Exp Econ (2012) 15:418–443
DOI 10.1007/s10683-011-9306-4

Random incentive systems in a dynamic choice experiment

Guido Baltussen · G. Thierry Post · Martijn J. van den Assem · Peter P. Wakker

Received: 18 May 2010 / Accepted: 30 September 2011 / Published online: 22 October 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract Experiments frequently use a random incentive system (RIS), where only tasks that are randomly selected at the end of the experiment are for real. The most common type pays every subject one out of her multiple tasks (within-subjects randomization). Recently, another type has become popular, where a subset of subjects is randomly selected, and only these subjects receive one real payment (between-subjects randomization). In earlier tests with simple, static tasks, RISs performed well. The present study investigates RISs in a more complex, dynamic choice experiment. We find that between-subjects randomization reduces risk aversion. While within-subjects randomization delivers unbiased measurements of risk aversion, it does not eliminate carry-over effects from previous tasks. Both types generate an increase in subjects' error rates. These results suggest that caution is warranted when applying RISs to more complex and dynamic tasks.

Keywords Random incentive system · Incentives · Experimental measurement · Risky choice · Risk aversion · Dynamic choice · Tremble · Within-subjects design · Between-subjects design

JEL Classification C91 · D81

Electronic supplementary material The online version of this article (doi:10.1007/s10683-011-9306-4) contains supplementary material, which is available to authorized users.

G. Baltussen · M.J. van den Assem (✉) · P.P. Wakker
Erasmus School of Economics, Erasmus University of Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
e-mail: vandenassem@ese.eur.nl

G.T. Post
Koç University Graduate School of Business, Istanbul, Turkey

https://doi.org/10.1007/s10683-011-9306-4 Published online by Cambridge University Press

Individual choice experiments commonly use a random incentive system (RIS) to implement real incentives. In the most common RIS, each subject performs a series of individual tasks, knowing that only one of these tasks will be randomly selected at the end to be for real. Although there have been several debates about the validity of the method, it is now widely accepted in studies of individual choice (Holt 1986; Lee 2008; Myagkov and Plott 1997; Starmer and Sugden 1991).

If a subject performs multiple tasks in an experiment where each task is for real, then income and portfolio effects will arise (Cho and Luce 1995; Cox and Epstein 1989). The RIS is the only incentive system known today that can avoid such effects. In addition, for a given research budget and with the face values of the monetary amounts kept the same, RISs allow for a larger number of observations. Early studies that implemented the RIS include Allais (1953), Grether and Plott (1979), Reilly (1982), Rosett (1971), Smith (1976), Tversky (1967a, 1967b), and Yaari (1965). Savage (1954, p. 29) credits W. Allen Wallis for first proposing the RIS.

More recently, a second, more extreme type of RIS has been used, where not every subject is paid. A subset of the subjects is randomly selected, and one of their tasks, randomly selected, will be for real. A drawback of this procedure is that the probability of real payoff is further reduced for every task, possibly inducing lower task motivation. In return, however, higher prizes can be awarded to the subjects selected, which may improve motivation. Studies that apply this incentive method include Bettinger and Slonim (2007), Camerer and Ho (1994), Cohen et al. (1987), and Schunk and Betsch (2006).

In the more extreme type of RIS, there is a selection of both tasks and subjects. To investigate the effects of these two elements of the randomization process in isolation, our study will consider, besides the most common RIS, a RIS where there is no selection of tasks. That is, each subject performs only one single task, after which a subset of subjects is randomly selected for whom the outcome of their task will be for real. We call this design the *between-subjects* RIS (BRIS), and the other, common design the *within-subjects* RIS (WRIS). The recent and more radical design where both tasks and subjects are selected is called the *hybrid* RIS. Studies using a pure BRIS include Tversky and Kahneman (1981, endnote 11) and Langer and Weber (2008).

Most tests of RISs are based on static choice tasks (Sect. 1 provides references). However, many experimental and real-world decision problems are dynamic. We analyze the effects of RISs in a dynamic choice experiment that is based on the popular TV game show *Deal or No Deal* (DOND). This show has received substantial attention from researchers, and is widely recognized as a natural laboratory for studying risky choice behavior with large real stakes (Blavatskyy and Pogrebna 2008, 2010a, 2010b; Brooks et al. 2009a, 2009b; Deck et al. 2008; Post et al. 2008). DOND is dynamic because it uses multiple game rounds and in each round the choice problem depends on the outcomes of earlier rounds. Section 2 provides details on the game.

We investigate three different treatments. In the first, called *basic* or *guaranteed payment* treatment, each subject plays the game only once and for real. Because every subject faces only one task and knows that it is for real, the observed choices represent an unbiased benchmark for analyzing possible distortion effects of RISs. We therefore use it as the gold standard. In the second treatment (WRIS), subjects play the game ten

https://doi.org/10.1007/s10683-011-9306-4 Published online by Cambridge University Press

times, one of which is then randomly selected for real payment. In the third treatment (BRIS), each subject plays the game once, with a ten percent chance of real payment. The hybrid RIS was not implemented in a separate treatment because its incentives, 100 times smaller per task than in the basic treatment, would be insufficiently salient. This point is further discussed in Sect. 4.

We only vary the incentive system across these three treatments. All other factors, in particular the face values of the prizes in the game, are held constant. This implies that the expected payoffs per task are not similar across treatments. Holding face values constant is, however, exactly what experimenters do when using the RIS: subjects are assumed to treat each task in isolation and as if it were for real. Keeping the face values constant is thus precisely in line with the purpose of our study to examine the validity of RISs. By comparing the choices in the RIS treatments with those in the basic treatment, we investigate whether between- and within-subjects randomization lead to biased estimates of risk aversion. In addition, by comparing choices after tasks that ended favorably with choices after tasks that ended unfavorably, we analyze possible carry-over effects in the WRIS treatment.

We find that risk aversion in the WRIS treatment is not different from that in the basic treatment. However, we observe strong carry-over effects from prior tasks: the more favorable the outcomes in the two preceding games are, the less risk aversion there is in the current game. The BRIS is based on one task per subject and thus avoids such carry-over effects. However, risk aversion is substantially lower in this treatment than in the basic treatment. Furthermore, we find some evidence of errors that are unrelated to the characteristics of the choice problem in both RIS treatments.

1 Background and motivation

RISs are known under several names, including *random lottery incentive system* (Starmer and Sugden 1991), *random lottery selection method* (Holt 1986), *random problem selection procedure* (Beattie and Loomes 1997), and *random round payoff mechanism* (Lee 2008). The different names apply to particular types of experiments (risky choice or social dilemma), rewards (lotteries or outcomes), or tasks (composite or single-choice). We use *random incentive system*, because it can be used for any type of experiment, reward, and task.

Holt (1986) raised a serious concern about WRISs and stated that subjects may not perceive each choice in the experiment in isolation. Rather, they may perceive the choices together as a meta-lottery, or a probability distribution over the different choices in the whole experiment and their resulting outcomes. Such a perception may lead to contamination effects between tasks if subjects violate the independence condition of expected utility. A large body of research indicates that people indeed systematically violate this condition (Allais 1953; Carlin 1992; Starmer 2000). In the literature, the extreme and implausible case where subjects perfectly integrate all choices and the RIS lottery into one meta-lottery is known as *reduction*. Milder forms are also conceivable, where subjects do not combine all choices and the RIS lottery precisely, but where they do take some properties of the meta-lottery into account.

Contrary to what has sometimes been thought, independence (together with appropriate dynamic principles) is sufficient but not necessary for the validity of RISs.

https://doi.org/10.1007/s10683-011-9306-4 Published online by Cambridge University Press

The case where subjects take each choice in the experiment as a single real choice is called *isolation*. Isolation leads to proper experimental measurements under RISs, also if independence is violated on other occasions.

The validity of the WRIS has been investigated in several studies. In a cleverly designed experiment based on simple, pairwise decision problems, Starmer and Sugden (1991) found isolation verified. However, in a direct comparison of the choices in a RIS treatment with those in a sample with guaranteed payment, they found a marginally significant difference. This difference has not been confirmed in later studies. Using more subjects, Beattie and Loomes (1997) and Cubitt et al. (1998a) concluded that there is no evidence of contamination for simple, pairwise decisions. Camerer (1989) also found that WRISs elicit true preferences. After the gamble to be implemented for real had been determined, virtually every subject in Camerer's experiment abided by her earlier decision when given the opportunity to change it. Hey and Lee (2005a, 2005b) compared the empirical fit of various preference specifications under reduction with the fit under isolation, and concluded in favor of isolation. All in all, these studies are supportive of the WRIS for simple binary choices.

In a pure between-subjects experimental design, each subject performs only one single task. In some cases, such a design is more desirable than a within-subjects design (Ariely et al. 2006; Greenwald 1976; Kahneman 2002; Keren and Raaijmakers 1988). When a RIS is employed in a between-subjects experiment, only a fraction of the subjects are paid for their task. Holt's (1986) concern about meta-lottery perception can also be raised for BRISs: biased risk preferences may similarly result if subjects integrate the choice problem they face with the RIS lottery. The BRIS is particularly susceptible to reduction, which here only involves a straightforward multiplication of the probabilities of the choice alternatives by the probability of real payment. Studies into the performance of the BRIS for risky choices are scarce. The only test we are aware of is in Cubitt et al. (1998b). Using a simple binary choice problem, they found a marginally significant difference, with lower risk aversion in the RIS treatment. Harrison et al. (2007, footnote 16) found no difference between a hybrid RIS and a WRIS for static risky choice.

A concern about WRISs is the possibility of carry-over effects from outcomes of prior tasks. With multiple tasks per subject, the outcomes of prior tasks may affect a subject's behavior in the current task in several ways. First, any experiment with multiple tasks and real payment of each task is vulnerable to an income effect. Outcomes of completed tasks accumulate and may distort subsequent choices (Cox and Epstein 1989). This effect may be limited when outcomes are revealed only at the end of the experiment. However, this is usually not possible. Many actual choice problems are dynamic, consisting of multiple sub-problems with intermediate decisions and intermediate outcomes. In this respect, the WRIS has a clear advantage over using rewards for every task. Because only one task is for real, there is no accumulation of payoffs, and thus no income effect. Grether and Plott (1979, p. 630), however, took the income effect one step further and argued that a subject's risk attitude may even be influenced by a pseudo-income effect from changes in the expected value of a subject's payment from the experiment.

Second, modern reference-dependent decision theories such as prospect theory suggest that the outcomes of prior tasks can also generate a reference-point effect in

https://doi.org/10.1007/s10683-011-9306-4 Published online by Cambridge University Press

a WRIS experiment. If subjects continue to think about results from a previous task, they may evaluate their current choice options relative to their previous winnings. That is, a previous task would set a reference point or anchor for the current task. A favorably completed prior task, for example, then places many outcomes in the domain of losses or makes many current possible outcomes seem relatively small, and consequently encourages risk taking.[1] Gächter et al. (2009) showed that even experimental economists can be subject to framing effects.

Third, in a design with multiple tasks and where all tasks are paid for real, a subject may also be more willing to gamble in order to compensate previously experienced losses or because she feels that she is playing with someone else's money after previous gains (Thaler and Johnson 1990; Kameda and Davis 1990). Under a WRIS, however, subjects know that only one task is for real. Logically, they would thus understand that gains or losses experienced in prior tasks cannot be lost, undone or enlarged in the current task. In this sense, the WRIS would avoid this kind of reference dependence. Still, even if subjects understand the separation, we cannot exclude that they carry over their experiences and that behavior is affected.

Fourth and last, carry-over effects may also result from misunderstanding of randomness. It is well documented in the literature that subjects' subjective perceptions of chance can be influenced by sequences or patterns of outcomes they observe (see, for example, Rabin 2002). Like the tendency of basketball spectators to overstate the degree to which players are streak shooters (Gilovich et al. 1985; Wardrop 1995; Aharoni and Sarig 2008), subjects in our WRIS treatment who avoided the elimination of large prizes in a previous game may be too confident about their chances of avoiding the elimination of large prizes in the current game.

Hardly any empirical research has been done on potential carry-over effects from outcomes of previous tasks in a RIS experiment. Only Lee (2008) partially touched upon this topic, and found that the WRIS avoided an income effect.

Our study examines the effects of RISs for risky choices. Various other studies analyzed RISs in other fields. Bolle (1990) reported that behavior under a BRIS is not different from that for real tasks in ultimatum games. Sefton (1992) found that a BRIS does affect behavior in dictator games. Armantier (2006) concluded that ultimatum game behavior under a hybrid RIS is similar to that under a WRIS. Stahl and Haruvy (2006) found that a hybrid RIS does lead to differences in dictator games.

Our study of RISs differs from previous studies in three respects. First, we examine whether outcomes from prior tasks in a WRIS experiment affect choice behavior in subsequent tasks. Second, we use a dynamic task that requires more mental effort than the choice problems of previous studies, allowing us to explore whether RISs increase decision errors. Prior analyses of the validity of RISs typically concerned static risky choice problems, in which each task requires the subjects to choose between two

[1] Behavioral research shows that people do indeed use relative judgments of size (comparing to other sizes encountered) rather than absolute values (see, for example, Ariely et al. 2003; Green et al. 1998; Johnson and Schkade 1989; Simonson and Drolet 2004; van den Assem et al. 2011). For meta-analyses of empirical work on the influence of gain/loss framing on risky decisions, see Kühberger (1998) and Kühberger et al. (1999). Kühberger et al. (2002) discuss how framing depends on incentives.

https://doi.org/10.1007/s10683-011-9306-4 Published online by Cambridge University Press

simple lotteries.[2] Finally, we also consider the BRIS. The validity of this design has hardly been investigated before.

2 The experiment

Our laboratory experiment mimics the choice problems in the TV game show *Deal or No Deal* (DOND); see Fig. 1. In every game round of DOND, a contestant has to choose repeatedly between a sure alternative and a risky lottery with known probabilities. DOND requires no special skills or knowledge. At the start of a game, the contestant chooses one (suit)case out of a total of 26 numbered cases, each hiding one out of 26 randomly distributed amounts of money. The content of the chosen case, which she then owns, remains unknown until the end of the game. Next, the first round starts and she has to select 6 of the other 25 cases to be opened, revealing their prizes, and revealing that these prizes are not in her own case. Then, the banker specifies a price for which he is willing to buy the contestant's case. If the contestant chooses "No Deal", she enters the second round and has to open 5 additional cases, followed by a new bank offer. The game continues this way until the contestant either accepts an offer ("Deal"), or rejects all offers and receives the contents of her own case. The maximum number of game rounds to be played is 9, and the number of cases to be opened in each round is 6, 5, 4, 3, 2, 1, 1, 1, and 1, reducing the number of remaining cases from 26 to 20, 15, 11, 8, 6, 5, 4, 3, and, finally, 2.

DOND has been aired in dozens of countries, sometimes under an alternative name. Each edition has its own set of prizes, and some also employ a different number of prizes and game rounds. The basic structure, however, is always the same.

Our experiment uses the 26 prizes of the original Dutch edition, scaled down by a factor of 10,000, with the lowest amounts rounded up to one cent. The resulting set of prizes is: €0.01 (9 times); €0.05; €0.10; €0.25; €0.50; €0.75; €1; €2.50; €5; €7.50; €10; €20; €30; €40; €50; €100; €250; €500. The distribution of prizes is clearly positively skewed, with a median of €0.63 and a mean of €39.14.

In the TV episodes, the bank offer starts from a small fraction of the average remaining prize in the early rounds, but approaches the average remaining prize in the last few rounds. Although the bank offers can be predicted accurately (Post et al. 2008), we eliminate any ambiguity by fixing the percentage bank offer for each game round and by including these fixed percentages in the instructions for the subjects. The percentages for round 1 to 9 are 15, 30, 45, 60, 70, 80, 90, 100, and 100, respectively. The resulting monetary offers are rounded to the nearest cent.

As discussed in the introduction, DOND has been used for many studies of risky choice behavior. There are several advantages of using this game for our purposes as well. First, it is a dynamic game. Because many experiments use dynamic choice

[2] Wilcox (1993) found that the probability that a task is selected to be for real is not important if choices concern simple, one-stage lotteries. However, he found that an increased probability did improve decisions in more complex two-stage lotteries that had exactly the same distributions as the one-stage lotteries. Apparently, the higher expected payoff per task encouraged subjects to spend more effort. Moffatt (2005) also confirmed that higher incentives generate an increase in efforts and that subjects need such motivation for complex tasks.

https://doi.org/10.1007/s10683-011-9306-4 Published online by Cambridge University Press

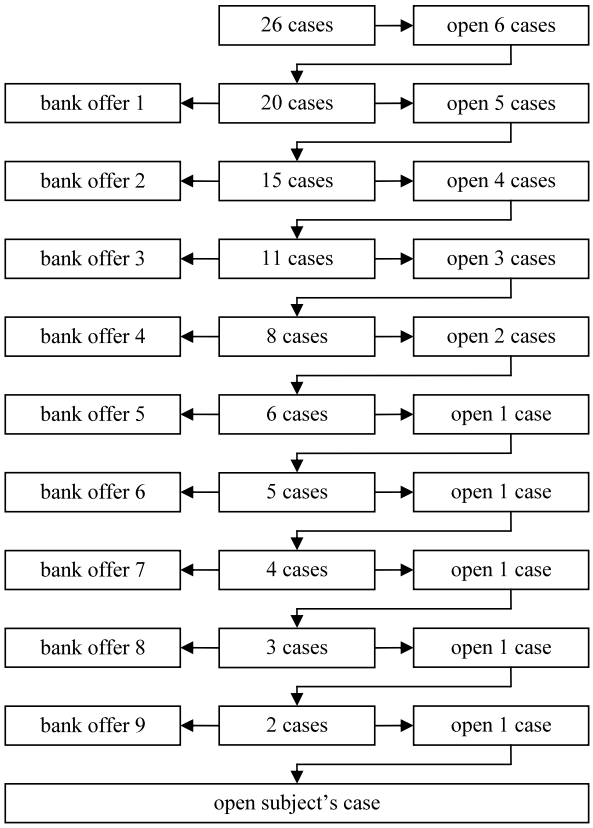


Fig. 1 Flow chart of the Deal or No Deal game. In each of a maximum of nine game rounds, the subject chooses a number of cases to be opened. After the monetary amounts in the chosen cases are revealed, a bank offer is presented. If the subject accepts the offer ("Deal"), she receives the amount offered and the game ends. If the subject rejects the offer ("No Deal"), play continues and she enters the next round. If the subject decides "No Deal" in the ninth round, she receives the amount in her own case. (Taken from Post et al. 2008)

tasks, it is desirable to test RISs in such an environment. Second, DOND requires a more than basic degree of mental effort, allowing us to explore whether RISs result in more error-driven choices. Third, although the choice problems are more difficult than conventional ones, the game itself is well understood by subjects. Most subjects are familiar with the game because of its great popularity on TV. A final advantage of DOND is that it promotes the involvement of subjects. A necessary condition for analyzing if outcomes of prior tasks influence current decisions is that subjects, at least broadly, remember what happened in prior games. Each game of DOND normally lasts for several rounds, and its length and dynamic nature increase the likelihood that subjects remember the course of their previous task and their experience. This is facilitated because the amounts not yet revealed in previous rounds are visually displayed throughout the game.

In our basic treatment, subjects play only once and for real. In the WRIS treatment, subjects play the game ten times, one of which for real payment. In the BRIS treatment, subjects play the game only once with a one-in-ten chance of real payment. The tasks used in the three treatments are identical: everything, including the face values of the prizes, is held constant. Each difference between the three treatments (randomized or guaranteed payment, expected reward per subject and per task, and

https://doi.org/10.1007/s10683-011-9306-4 Published online by Cambridge University Press

number of tasks per subject) is implied by the requirement of identical prizes and by the differences between the incentive systems.

We randomly selected first-year economics students at Erasmus University Rotterdam for each treatment. Students were not allowed to participate more than once. Given the random allocation of subjects and the homogeneous population of students, the groups are likely to be very similar.

A research assistant developed computer software that randomly distributes the prizes across the cases (independent draws for each game), displays the game situations, provides a user interface, and stores the game situations and choices of the subjects. All treatments of the experiment were conducted in a computerized laboratory and run in sessions with about 20 subjects sitting at computer terminals. They were separated by at least one unoccupied computer to provide a quiet environment and to prevent them from observing each other's choices. We did not impose any time constraints and informed subjects that we would finalize the experiment only after everyone had finished.

Before the experiment actually started, subjects were given ample time to read the instructions (we also allowed subjects to consult the instructions during the experiment), and to ask questions. We next explained the payment procedure. For the RIS treatments, we explained that they should make their choices as if the payment of outcomes were for sure. At the end of each session, the relevant payment procedure was implemented. In both RIS treatments, a ten-sided die was thrown individually by each subject to determine her payment. There was no show-up fee. Accordingly, subjects who were not selected for payment in the BRIS treatment earned nothing. The instructions for the WRIS treatment are available as a web appendix. The instructions for the other two treatments were similar, apart from the details about the number of games to be played and the incentive scheme.

3 Analyses and results

A total of 97 subjects participated in the basic treatment, 100 took part in the BRIS treatment, and 88 in the WRIS treatment. On average, subjects earned about €50, €5, and €38, respectively. In what follows, we first explore possible treatment effects by looking at simple risk aversion measures from the game (Sect. 3.1). We then present more rigorous probit regression analyses (Sect. 3.2). The last subsection examines our data using structural choice models (Sect. 3.3).

3.1 Preliminary analysis

A crude way to compare risk aversion across the three treatments is by analyzing the round in which subjects accept a bank offer ("Deal"). After all, as the game progresses the expected return from continuing play ("No Deal") decreases and the risk generally increases. Therefore, the longer a subject continues to play, the less risk averse she probably is. Another but similar way is to look at the difference between the bank offer and the average remaining prize in the last two rounds of a subject's game. The greater the discounts are at which subjects are indifferent between accepting the bank

https://doi.org/10.1007/s10683-011-9306-4 Published online by Cambridge University Press

offer and continuing play, the greater is their risk aversion. We define the certainty discount as the difference between the bank offer and the average remaining prize, expressed as a fraction of the average remaining prize, and take the average of the ratios for the ultimate and penultimate game round as our estimate of the certainty discount for a given game. The p-values in this subsection are based on two-sided t-tests. Non-parametric Wilcoxon rank-sum tests yield similar values and identical conclusions unless reported otherwise.

Table 1 reports summary statistics of both the stop round and the estimated certainty discount for the three treatments (Panel A), including separate statistics for each of the ten successive games in the WRIS treatment (Panel B). For subjects who rejected all offers and played the game until the end, we set the stop round and the certainty discount equal to 10 and 0, respectively. We excluded games that ended up with trivial choice problems involving prizes of one cent only. Figure 2 shows histograms of the stop round in the basic treatment, in the WRIS treatment, and in the BRIS treatment. To investigate the effects of the various RISs on the degree of risk aversion, we compare the two RIS treatments with the basic treatment.

Measured across the ten games in the WRIS treatment, the average subject accepts a bank offer ("Deal") in round 7.37, compared to 7.80 in the basic treatment. The difference is statistically insignificant ($p = 0.105$). The average certainty discounts in the two treatments are 16.6 and 13.5 percent. The difference is, again, insignificant ($p = 0.162$). Possibly, decisions in the last nine games of the WRIS treatment are influenced by carry-over effects from the outcomes of earlier games, or affected by more familiarity with the task or by boredom. If we drop the last nine games from the comparison and use the first game only, then the treatment differences are larger and marginally significant (stop round: $p = 0.058$; certainty discount: $p = 0.070$).[3]

The BRIS treatment, on the contrary, does yield significantly different values. Subjects' average stop round is 8.58, compared to 7.80 for the basic treatment ($p = 0.013$), and their average certainty discount is only 7.1 percent, about half the discount in the basic treatment ($p = 0.016$). Strikingly, nearly two-thirds of the subjects display risk-seeking behavior by rejecting actuarially fair bank offers. This first and crude analysis therefore suggests that employing a BRIS has a considerable effect on risk behavior.

The preceding analyses of the average stop round do not correct for the potential influence of errors. Errors are likely to have an asymmetric effect in our experiment, reducing the average stop round and increasing the average certainty discount: erroneous "Deal" decisions immediately end the game, whereas erroneous "No Deal" decisions may lead to only one extra round because the subject can stop the round after it. If errors are more likely to occur under a RIS, this may have biased the above comparisons.

Because payment of the outcome of a task is not sure under a RIS, a subject's expected reward for solving a decision problem is smaller than her reward in the case of guaranteed payment. For tasks with, for example, a one-in-ten chance of being selected, the expected reward is ten times smaller when the nominal stakes are held

[3] According to the Wilcoxon rank-sum test, the certainty discount for the first task is not significantly different from that for the basic treatment ($p = 0.237$).

https://doi.org/10.1007/s10683-011-9306-4 Published online by Cambridge University Press

Table 1 Summary statistics. The table shows summary statistics of the stop round and the estimated certainty discount for the various treatments (Panel A), separate statistics for the ten successive games in the WRIS treatment (Panel B), and separate statistics for games in the WRIS treatment subdivided on the basis of the outcome of a prior task (Panel C). The stop round is the round number in which the bank offer is accepted ("Deal"), or 10 for subjects who rejected all offers. The certainty discount is estimated as the average difference between the bank offer and the average remaining prize (scaled by the average remaining prize) for the ultimate and penultimate game round, or 0% for subjects who rejected all offers. In the basic treatment (Basic), subjects play the game once and for real. In the between-subjects RIS treatment (BRIS), subjects play the game once with a ten percent chance of real payment. In the within-subjects RIS treatment (WRIS), subjects play the game ten times with a random selection of one of the ten outcomes for real payment. Shown are the mean, the median, the standard deviation (Stdev), and the number of observations (No. obs.). The p-values refer to t-tests for the mean being equal to the mean of the basic treatment (Panel A and B), or to t-tests for equality of the means of the two subsamples divided on the basis of the outcome of a prior task (Panel C). EV_{-k} ($k = 1, 2, 3, 4$) is the average remaining prize in the last round of the game played k games before the current game

	Stop Round				Certainty Discount (%)				No. obs.
	Mean	Median	Stdev	p-value	Mean	Median	Stdev	p-value	
A. Overall									
Basic	7.80	8.00	2.29		13.5	5.0	19.5		88
WRIS	7.37	8.00	2.34	0.105	16.6	5.0	21.1	0.162	796
BRIS	8.58	9.00	1.85	0.013	7.1	0.0	15.9	0.016	92
B. WRIS conditional on task number									
Game 1	7.05	8.00	2.63	0.058	20.1	5.0	25.4	0.070	76
Game 2	7.33	8.00	2.50	0.208	17.7	5.0	22.6	0.200	82
Game 3	7.26	7.00	2.29	0.142	17.4	15.0	20.3	0.218	73
Game 4	7.36	8.00	2.37	0.231	16.8	5.0	21.8	0.301	80
Game 5	7.28	8.00	2.41	0.162	17.7	5.0	21.3	0.193	78
Game 6	7.31	8.00	2.28	0.165	16.7	5.0	20.7	0.303	84
Game 7	7.78	8.00	2.38	0.977	14.0	5.0	21.2	0.884	79
Game 8	7.71	8.00	2.14	0.808	13.3	5.0	18.1	0.942	80
Game 9	7.28	8.00	2.16	0.137	16.4	5.0	19.5	0.347	81
Game 10	7.35	8.00	2.28	0.203	16.5	5.0	20.0	0.323	83
C. WRIS conditional on outcome of prior task									
$EV_{-1} \leq 39.14$	7.23	8.00	2.30	0.031	17.5	5.0	20.8	0.121	370
$EV_{-1} > 39.14$	7.62	8.00	2.34		15.0	5.0	20.6		287
$EV_{-2} \leq 39.14$	7.33	8.00	2.28	0.111	16.7	5.0	20.7	0.210	323
$EV_{-2} > 39.14$	7.63	8.00	2.28		14.6	5.0	19.8		260
$EV_{-3} \leq 39.14$	7.44	8.00	2.28	0.963	15.8	5.0	20.7	0.732	284
$EV_{-3} > 39.14$	7.43	8.00	2.37		16.4	5.0	20.9		230
$EV_{-4} \leq 39.14$	7.44	8.00	2.32	0.701	15.9	5.0	20.8	0.814	236
$EV_{-4} > 39.14$	7.36	8.00	2.26		16.4	5.0	20.0		205

constant, whereas subjects' costs of discovering optimal choices are not affected. As a result, subjects might be less motivated to consider choice problems profoundly, leading to more choices that are driven by errors (Camerer and Hogarth 1999;

https://doi.org/10.1007/s10683-011-9306-4 Published online by Cambridge University Press

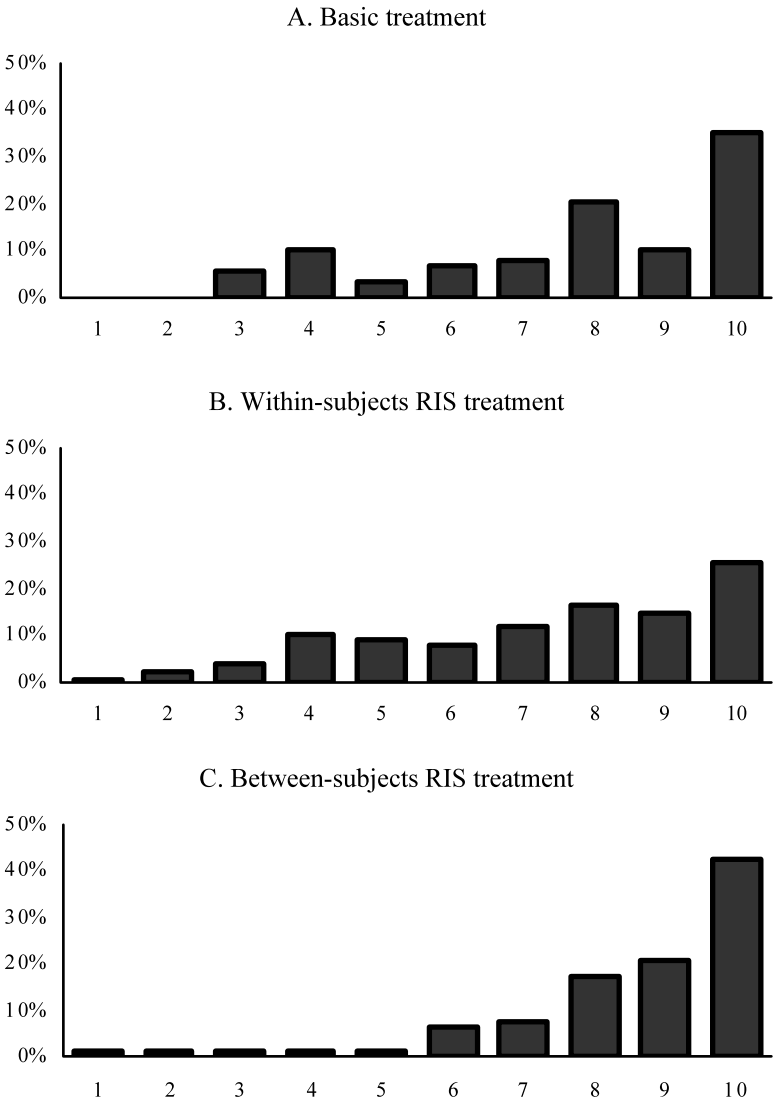


Fig. 2 Histograms of the stop round. The figure shows histograms of the stop round in the basic treatment (Panel **A**), in the WRIS treatment (Panel **B**) and in the BRIS treatment (Panel **C**). The stop round is the round number in which the bank offer is accepted ("Deal"), or 10 for subjects who rejected all offers. In the basic treatment, subjects play the game once and for real. In the WRIS treatment, subjects play the game ten times with a random selection of one of the ten outcomes for real payment. In the BRIS treatment, subjects play the game once with a ten percent chance of real payment

Smith and Walker 1993; Wilcox 1993). For the BRIS treatment, controlling for an asymmetric effect of errors would aggravate the difference with the basic treatment. For the WRIS treatment, however, doing so would reduce the difference.

The histograms (Fig. 2A–C) indeed suggest increased error rates in the RIS treatments. Interestingly, some decisions in the RIS treatments seem to have been made

https://doi.org/10.1007/s10683-011-9306-4 Published online by Cambridge University Press

without any regard to the attractiveness of the alternative choice option. For example, accepting a bank offer of 15 or 30 percent (in round 1 and 2, respectively) of the mean prize implies an implausible degree of risk aversion, assuming that this decision is really carefully considered. In the probit regression analyses of the next paragraph, we will analyze the likelihood and effect of increased errors more thoroughly.

To obtain a first indication of carry-over effects, or the dependence of subjects' choices on the outcomes of preceding tasks, we compare the stop round and the certainty discount for tasks preceded by relatively favorable outcomes with those preceded by relatively unfavorable outcomes. To classify the outcomes of prior tasks, we focus on the average remaining prize in the last game round (or the prize in the subject's own case if she rejected all bank offers), and use the statistical average of the prizes present at the start of each game (€39.14) to create a division into "good" and "bad" outcomes.

Panel C of Table 1 presents the results. When the previous task ended with stakes below the initial level, the average stop round is 7.23, whereas when it ended with larger stakes, the average stop round is 7.62. The difference is significant ($p = 0.031$). The difference for the penultimate prior task is of the same sign, but insignificant ($p = 0.111$). For the average certainty discount, the signs of the differences similarly suggest that subjects take more risk after a favorable outcome, but here both differences (2.5 and 2.1 percent) are insignificant.[4] There is no evidence of an effect from games played before the last two.

This approach of classifying subjects may also pick up differences in subjects' risk attitudes: more adventurous subjects play longer, and, due to the skewness of the prizes, they are more likely to end up with below-average stakes (expressed in terms of the average remaining prize). However, based on this argument, we would expect subjects with below-average prior outcomes to take more risk, i.e., play more game rounds, which is opposite to what we observe. Heterogeneity would thus strengthen our first indications that subjects take more risk after a recent favorable result.

To summarize, the preliminary analysis suggests that the BRIS generates lower risk aversion. The average degree of risk aversion under a WRIS is roughly unaffected, but differences in risk aversion are related to the outcomes of recent preceding tasks. Visual inspection suggests that both RISs increase error rates.

3.2 Probit regression analysis

The preceding analyses were crude: they only used subjects' last choices and did not control for differences between the various choice problems such as differences in the amounts at stake. This subsection uses probit regression analysis to analyze the different treatments while controlling for the characteristics of the choice problems.

The dependent variable is the subject's decision, with a value of 1 for "Deal" and 0 for "No Deal". We explain the various choices with the following set of variables:

[4] The stop round difference for the penultimate task is marginally significant when we use a Wilcoxon rank-sum test ($p = 0.068$). Wilcoxon rank-sum tests also point at a significant ($p = 0.039$) and a marginally significant ($p = 0.084$) certainty discount difference for the previous and penultimate task, respectively.

https://doi.org/10.1007/s10683-011-9306-4 Published online by Cambridge University Press

- D^{WS}: dummy variable indicating that the choice is made in the WRIS treatment ($1 = \text{WRIS}$);
- D^{BS}: dummy variable indicating that the choice is made in the BRIS treatment ($1 = \text{BRIS}$);
- $EV/100$: stakes, measured as the current average remaining prize divided by 100;
- EV/BO: expected relative return ($+1$) from rejecting the current and subsequent bank offers, or the average remaining prize divided by the bank offer;
- $Stdev/EV$: standard deviation ratio, or the standard deviation of the distribution of the average remaining prize in the next round divided by the average remaining prize.

The dummy variables D^{WS} and D^{BS} measure the effects of the different treatments. The stakes are divided by 100 to obtain more convenient regression coefficients. To control for the attractiveness of the bank offer, we use the expected return from rejecting the current and subsequent bank offers. The standard deviation ratio measures the risk of continuing to play ("No Deal") for one additional round. We did not include the common demographic variables *Age* and *Gender*. Subjects in our sample are all first-year Economics students of about the same age, and *Gender* does not have significant explanatory power. To allow for the possibility that the errors of individual subjects are correlated, we perform a cluster correction on the standard errors (Wooldridge 2003). For the WRIS treatment, we use subjects' first game only in order to avoid confounding effects of outcomes of prior tasks. We also exclude the trivial choices from games that ended up with prizes of one cent only (nontrivial choices from such games are not removed). The sample used for the regression analyses consists of 1977 choice observations, of which 677 are from the basic treatment, 766 from the BRIS treatment, and 534 from the WRIS treatment.

The first column of Table 2 shows the probit estimation results. As expected under non-satiability and risk aversion, the "Deal" propensity decreases with the expected return from continuing play and increases with the dispersion of the outcomes. The "Deal" propensity also increases with the stakes. The signs of the treatment effects are similar to those in the previous subsection. The WRIS dummy is significantly larger than zero ($p = 0.011$), indicating a higher deal propensity in the WRIS treatment than in the basic treatment. The BRIS dummy is negative, but statistically only marginally significant ($p = 0.094$).

This analysis does, however, not account for the possible effects of tremble, or the subject losing concentration and choosing completely at random. Tremble can explain "Deal" decisions in the early game rounds, when the bank offers are very conservative and "Deals" cannot reasonably be accounted for by risk aversion or errors in weighing the attractiveness of "Deal" and "No Deal" against each other. The relatively large number of early "Deals" suggests that tremble is indeed relevant in the RIS treatments.

To account for tremble, we extend our probit model by adding a fixed tremble probability (Harless and Camerer 1994). The standard model $P(y_i = 1) = \Phi(x_i'\beta)$ then becomes $P(y_i = 1) = (1-\omega)\Phi(x_i'\beta) + 0.5\omega$, where ω is the tremble probability ($0 \leq \omega \leq 1$) and the other parameters have the usual meaning. In fact, a subject is now assumed to choose according to the standard model with probability $(1 - \omega)$ and at random with probability ω, or, put otherwise, the likelihood of each choice is

https://doi.org/10.1007/s10683-011-9306-4 Published online by Cambridge University Press

Table 2 Probit regression results: treatment effects. The table displays the results from the probit regression analyses of the DOND decisions in the three different treatments. In the basic treatment, subjects play the game once and for real. In the BRIS treatment, subjects play the game once with a ten percent chance of real payment. In the WRIS treatment, subjects play the game ten times with a random selection of one of the ten outcomes for real payment. The dependent variable is the subject's decision, with a value of 1 for "Deal" and 0 for "No Deal". *EV* is the current average remaining prize in Euros. *BO* is the bank offer. *Stdev* measures the standard deviation of the distribution of the average remaining prize in the next game round. D^{WS} (D^{BS}) is a dummy variable that takes a value of 1 for observations from the WRIS (BRIS) treatment. Tremble is the estimated probability that a choice is made at random. Apart from the maximum likelihood estimates for the regression coefficients, the table reports the log-likelihood (LL), the mean log-likelihood (MLL), McFadden's R-squared, and the number of observations (No. obs.). The p-values (within parentheses) for the regression coefficients are corrected for correlation between the responses of a given subject (subject-level cluster correction). The p-values for the tremble probabilities are based on likelihood ratio tests

	Model 1	Model 2
Constant	−2.284 (0.000)	−1.843 (0.000)
D^{WS}	0.288 (0.011)	0.230 (0.104)
D^{BS}	−0.184 (0.094)	−0.258 (0.030)
$EV/100$	0.753 (0.000)	0.767 (0.000)
EV/BO	−0.244 (0.002)	−0.496 (0.001)
$Stdev/EV$	1.964 (0.000)	1.801 (0.000)
Tremble:		
Constant		0.000 (0.500)
D^{WS}		0.027 (0.003)
D^{BS}		0.009 (0.007)
LL	−458.2	−453.4
MLL	−0.232	−0.229
McFadden R^2	0.213	0.221
No. obs.	1977	1977

now a weighted average of the standard likelihood and 0.5. For a further discussion about the inclusion of a tremble probability (or constant error) term in a probit model, see Moffatt and Peters (2001). We allow for different tremble probabilities in our three treatments by modeling the tremble probability as $\omega = \omega_0 + \omega_1\ D^{WS} + \omega_2 D^{BS}$. The constant ω_0 represents the tremble probability in the basic treatment, and the parameters ω_1 and ω_2 represent the deviations of the tremble probabilities for the WRIS and BRIS treatment, respectively. Following the recommendation of Moffatt and Peters, we calculate the p-values for the tremble probabilities on the basis of likelihood ratio tests. Because the tremble parameter is restricted to be nonnegative, a test for tremble is one-sided, and the restricted p-value is obtained by dividing the unrestricted p-value by two.

The second column of Table 2 presents the new results. The tremble term in the basic treatment is virtually zero. In both RIS treatments, however, the tremble probability is nonzero: 2.7 percent ($p = 0.003$) in the WRIS treatment and 0.9 percent ($p = 0.007$) in the BRIS treatment. The difference between the two is significant ($p = 0.040$; not tabulated). Adding tremble improves the fit of the regression model

https://doi.org/10.1007/s10683-011-9306-4 Published online by Cambridge University Press

($LR = 9.622$; $p = 0.022$). Interestingly, after correcting for trembles, the WRIS dummy is no longer significantly different from zero ($p = 0.104$), indicating that we cannot reject the null hypothesis of unbiased risk aversion in this treatment. The BRIS dummy, on the other hand, now is significantly negative ($p = 0.030$), implying less risk aversion under the BRIS.[5]

We also estimated a model that allows for differences in the standard noise term between the three treatments, but found no significant improvement of the fit. The standard noise term represents errors in weighing the attractiveness of "Deal" and "No Deal" against each other, and particularly helps to understand decision errors that occur when the subject is almost indifferent between the two choices. It can explain for example why a subject stops in game round 6 or 8 when stopping in round 7 would be optimal for her. Tremble helps to understand why a moderately risk averse subject would sometimes stop in the early game rounds when the bank offers are still very conservative. For further discussions of the interpretation and modeling of tremble and other stochastic elements in risky choice experiments we refer to Harless and Camerer (1994), Hey (1995), Hey and Orme (1994), Loomes and Sugden (1995), Luce and Suppes (1965), and Wilcox (2008).

Carry-over effects As discussed in Sect. 1, intermediate or final outcomes of earlier tasks in a WRIS may influence a subject's choices in a given task. Our preliminary analyses of the stop round and the certainty discount indeed suggested such an effect. Here, we will perform a probit regression analysis on the WRIS data with variables that capture the outcomes of previous tasks.

To quantify a subject's winnings in the game played k games before the current game, we take the average remaining prize in the last game round (or the prize in the subject's own case if she rejected all bank offers), EV_{-k}, and we divide this variable by 100 to obtain convenient coefficients. We intentionally use the average remaining prize and not the actual winnings (the accepted bank offer), because the latter picks up heterogeneity in risk attitudes between subjects and would introduce a bias in the regression coefficients: a more prudent subject is more likely to say "Deal" in a given round of the current game, and, at the same time, is expected to have won smaller amounts in prior games because she is more inclined to say "Deal" in early game rounds (when the percentage bank offers are lower). We include the outcomes of the four most recent prior games, i.e., $k = 1, \ldots, 4$. Missing values are set equal to the sample average. (We also ran the regression on a smaller sample that excludes the observations with missing values for prior outcomes. The results are similar.)

The first column of Table 3 presents the probit regression results. The outcomes of the two most recent tasks, EV_{-1} and EV_{-2}, strongly influence the "Deal" propensity in the current task: the larger the prior winnings, the less a subject is inclined to accept the sure alternative, i.e., the more risk she takes. The other two lags have no effect.

As discussed before, using the average remaining prize to measure the outcome of a prior task (EV_{-k}) avoids the increasing trend that the accepted bank offer exhibits.

[5] The absolute values of both coefficients are roughly equal, suggesting opposite biases of similar strength. Statistical inference for the WRIS dummy seems to be affected by overlap in the effects of increasing the dummy coefficient and increasing the tremble probability. Both yield a decrease in the predicted stop round.

https://doi.org/10.1007/s10683-011-9306-4 Published online by Cambridge University Press

Table 3 Probit regression results: carry-over and tremble effects across tasks. The table displays the results from the probit regression analyses of the DOND decisions in the ten different games of the WRIS treatment. EV_{-k} $(k = 1, 2, 3, 4)$ is the average remaining prize in the last round of the game played k games before the current game. $Fortune_{-k}$ $(k = 1, 2, 3, 4)$ is the probability of an average remaining prize that is smaller than or equal to the actual average in the last game round of the game played k games before the current game. Missing values for both variables are set equal to the sample average. Models 1 and 2 assume a constant tremble probability across the different tasks, and Models 3 and 4 assume a log-linear pattern. Other definitions are as in Table 2

	Model 1	Model 2	Model 3	Model 4
Constant	−1.182 (0.000)	−0.366 (0.165)	−1.170 (0.000)	−0.377 (0.100)
$EV/100$	0.812 (0.000)	0.811 (0.000)	0.805 (0.000)	0.793 (0.000)
EV/BO	−0.613 (0.000)	−0.593 (0.000)	−0.616 (0.005)	−0.575 (0.000)
$Stdev/EV$	1.610 (0.000)	1.655 (0.000)	1.594 (0.000)	1.631 (0.000)
$EV_{-1}/100$	−0.204 (0.001)		−0.200 (0.001)	
$EV_{-2}/100$	−0.197 (0.001)		−0.193 (0.001)	
$EV_{-3}/100$	−0.052 (0.391)		−0.053 (0.385)	
$EV_{-4}/100$	−0.029 (0.683)		−0.025 (0.720)	
$Fortune_{-1}$		−0.583 (0.000)		−0.567 (0.000)
$Fortune_{-2}$		−0.605 (0.000)		−0.601 (0.000)
$Fortune_{-3}$		−0.250 (0.082)		−0.269 (0.067)
$Fortune_{-4}$		−0.291 (0.073)		−0.268 (0.092)
Tremble:				
Constant	0.011 (0.000)	0.009 (0.000)	0.028 (0.000)	0.026 (0.000)
Ln(*Task*)			−0.011 (0.049)	−0.011 (0.033)
LL	−1531.2	−1518.8	−1529.4	−1516.7
MLL	−0.266	−0.264	−0.266	−0.264
McFadden R^2	0.200	0.206	0.201	0.207
No. obs.	5756	5756	5756	5756

Still, one may wonder if this variable is completely exogenous from a statistical perspective. Although the expected value of EV_{-k} is the same for every game round, the variance and higher moments of EV_{-k} are not. For example, extreme values are more likely to occur in the final game rounds. As a robustness check, we therefore also run the regression with an alternative proxy variable. We define $Fortune_{-k}$ as the probability of an average remaining prize that is smaller than or equal to the actual average in the last game round of the game played k games before the current one.

For example, suppose that a subject has reached round 9 in the previous game ($k = 1$), with only €1 and €500 remaining. The number of combinations of 2 prizes from 26 is 325. Across these combinations, the average prize ranges from €0.01 (two cases with €0.01 remain) to €375 (€250 and €500 remain). If we rank all possible combinations by their average in ascending order, then our subject's average (€250.50) ranks 315. The probability of an average prize that is smaller than or equal to this value is therefore equal to 315/325, or $Fortune_{-1} = 0.969$.

For every game round, $Fortune_{-k}$ has the standard uniform distribution. From a statistical perspective, this variable seems to be more appealing than EV_{-k}, although

https://doi.org/10.1007/s10683-011-9306-4 Published online by Cambridge University Press

we may question whether the average subject is always able to reliably approximate the relevant value and whether the different values adequately express how subjects value or experience the different outcomes from an economic perspective. The second column of Table 3 shows that the results are robust for the use of this alternative proxy. Again, the outcomes of the two most recent tasks influence the "Deal" propensity: the higher the relative rank of the average remaining prize in the two most recent tasks are, the more risk subjects take in the current task ($p < 0.001$). Strikingly, using $Fortune_{-k}$ instead of EV_{-k} leads to marginal significance of the outcomes of the third and fourth preceding game.

The tremble probability is again significant in both analyses (1.1 and 0.9 percent, respectively). If reward dilution is an issue in a RIS design and indeed leads to an increased error rate, then this rate does not need to be constant across the different tasks. With every task, subjects gain experience in making decisions, and, with experience, decision costs decrease, resulting in fewer errors (Smith and Walker 1993). Subjects also become more familiar with the software and devices, which may further decrease the likelihood of errors. On the other hand, subjects in experiments with repeated tasks may become bored, resulting in reduced concentration and increased errors.

The two estimates for the tremble probability across the ten different tasks are clearly smaller than the tremble probability estimated for the first task only (2.7 percent, see Table 2). This suggests that the likelihood of trembles indeed decreases during the experiment. To further investigate this possibility, we decompose the tremble probability into a constant that represents the tremble in the first task, and a term that varies log-linearly with the task number, i.e., $\omega = \omega_0 + \omega_1 \log(Task)$. Column 3 of Table 3 shows the estimation results when EV_{-k} is used as a measure for prior outcomes. The log-linear term is negative ($p = 0.049$), confirming a decreasing pattern of the tremble probability. The constant (2.8 percent) is almost equal to the tremble probability estimated for the first task separately (2.7 percent). If we replace the log-linear term by a linear term, i.e., $\omega = \omega_0 + \omega_1 (Task)$, then the significance of the linear component decreases ($p = 0.068$; not tabulated), suggesting that the largest effect of gaining experience occurs during the first few tasks. Using $Fortune_{-k}$ as a measure for prior outcomes yields a similar decreasing pattern for the tremble parameter (Column 4).

3.3 Structural model approach

To obtain an impression of how the differences in choice behavior across treatments correspond to differences in risk parameters, this subsection presents the results of a structural modeling approach.[6] We implement two simple representations of expected utility of final wealth theory (EU) and prospect theory (PT), and estimate the size of treatment effects and carry-over effects on the risk aversion and loss aversion parameters.

A difficulty with structural choice model estimations are the many specification choices and the impact that these choices may have. Examples include the shape of

[6] Adding this analysis was recommended by one of the reviewers.

https://doi.org/10.1007/s10683-011-9306-4 Published online by Cambridge University Press

the utility function, the number of game rounds subjects are assumed to look ahead, the modeling of the error term (general distribution, dependence on choice problem difficulty), the dynamics of the reference point, probability weighting, and the treatment of potential outliers.

We will follow the methodology used in the earlier DOND-based study by Post et al. (2008).[7] Obviously, we also tried alternative specifications, and although the precise parameter estimates and their standard errors are indeed often affected, our findings that between-subjects randomization reduces risk aversion and that a WRIS entails carry-over effects turned out to be very robust. The next few paragraphs summarize our approach. For further methodological details, background and discussion we refer to Post et al.

For EU, we employ the following general, flexible-form expo-power utility function:

$$u(x) = \frac{1 - \exp(-\alpha(W + x)^{1-\beta})}{\alpha} \tag{1}$$

where α and β are the risk aversion coefficients and W is the initial wealth parameter (in Euros). A CRRA (constant relative risk aversion) power function arises as the limiting case of $\alpha \to 0$ and a CARA (constant absolute risk aversion) exponential function arises when $\beta = 0$. For our experimental data, the optimal expo-power utility function always reduces to a CARA exponential function (where W can take any value), leaving in fact just one unknown parameter: $u(x) = 1 - \exp(-\alpha x)$. For brevity, we therefore present treatment and carry-over effects for α only, but it should be kept in mind that we actually estimated a three-parameter function for EU.

For PT, we use a simple specification that can explain break-even and house-money effects (Thaler and Johnson 1990). The utility function (also called value function in the literature) is defined as follows:

$$v(x|RP) = \begin{cases} (x - RP)^{\alpha} & x > RP \\ -\lambda(RP - x)^{\alpha} & x \leq RP \end{cases} \tag{2}$$

where $\lambda > 0$ is the loss-aversion parameter, RP is the reference point that separates losses from gains, and $\alpha > 0$ measures the curvature of the value function. Because we take the same power for gains and losses, loss aversion is well defined (Wakker 2010, end of Sect. 9.6.1). The reference point in the current round r, RP_r, is determined by the current bank offer, B_r, and—to allow for partial adjustment or "stickiness"—by the relative increase in the average remaining prize $d_r^{(j)} = (\bar{x}_r - \bar{x}_j)/\bar{x}_r$ across the last two rounds, $d_r^{(r-2)}$, and across the entire game, $d_r^{(0)}$:

$$RP_r = \left(\theta_1 + \theta_2 d_r^{(r-2)} + \theta_3 d_r^{(0)}\right) B_r \tag{3}$$

[7] In fact, our methodology is a carbon copy of that in Post et al. (2008) with three exceptions: (i) we do not exclude the trivial decisions from the first game round, simply because we do not need to align different game formats, (ii) our percentage bank offers are fixed across rounds and do not need to be estimated, and (iii) we limit the influence of individual observations by trimming the likelihood of each decision at a minimum of one percent.

https://doi.org/10.1007/s10683-011-9306-4 Published online by Cambridge University Press

For $r = 1$ we set $d_r^{(r-2)} = 0$. Note that the reference point sticks to earlier values when $\theta_2 < 0$ or $\theta_3 < 0$. We ignore probability weighting and use the true probabilities as decision weights.

We estimate the unknown parameters using a maximum likelihood procedure, where the likelihood of each decision is based on the difference in utility between the current bank offer and the (expected) utility of the possible winnings from rejecting the offer, plus some normally distributed error. We assume that subjects look only one round ahead, implying that the possible winnings from continuing play are the possible bank offers in the next round. The error standard deviation is modeled as being proportional to the difficulty of a decision, where difficulty is measured as the standard deviation of the utility values of the possible winnings from continuing play. To reduce the potential weight of individual observations, we truncate the likelihood of each decision at a minimum of one percent. We also looked at an alternative specification that includes a fixed tremble probability rather than truncates the likelihood of individual observations, but found that this distorts the estimation results for our structural models. A few observations are near-violations of stochastic dominance, and their extremely low likelihood triggers relatively large tremble probabilities. Leaving out those few extreme observations or trimming the likelihood of observations makes the tremble term redundant, indicating that such large tremble probability estimations would be misrepresentations.

We use the dummy variables D^{WS} and D^{BS} to measure the effects of the different treatments on the risk aversion parameter α in our EU model and on the loss aversion parameter λ in our PT model. That is, we replace α by $\alpha_0 + \alpha_1 D^{WS} + \alpha_2 D^{BS}$ and λ by $\lambda_0 + \lambda_1 D^{WS} + \lambda_2 D^{BS}$ in (1) and (2), respectively. In a similar fashion, we incorporate EV_{-k} and $Fortune_{-k}$ to measure the effects of a subject's winnings in the game played k games before the current game.

The first column of Table 4 presents the treatment effects. Consistent with previous results, both the EU (Panel A) and the PT estimation results (Panel B) yield evidence that subjects take more risk under a BRIS. For EU, there is a marginally significant difference ($p = 0.091$) between the risk aversion coefficient α in our BRIS treatment ($\alpha = 0.0002$) and that in our basic treatment ($\alpha = 0.0016$). Remarkably, preferences under the BRIS are not different from risk neutrality ($p = 0.709$; not tabulated). Under the WRIS, the risk aversion parameter is not different from that in the basic treatment ($p = 0.203$).

For PT, the loss aversion parameter in the basic treatment is 1.55. The degree of loss aversion is lower in the BRIS and higher in the WRIS treatment: 1.26 ($p = 0.020$) and 1.96 ($p = 0.019$), respectively.[8] However, the increased loss aversion under a WRIS does not appear to be very robust, because it holds for the first of the ten games in the WRIS treatment only: for the choices from all ten games combined, the degree of loss aversion is similar to that in the basic treatment ($\lambda = 1.50$).

[8] θ_3 is smaller than zero ($p < 0.001$), indicating that subjects' reference point sticks to their expectations at the start of a game. Such stickiness was also found by Post et al. (2008), and can yield break-even and house-money effects, or a lower risk aversion after losses and after gains (Thaler and Johnson 1990). The economically and statistically insignificant value of θ_2 indicates that changes during the last few rounds have no distinct impact here.

https://doi.org/10.1007/s10683-011-9306-4 Published online by Cambridge University Press

Table 4 Structural model estimation results. The table shows the maximum likelihood estimation results for our expected utility theory (EU; Panel A) and prospect theory models (PT; Panel B). For EU, the optimal expo-power utility function always reduces to a CARA exponential function and Panel A therefore reports risk aversion parameter α only. For PT, Panel B shows the loss aversion (λ) and curvature (α) parameters of the value function and the three parameters of the reference point model θ_1, θ_2, and θ_3, The noise parameter is represented by σ. In both panels, the first column with parameters and cluster-corrected p-values (within parentheses) represents the estimation results across the three different treatments, where α_1 and α_2, and λ_1 and λ_2 measure treatment effects on the risk aversion and loss aversion parameters α and λ, respectively. The other two columns represent the estimation results across subjects' ten different games in the WRIS treatment, where $\alpha_3, \alpha_4, \ldots, \alpha_{10}$ and $\lambda_3, \lambda_4, \ldots, \lambda_{10}$ capture the impact of outcomes of prior tasks on α and λ, respectively. Other definitions are as in previous tables

	Model 1A & 1B: Treatment effects	Model 2A & 2B: Carry-over effects	Model 3A & 3B: Carry-over effects
A. Expected utility theory			
α_0	0.0016 (0.012)	0.0054 (0.000)	0.0151 (0.000)
$\alpha_1 \cdot D^{WS}$	0.0023 (0.203)		
$\alpha_2 \cdot D^{BS}$	−0.0014 (0.091)		
$\alpha_3 \cdot EV_{-1}/100$		−0.0022 (0.003)	
$\alpha_4 \cdot EV_{-2}/100$		−0.0026 (0.002)	
$\alpha_5 \cdot EV_{-3}/100$		0.0002 (0.796)	
$\alpha_6 \cdot EV_{-4}/100$		0.0000 (0.991)	
$\alpha_7 \cdot Fortune_{-1}$			−0.0056 (0.006)
$\alpha_8 \cdot Fortune_{-2}$			−0.0080 (0.000)
$\alpha_9 \cdot Fortune_{-3}$			−0.0019 (0.514)
$\alpha_{10} \cdot Fortune_{-4}$			−0.0032 (0.194)
σ	0.209 (0.000)	0.242 (0.000)	0.241 (0.000)
LL	−514.4	−1641.0	−1636.5
MLL	−0.260	−0.285	−0.284
B. Prospect theory			
λ_0	1.545 (0.000)	1.715 (0.000)	2.669 (0.000)
α	0.599 (0.000)	0.711 (0.000)	0.700 (0.000)
θ_1	1.030 (0.000)	1.056 (0.000)	1.055 (0.000)
θ_2	0.003 (0.629)	0.008 (0.540)	0.007 (0.469)
θ_3	−0.038 (0.000)	−0.086 (0.000)	−0.082 (0.000)
$\lambda_1 \cdot D^{WS}$	0.412 (0.019)		
$\lambda_2 \cdot D^{BS}$	−0.282 (0.020)		
$\lambda_3 \cdot EV_{-1}/100$		−0.248 (0.000)	
$\lambda_4 \cdot EV_{-2}/100$		−0.200 (0.003)	
$\lambda_5 \cdot EV_{-3}/100$		−0.062 (0.478)	
$\lambda_6 \cdot EV_{-4}/100$		−0.061 (0.520)	
$\lambda_7 \cdot Fortune_{-1}$			−0.656 (0.000)
$\lambda_8 \cdot Fortune_{-2}$			−0.636 (0.000)
$\lambda_9 \cdot Fortune_{-3}$			−0.258 (0.188)
$\lambda_{10} \cdot Fortune_{-4}$			−0.374 (0.100)
σ	0.269 (0.000)	0.314 (0.000)	0.320 (0.000)
LL	−429.6	−1491.7	−1484.9
MLL	−0.217	−0.259	−0.258
No. obs.	1977	5756	5756

https://doi.org/10.1007/s10683-011-9306-4 Published online by Cambridge University Press

The weak evidence in our EU analysis for lower risk aversion under a BRIS seems to be related to the inadequacy of the EU model to explain the data. The PT model explains subjects' choices substantially better, as appears from the large difference in the overall log-likelihood (−515.4 vs. −429.6). The difference in the number of free parameters puts the EU model at a disadvantage, but the empirical fit of PT is clearly superior, also after controlling for this difference.

The second and third columns of Table 4 show the impact of outcomes of prior tasks on the risk parameters. As in the probit regression results, the two most recent tasks strongly influence risk taking in the current task: the larger the prior winnings, the smaller both the EU risk aversion parameter α and the PT loss aversion parameter λ (with p-values ranging from 0.000 to 0.003 for EV_{-k}). To illustrate the marginal effects: each additional €10 in expected value at the end of the previous (penultimate) task decreases α by $2.2 \cdot 10^{-4}$ ($2.6 \cdot 10^{-4}$) and λ by 0.025 (0.020). Likewise, outcomes of the third and fourth preceding game have no significant impact on preferences. Replacing the expected value proxy variable by the alternative proxy for outcomes of prior tasks ($Fortune_{-k}$; third column) leads to similar results.

4 Summary of our results and discussion

The general degree of risk aversion in our WRIS treatment is roughly similar to that in our basic treatment. This result is consistent with most earlier investigations of the WRIS. There is a weak sign of increased risk aversion in our data only for subjects' first task. This is probably attributable to sampling error because no such deviation is observed for the other nine tasks. In contrast, the BRIS treatment yields a bias towards less risk aversion. This result may be explained by a reduction of compound lotteries. In a between-subjects design with a RIS, integration of the choice problem and the RIS lottery is relatively easy. Only one single choice problem is involved, and reduction requires just straightforward multiplication of the probabilities in the choice problem with the probability of payout. With a payout probability of 10 percent, a 50 percent chance at €100 then becomes a 5 percent chance at €100. Because small probabilities tend to be overweighted, reduction encourages risk taking in BRIS designs. Reduction is more complex under a WRIS, due to the many tasks and possible outcomes involved. It is virtually impossible for an average subject to solve the complex problem of reduction in a few seconds. Hence, subjects will rather process each task in isolation.

Subjects in our WRIS treatment are clearly influenced by the outcomes of prior tasks: a substantial part of the variation in risk attitudes that we observed across subjects and games can be explained by the outcomes of the two preceding games. The larger a subject's winnings in those games, the lower her risk aversion is. This finding is robust with respect to the use of an instrument variable that avoids any conceivable effect of subject heterogeneity. Possible explanations are that subjects evaluate outcomes from their current game relative to those from previous games and take more risk when current prospects are perceived as losses or relatively small amounts, or that they misunderstand randomness and become overly optimistic about their chances after a game went well. The cross-task contamination does not seem to reflect a pseudo-income effect from changes in the expected value of a subject's payment from the

https://doi.org/10.1007/s10683-011-9306-4 Published online by Cambridge University Press

experiment, because an income effect cannot explain why only the last two games affect decisions. A subject's expected income from the experiment is determined by the outcomes of every game.

Carry-over effects are not necessarily an important drawback for average results in a large sample if the effects of favorable and unfavorable prior outcomes cancel out. In our experiment, we indeed cannot reject the hypothesis that the average degree of risk aversion is stable and unbiased across the ten repetitions of the task. With static choice problems, carry-over effects from outcomes of recently performed tasks can easily be avoided by simply postponing the presentation of outcomes until all tasks have been completed. Intermediate outcomes are, however, inevitable for dynamic choice problems. Subjects in our DOND experiment, for example, need to know which prizes are eliminated and which prizes still remain at every stage of the game. Our results indicate that the carry-over effects from prior tasks are short-lived. When analyzing dynamic risky choice in a within-subjects design, carry-over effects could therefore be minimized by interposing dummy tasks between the tasks of interest.

The distributions of subjects' stop round and certainty discount, and the results of our probit analyses suggest that the likelihood of random choice is larger under a RIS than in a design where every task is paid. A smaller expected reward per task may discourage a subject's mental efforts and may lead to increased lapses of concentration. Of course, trembles in RIS treatments need not be a consequence of RISs *per se*, and they might also emerge if we were to decimate the prizes in our basic treatment. Most prior investigations of RISs used simple choice problems that require little mental effort, and unintentionally they may thus have avoided effects of reward dilution. Our results suggest that experimenters considering a RIS should be aware that RISs may increase error rates, and that it may be worthwhile to make sure that the expected payoffs after dilution from the RIS sufficiently counterbalance subjects' cognitive efforts. Decision errors can distort results, especially if they do not cancel out in a large sample. For example, if the least risky alternative is optimal for most subjects, then errors will decrease observed risk aversion. Harrison (1989, 1994) raised concerns that low rewards may even have resulted in inaccurate inferences about the (in)validity of expected utility theory.

We chose not to include a treatment with a hybrid RIS (with random selection of both subject and task). Given the requirement of identical face values of prizes, the expected rewards would be very small and too far off from our basic treatment. In our experiment, the selection probabilities are 0.1 (between subjects as well as within subjects). Then, for each task, the probability of real payment is 0.01 and the expected reward is below €0.40. Such incentives do not satisfy Smith's (1982) saliency requirement, and subjects would not take them seriously. The study of a hybrid RIS, therefore, requires a different design with either larger face values or larger selection probabilities in every treatment. We leave this as a topic for future research. Our study does, however, shed indirect light on the hybrid design. The two randomization factors combined in a hybrid RIS are analyzed separately in our WRIS and BRIS treatments. The hybrid design can be expected to display a combination of the effects that we found there, with carry-over effects as under the WRIS, lower risk aversion as under the BRIS, and a higher error rate.

https://doi.org/10.1007/s10683-011-9306-4 Published online by Cambridge University Press

Future studies may also investigate whether higher error rates in a RIS result from lower expected payoffs alone or also from the layer of complexity introduced by the RIS. New designs could be developed that test for trembles in a more explicit way, for example by using choice tasks where one option is dominated by another option.

5 Conclusions

We compared two random incentive systems (RISs) with another, basic design. The latter, serving as our gold standard, had only one task per subject and the outcome of the task was always paid out. Our experiment considered dynamic risky choices, which are more complex and more realistic than the simple choice tasks usually considered. We find that the within-subjects RIS (WRIS; random selection of one task per subject for real payment) entails carry-over effects from outcomes of previous tasks. Risk aversion increases after unfavorable recent results and decreases after favorable recent results. On average, however, risk aversion in the WRIS is similar to that in the basic treatment. Concerns about the WRIS may be subordinate to the opportunity to perform within-subjects analyses and to obtain a large number of observations. Researchers using this method for static choice experiments can easily avoid carry over-effects by playing out the tasks at the end of the experiment, after subjects have made all their choices. Our between-subjects RIS (BRIS; random selection of subjects for real payment) requires each subject to complete just one single task, thus avoiding carry-over effects. However, the BRIS decreases risk aversion. This bias can be explained by the relative ease of reduction together with the well-documented violations of independence. Finally, RISs may increase error rates. Thus, whereas RISs have not been found to generate biases in simple choice tasks, in more complex and more realistic dynamic choice tasks we do find biases. They call for cautious implementations and sufficiently salient payoffs.

Acknowledgements We thank conference participants at FUR 2008 Barcelona, Dennie van Dolder and the anonymous reviewers for useful comments and suggestions, and Nick de Heer for his skillful research assistance. We gratefully acknowledge support from the Erasmus Research Institute of Management and the Tinbergen Institute.

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

References

Aharoni, G., & Sarig, O. H. (2008). *Hot hands in basketball and equilibrium* (Working paper). Available at http://ssrn.com/abstract=883324.

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'École Américaine. *Econometrica*, *21*(4), 503–546.

Ariely, D., Loewenstein, G., & Prelec, D. (2003). Coherent arbitrariness: stable demand curves without stable preferences. *The Quarterly Journal of Economics*, *118*(1), 73–105.

Ariely, D., Loewenstein, G., & Prelec, D. (2006). Tom Sawyer and the construction of value. *Journal of Economic Behavior & Organization*, *60*(1), 1–10.

https://doi.org/10.1007/s10683-011-9306-4 Published online by Cambridge University Press

Armantier, O. (2006). Do wealth differences affect fairness considerations? *International Economic Review*, *47*(2), 391–429.
Beattie, J., & Loomes, G. (1997). The impact of incentives upon risky choice experiments. *Journal of Risk and Uncertainty*, *14*(2), 155–168.
Bettinger, E., & Slonim, R. (2007). Patience among children. *Journal of Public Economics*, *91*(1–2), 343–363.
Blavatskyy, P., & Pogrebna, G. (2008). Risk aversion when gains are likely and unlikely: evidence from a natural experiment with large stakes. *Theory and Decision*, *64*(2–3), 395–420.
Blavatskyy, P., & Pogrebna, G. (2010a). Models of stochastic choice and decision theories: why both are important for analyzing decisions. *Journal of Applied Econometrics*, *25*(6), 963–986.
Blavatskyy, P., & Pogrebna, G. (2010b). Endowment effects? 'Even' with half a million on the table! *Theory and Decision*, *68*(1–2), 173–192.
Bolle, F. (1990). High reward experiments without high expenditure for the experimenter. *Journal of Economic Psychology*, *11*(2), 157–167.
Brooks, R. D., Faff, R. W., Mulino, D., & Scheelings, R. (2009a). Deal or no deal, that is the question: the impact of increasing stakes and framing effects on decision-making under risk. *International Review of Finance*, *9*(1–2), 27–50.
Brooks, R. D., Faff, R. W., Mulino, D., & Scheelings, R. (2009b). Does risk aversion vary with decision-frame? An empirical test using recent game show data. *Review of Behavioral Finance*, *1*(1–2), 44–61.
Camerer, C. F. (1989). An experimental test of several generalized utility theories. *Journal of Risk and Uncertainty*, *2*(1), 61–104.
Camerer, C. F., & Ho, T.-H. (1994). Violations of the betweenness axiom and nonlinearity in probability. *Journal of Risk and Uncertainty*, *8*(2), 167–196.
Camerer, C. F., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: a review and capital-labor-production framework. *Journal of Risk and Uncertainty*, *19*(1–3), 7–42.
Carlin, P. S. (1992). Violations of the reduction and independence axioms in Allais-type and common ratio effect experiments. *Journal of Economic Behavior & Organization*, *19*(2), 213–235.
Cho, Y., & Luce, R. D. (1995). Tests of hypotheses about certainty equivalents and joint receipt of gambles. *Organizational Behavior and Human Decision Processes*, *64*(3), 229–248.
Cohen, M., Jaffray, J.-Y., & Said, T. (1987). Experimental comparison of individual behavior under risk and under uncertainty for gains and for losses. *Organizational Behavior and Human Decision Processes*, *39*(1), 1–22.
Cox, J. C., & Epstein, S. (1989). Preference reversals without the independence axiom. *American Economic Review*, *79*(3), 408–426.
Cubitt, R. P., Starmer, C., & Sugden, R. (1998a). On the validity of the random lottery incentive system. *Experimental Economics*, *1*(2), 115–131.
Cubitt, R. P., Starmer, C., & Sugden, R. (1998b). Dynamic choice and the common ratio effect: an experimental investigation. *Economic Journal*, *108*(450), 1362–1380.
Deck, C. A., Lee, J., & Reyes, J. A. (2008). Risk attitudes in large stake gambles: evidence from a game show. *Applied Economics*, *40*(1), 41–52.
Gächter, S., Orzen, H., Renner, E., & Starmer, C. (2009). Are experimental economists prone to framing effects? A natural field experiment. *Journal of Economic Behavior & Organization*, *70*(3), 443–446.
Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: on the misperception of random sequences. *Cognitive Psychology*, *17*(3), 295–314.
Green, D., Jacowitz, K. E., Kahneman, D., & McFadden, D. (1998). Referendum contingent valuation, anchoring, and willingness to pay for public goods. *Resource and Energy Economics*, *20*(2), 85–116.
Greenwald, A. G. (1976). Within-subjects designs: to use or not to use? *Psychological Bulletin*, *83*(2), 314–320.
Grether, D. M., & Plott, C. R. (1979). Economic theory of choice and the preference reversal phenomenon. *American Economic Review*, *69*(4), 623–638.
Harless, D. W., & Camerer, C. F. (1994). The predictive utility of generalized expected utility theories. *Econometrica*, *62*(6), 1251–1289.
Harrison, G. W. (1989). Theory and misbehavior of first-price auctions. *American Economic Review*, *79*(4), 749–762.
Harrison, G. W. (1994). Expected utility theory and the experimentalists. *Empirical Economics*, *19*(2), 223–253.
Harrison, G. W., Lau, M. I., & Rutström, E. E. (2007). Estimating risk attitudes in Denmark: a field experiment. *The Scandinavian Journal of Economics*, *109*(2), 341–368.

https://doi.org/10.1007/s10683-011-9306-4 Published online by Cambridge University Press

Hey, J. D. (1995). Experimental investigations of errors in decision making under risk. *European Economic Review*, *39*(3–4), 633–640.
Hey, J. D., & Lee, J. (2005a). Do subjects remember the past? *Applied Economics*, *37*(1), 9–18.
Hey, J. D., & Lee, J. (2005b). Do subjects separate (or are they sophisticated)? *Experimental Economics*, *8*(3), 233–265.
Hey, J. D., & Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica*, *62*(6), 1291–1326.
Holt, C. A. (1986). Preference reversals and the independence axiom. *American Economic Review*, *76*(3), 508–515.
Johnson, E. J., & Schkade, D. A. (1989). Bias in utility assessments: further evidence and explanations. *Management Science*, *35*(4), 406–424.
Kahneman, D. (2002). *Maps of bounded rationality: a perspective on intuitive judgment and choice*. Prize lecture for the Nobel Foundation, 8 December, 2002, Stockholm, Sweden.
Kameda, T., & Davis, J. H. (1990). The function of the reference point in individual and group risk decision making. *Organizational Behavior and Human Decision Processes*, *46*(1), 55–76.
Keren, G. B., & Raaijmakers, J. G. W. (1988). On between-subjects versus within-subjects comparisons in testing utility theory. *Organizational Behavior and Human Decision Processes*, *41*(2), 233–247.
Kühberger, A. (1998). The influence of framing on risky decisions: a meta-analysis. *Organizational Behavior and Human Decision Processes*, *75*(1), 23–55.
Kühberger, A., Schulte-Mecklenbeck, M., & Perner, J. (1999). The effects of framing, reflection, probability, and payoff on risk preference in choice tasks. *Organizational Behavior and Human Decision Processes*, *78*(3), 204–231.
Kühberger, A., Schulte-Mecklenbeck, M., & Perner, J. (2002). Framing decisions: hypothetical and real. *Organizational Behavior and Human Decision Processes*, *89*(2), 1162–1175.
Langer, T., & Weber, M. (2008). Does commitment or feedback influence myopic loss aversion? An experimental analysis. *Journal of Economic Behavior & Organization*, *67*(3–4), 810–819.
Lee, J. (2008). The effect of the background risk in a simple chance improving decision model. *Journal of Risk and Uncertainty*, *36*(1), 19–41.
Loomes, G., & Sugden, R. (1995). Incorporating a stochastic element into decision theories. *European Economic Review*, *39*(3–4), 641–648.
Luce, R. D., & Suppes, P. (1965). Preference, utility, and subjective probability. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 3, pp. 249–410). New York: Wiley.
Moffatt, P. G. (2005). Stochastic choice and the allocation of cognitive effort. *Experimental Economics*, *8*(4), 369–388.
Moffatt, P. G., & Peters, S. A. (2001). Testing for the presence of a tremble in economic experiments. *Experimental Economics*, *4*(3), 221–228.
Myagkov, M. G., & Plott, C. R. (1997). Exchange economies and loss exposure: experiments exploring prospect theory and competitive equilibria in market environments. *American Economic Review*, *87*(5), 801–828.
Post, G.T., van den Assem, M. J., Baltussen, G., & Thaler, R. H. (2008). Deal or no deal? Decision making under risk in a large-payoff game show. *American Economic Review*, *98*(1), 38–71.
Rabin, M. (2002). Inference by believers in the law of small numbers. *The Quarterly Journal of Economics*, *117*(3), 775–816.
Reilly, R. J. (1982). Preference reversal: further evidence and some suggested modifications in experimental design. *American Economic Review*, *72*(3), 576–584.
Rosett, R. N. (1971). Weak experimental verification of the expected utility hypothesis. *Review of Economic Studies*, *38*(4), 481–492.
Savage, L. J. (1954). *The foundations of statistics*. New York: Wiley (Second revised edition, 1972, New York: Dover Publications).
Schunk, D., & Betsch, C. (2006). Explaining heterogeneity in utility functions by individual differences in decision modes. *Journal of Economic Psychology*, *27*(3), 386–401.
Sefton, M. (1992). Incentives in simple bargaining games. *Journal of Economic Psychology*, *13*(2), 263–276.
Simonson, I., & Drolet, A. (2004). Anchoring effects on consumers' willingness-to-pay and willingness-to-accept. *Journal of Consumer Research*, *31*(3), 681–690.
Smith, V. L. (1976). Experimental economics: induced value theory. *American Economic Review*, *66*(2), 274–279.

https://doi.org/10.1007/s10683-011-9306-4 Published online by Cambridge University Press

Smith, V. L. (1982). Microeconomic systems as an experimental science. *American Economic Review*, *72*(5), 923–955.
Smith, V. L., & Walker, J. M. (1993). Monetary rewards and decision costs in experimental economics. *Economic Inquiry*, *31*(2), 245–261.
Stahl, D. O., & Haruvy, E. (2006). Other-regarding preferences: egalitarian warm glow, empathy, and group size. *Journal of Economic Behavior & Organization*, *61*(1), 20–41.
Starmer, C. (2000). Developments in non-expected utility theory: the hunt for a descriptive theory of choice under risk. *Journal of Economic Literature*, *38*(2), 332–382.
Starmer, C., & Sugden, R. (1991). Does the random-lottery incentive system elicit true preferences? An experimental investigation. *American Economic Review*, *81*(4), 971–978.
Thaler, R. H., & Johnson, E. J. (1990). Gambling with the house money and trying to break even: the effects of prior outcomes on risky choice. *Management Science*, *36*(6), 643–660.
Tversky, A. (1967a). Additivity, utility, and subjective probability. *Journal of Mathematical Psychology*, *4*(2), 175–201.
Tversky, A. (1967b). Utility theory and additivity analysis of risky choices. *Journal of Experimental Psychology*, *75*(1), 27–36.
Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458.
van den Assem, M. J., van Dolder, D., & Thaler, R. H. (2011). Split or steal? Cooperative behavior when the stakes are large. *Management Science*. doi:10.1287/MNSC.1110.1413.
Wakker, P. P. (2010). *Prospect theory: for risk and ambiguity*. Cambridge: Cambridge University Press.
Wardrop, R. L. (1995). Simpson's paradox and the hot hand in basketball. *The American Statistician*, *49*(1), 24–28.
Wilcox, N. T. (1993). Lottery choice: incentives, complexity and decision time. *Economic Journal*, *103*(421), 1397–1417.
Wilcox, N. T. (2008). Stochastic models for binary discrete choice under risk: a critical primer and econometric comparison. In J. C. Cox & G. W. Harrison (Eds.), *Research in experimental economics: Vol. 12. Risk aversion in experiments* (pp. 197–292). Bingley: Emerald.
Wooldridge, J. M. (2003). Cluster-sample methods in applied econometrics. *American Economic Review*, *93*(2), 133–138.
Yaari, M. E. (1965). Convexity in the theory of choice under risk. *The Quarterly Journal of Economics*, *79*(2), 278–290.

https://doi.org/10.1007/s10683-011-9306-4 Published online by Cambridge University Press