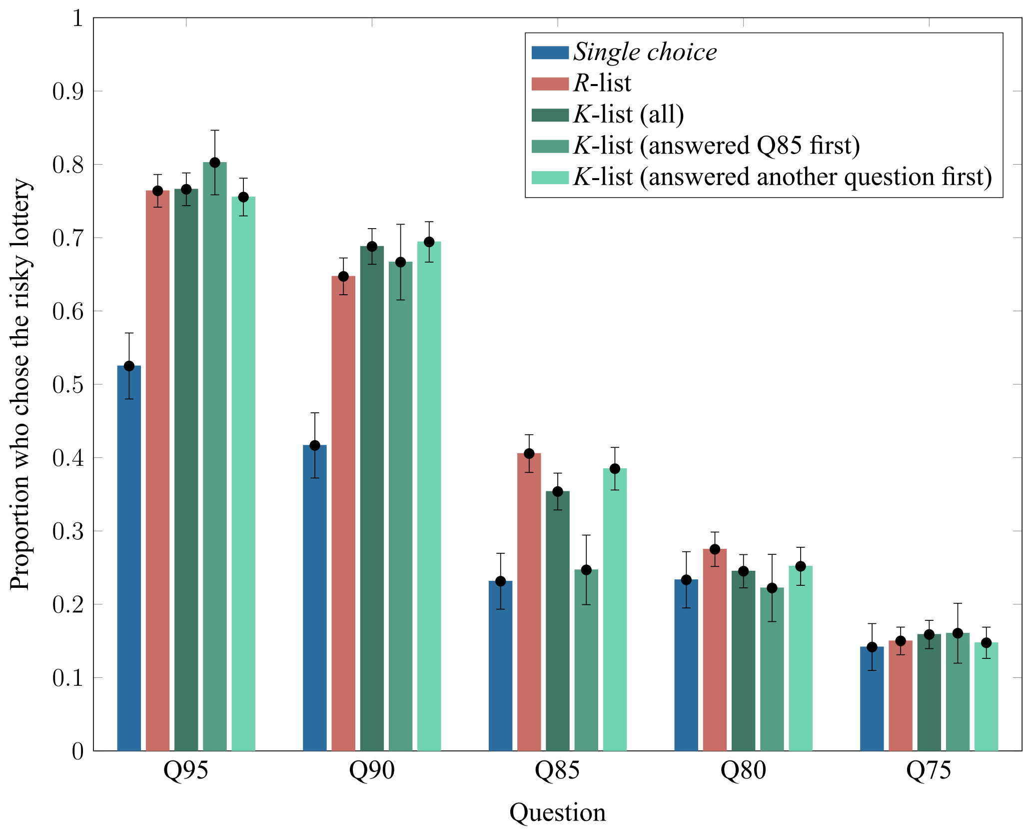


Figure 2: Risk taking by condition. Bars show the proportion of subjects who chose the risky option. Results for K-list are presented both in aggregate and broken down by whether subjects answered the 85% question first. Errors bars denote the standard error of the mean.

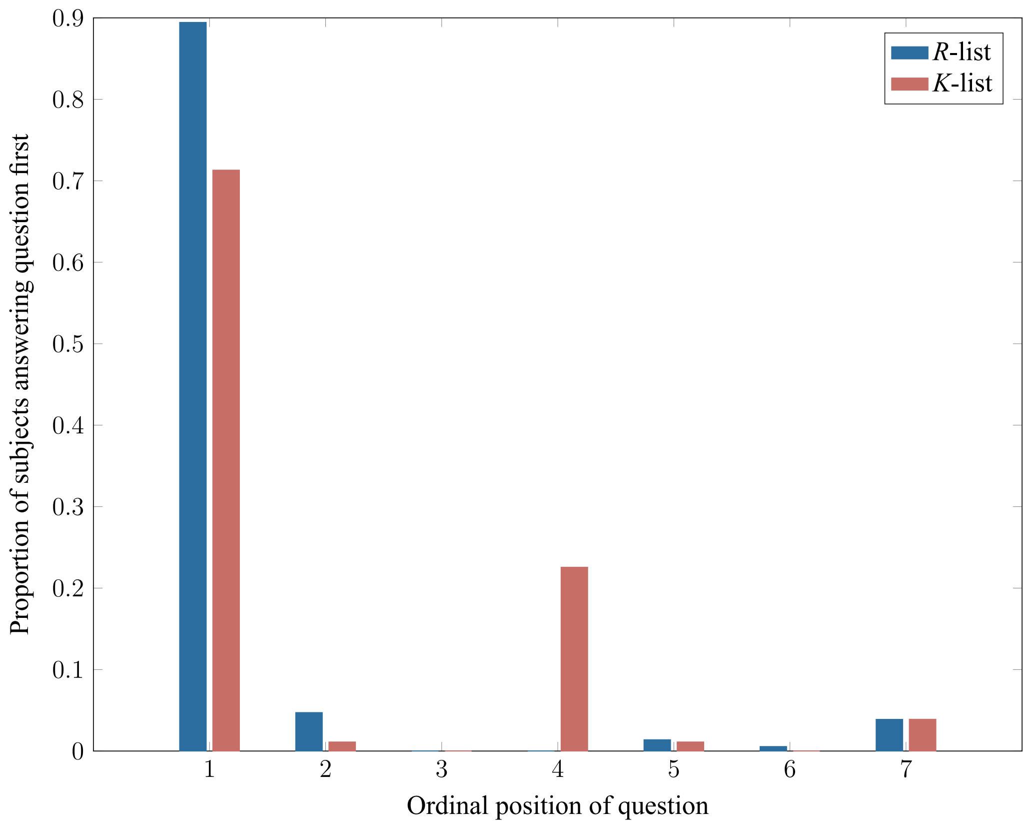


Figure 3: The first question tackled by *R*-list and *K*-list subjects.

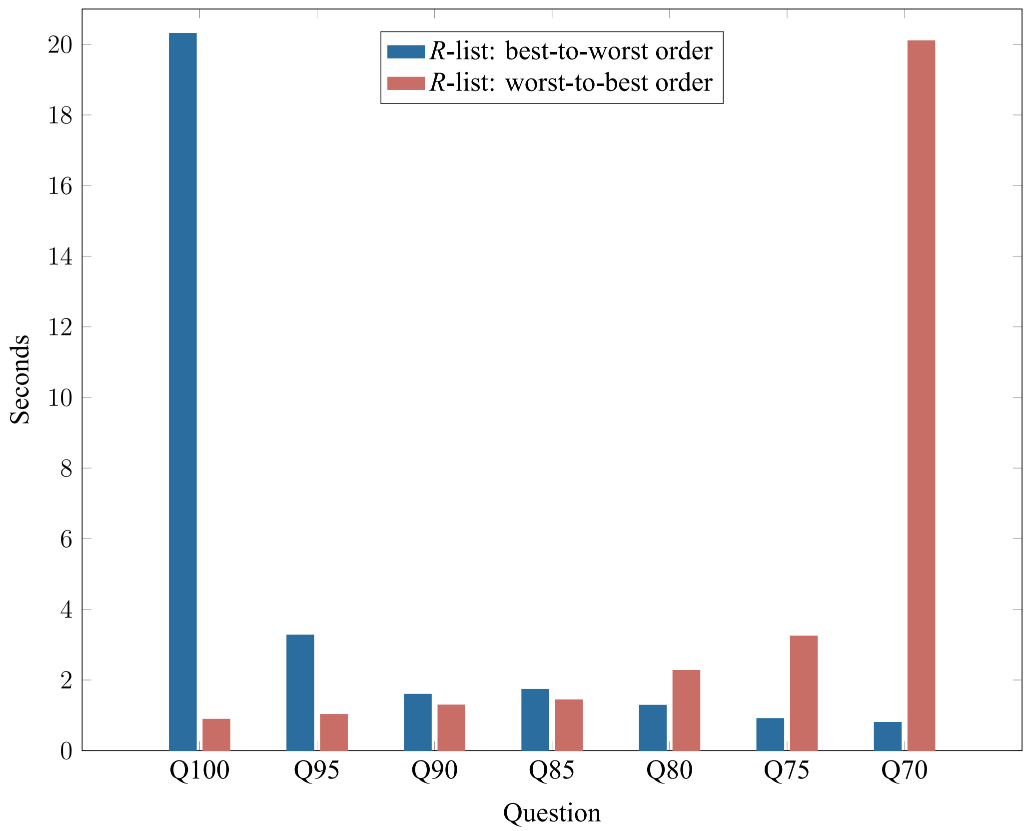
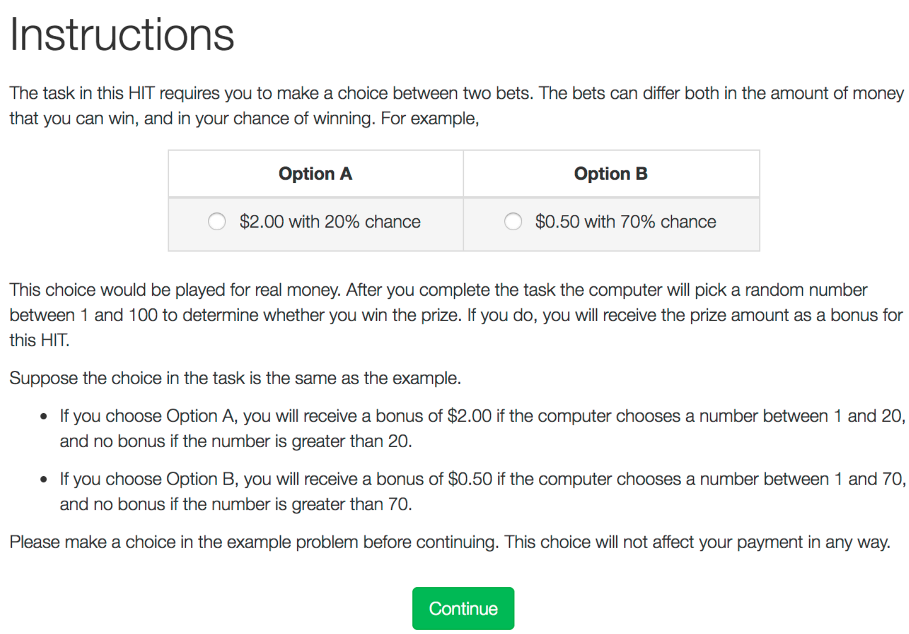


Figure 4: Median time to answer the different questions in the list. Sample limited to subjects who answered each question just once and in order.



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.

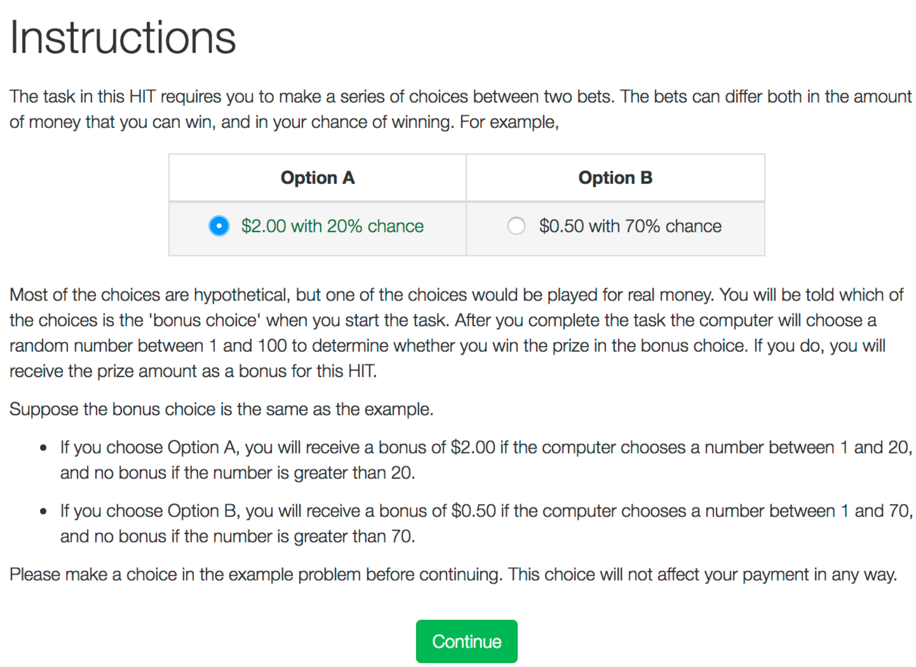
Suppose the bonus choice is the same as the example.

- 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

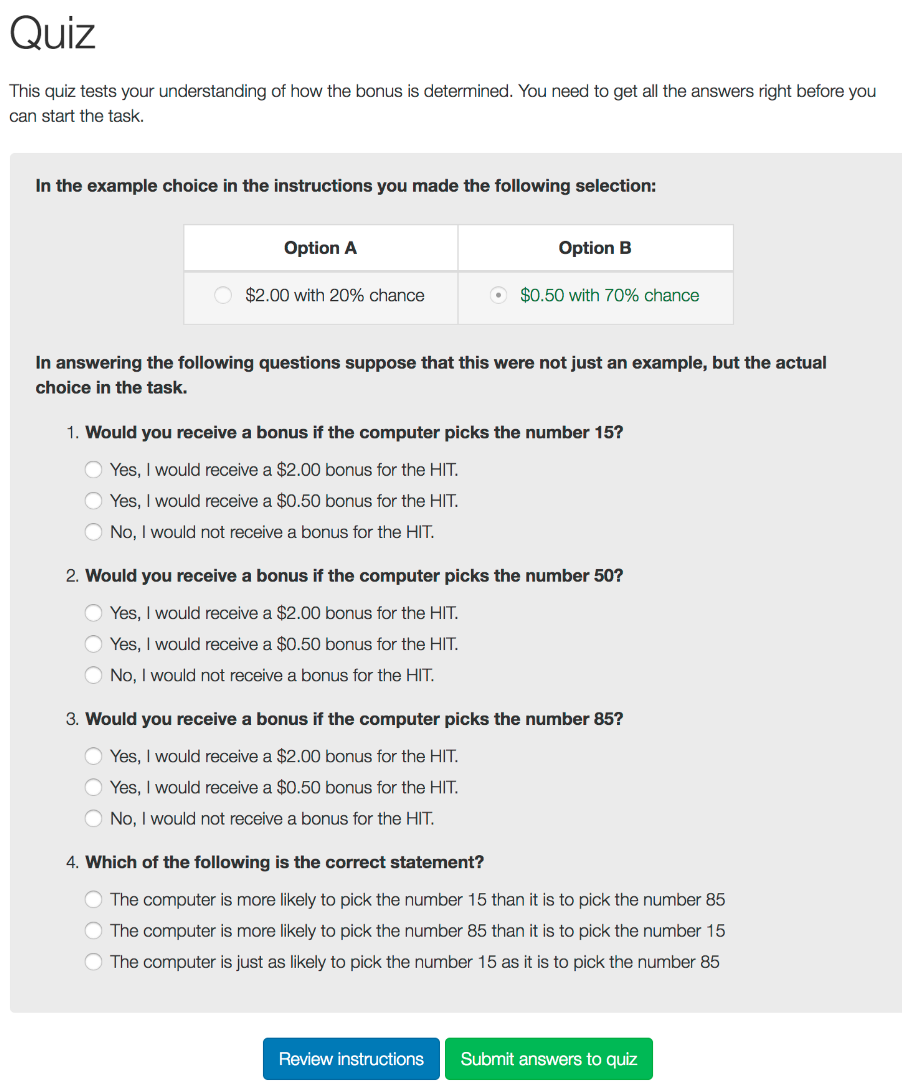
Subjects started the experimental task after completing the quiz. The questions differed between each single choice condition, as did the choice lists between each list conditions. The instructions for the task were, however, identical between all *SC* conditions (Figure A.5), all *R*-list conditions (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.

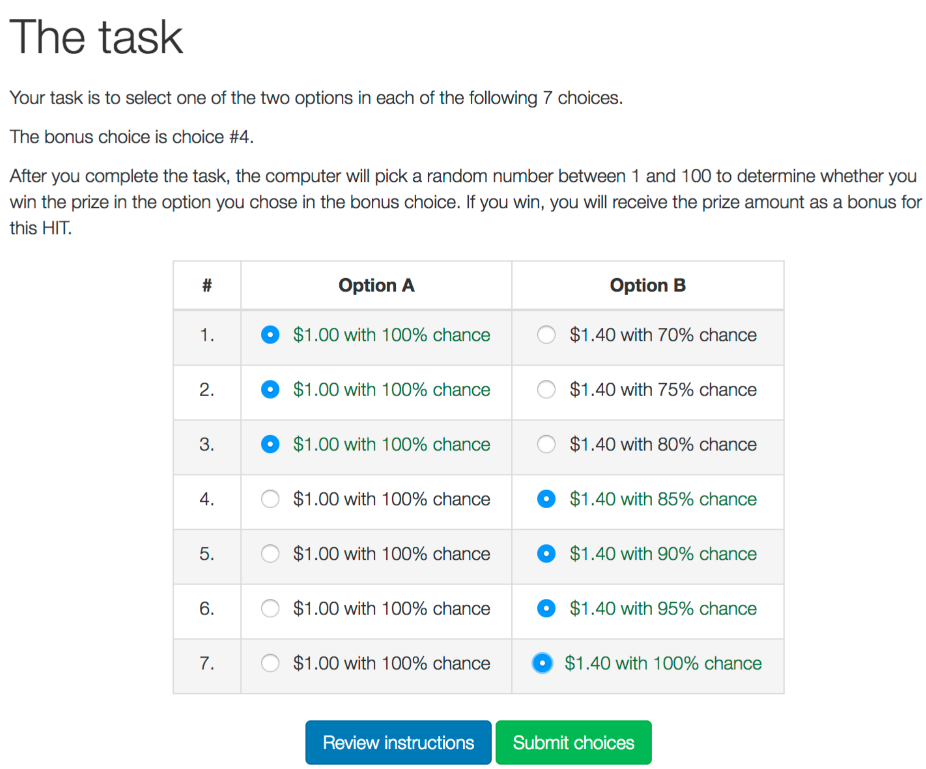
After you complete the task, the computer will randomly select one of these choices to be the bonus choice.

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.

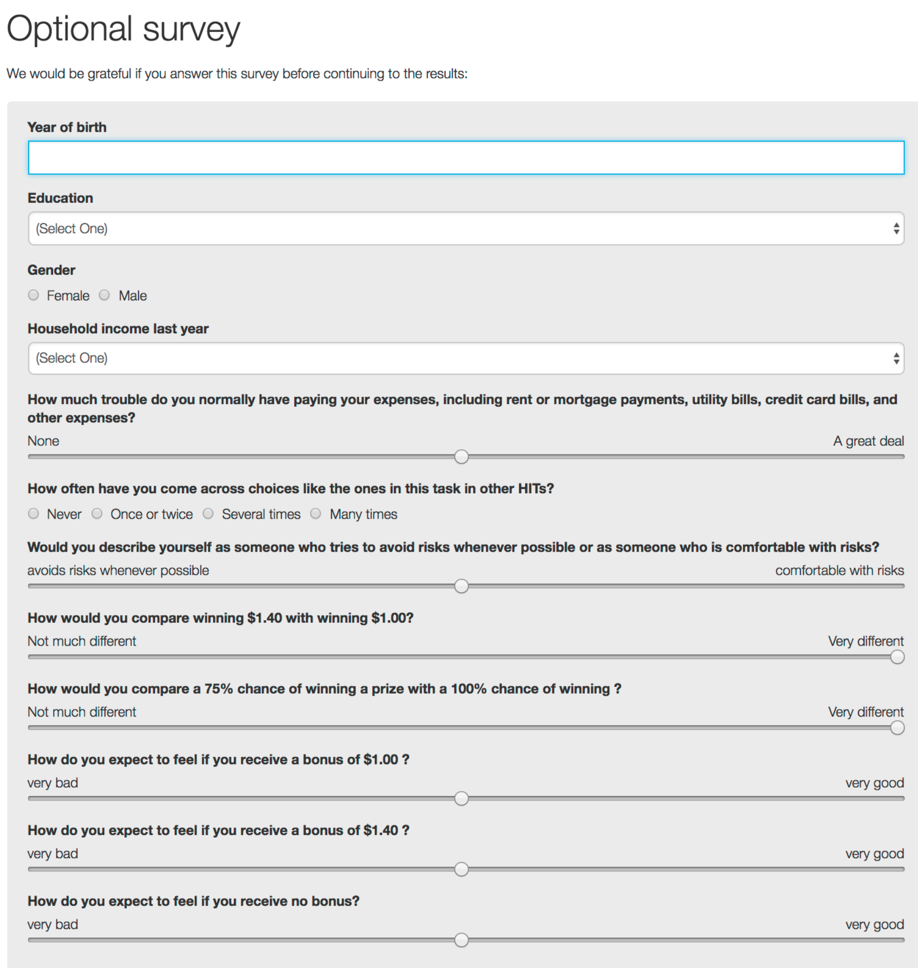
The bonus choice is choice #4.

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).