Experimental Economics, 1:115–131 (1998)

On the Validity of the Random Lottery Incentive System

ROBIN P. CUBITT, CHRIS STARMER AND ROBERT SUGDEN
School of Economic and Social Studies, University of East Anglia, Norwich NR4 7TJ, England

Abstract

The random lottery incentive system is widely used in experimental economics to motivate subjects. This paper investigates its validity. It reports three experiments which compare responses given to decision tasks which are embedded in random lottery designs with responses in 'single choice' designs in which each subject faces just one task for real. The experiments were designed to detect cross-task contamination effects in the random lottery treatment. No significant differences between treatments, and no significant contamination effects, were found. Over the three experiments, observed differences between the treatments are adequately explained as sampling variation.

Keywords: random lottery incentive system

JEL Classification: C90, C91

Economists generally maintain that, in order for experimental investigations of decision-making to generate reliable data, subjects should be given incentives. A widely used incentive mechanism for such experiments is the *random lottery* design. In an experiment with this design—we will call any such experiment a *random lottery experiment*—each subject faces a number of tasks. Depending on the experiment, each task might be a choice between consumption opportunities, between gambles, or between strategies in a game against other subjects; or it might be a valuation task; or participation in an auction; or some other form of decision problem. Whatever the individual task, there is some nominal reward structure, so that if the task were for real, the subject's response (often in interaction with other subjects' responses, or with some random device) would determine his reward. At the start of the experiment, the subject is told that, after all the tasks have been completed, one of them will be selected at random (this selection process is the 'random lottery'). The nominal reward for that task will then be the subject's payoff for the entire experiment.[1]

This incentive system has proved useful because experimenters often want each of their subjects to perform several tasks. In some cases, where within-subject comparisons are required, each subject's performing more than one task is an essential part of the experimental design. In other cases, it is a way of economizing on the substantial fixed costs of recruiting subjects. By using the random lottery system, it is possible to give an incentive for each task while avoiding the problem that if all tasks were for real, a subject's endowment for one task would vary according to the payment she received from other tasks. In this paper, we investigate whether the random lottery design generates reliable data. We focus on a potential problem first pointed out by Charles Holt (1986).

Holt gives reasons for expecting that, in a random lottery experiment, the attitude to risk that a subject reveals in one task may be affected by the other tasks in the experiment; and he shows how, in a particular class of experiments, such a ‘contamination effect’ could systematically generate spurious results. This effect has the potential to distort the results of any experiment that uses the random lottery design and in which the tasks involve decision-making under risk or uncertainty. For example, it could affect a game or auction task in which a subject’s payoff depends on the actions of other subjects. Clearly, however, the most controlled experimental environment in which to test for contamination effects is one in which individuals make non-interactive choices between risky options. We shall investigate whether, for decision problems of this kind, the random lottery design elicits individuals’ true preferences.

To pose this question in a way that can be tested, we need an operational definition of ‘true’ preferences. We define true preferences with respect to a given task as those that would be elicited by a *single choice* experimental design in which each subject faces only that task, and knows it to be for real. We investigate whether the responses to individual tasks in random lottery experiments are systematically different from true preferences. If such systematic differences exist, we shall say that the random lottery design is *biased*.

To date, very few decision-making experiments with single choice designs have been reported in the literature. This paper reports three experiments, with a total of over 550 subjects, which compare responses to random lottery and single choice treatments. These experiments continue a programme of research which began with an experiment reported by Starmer and Sugden (1991). To our knowledge, the only comparable research is that of Beattie and Loomes (1997), which also compares random lottery and single choice treatments, but does not specifically investigate contamination effects. We discuss Beattie and Loomes’s work in Section 6.

1. Theory: isolation, reduction and contamination

Consider a random lottery experiment in which a given subject faces N decision tasks; each task requires a choice between two *prospects*—that is, probability distributions over monetary payoffs. We denote each task i by $\{x_i, y_i\}$, and use $c_i \in \{x_i, y_i\}$ to denote the option actually chosen. We shall use the term ‘preference’ only for true preferences, i.e., those that are revealed in single choice experiments. We use $\succ$, $\succeq$ and $\sim$ to denote strict preference, weak preference and indifference, respectively. If the subject chooses x_i in task i in this random lottery experiment, can we infer that she has the preference $x_i \succeq y_i$?

One possibility is that the subject treats each task in isolation, just as if it were the only task, and for real. We shall call this the *isolation hypothesis*. Clearly, it implies that the random lottery design is unbiased. The isolation hypothesis is compatible with Daniel Kahneman and Amos Tversky’s (1979) *prospect theory*, in which individuals use various editing heuristics to simplify decision problems. One such operation applies to choices between two-stage lotteries which have a common first stage; the simplification is to eliminate the first stage. Behaviour that is consistent with that operation has been found in several experiments (Kahneman and Tversky, 1979; Tversky and Kahneman, 1981; Camerer, 1989). However, since none of those experiments used a single choice treatment,[2] this evidence gives only indirect support to the isolation hypothesis.

Holt (1986) considers the polar opposite of the isolation hypothesis: the hypothesis that the subject treats the whole experiment as a single decision problem, reducing all compound prospects to simple ones using the standard probability calculus. We shall call this the *reduction hypothesis*. Consider the position of a subject in a random lottery experiment, who has made all N decisions but does not yet know which task is to be for real. Suppose that all tasks have the same probability of being selected in the random lottery. Then the subject faces a compound prospect made up of the prospects $c_1, \ldots, c_N$, each with a probability of $1/N$. Such a prospect will be written as $(c_1, 1/N; \ldots; c_N, 1/N)$. Now consider the same subject at the start of the experiment. There are now 2^N different ways in which she could respond to the set of N decision tasks in the experiment, each of which leads to a particular compound prospect and thus (after reduction) to a particular simple prospect. The simple prospect which is generated by reducing any given compound prospect p will be written as $r(p)$. The reduction hypothesis is that the subject responds to the individual tasks so as to arrive at the most preferred of these 2^N simple prospects $r(c_1, 1/N; \ldots; c_N, 1/N)$.

Given this hypothesis, it is true, in general, that responses to individual tasks in random lottery experiments coincide with single-choice responses if and only if the subject's preferences satisfy the independence axiom of expected utility theory (EUT). To show this, suppose the reduction hypothesis is true. Take any task j in a random lottery experiment in which all tasks have equal probability of being selected, and consider what inferences can be drawn from the observation that (say) x_j was chosen in that task. Let c_{-j} be the compound prospect $[c_1, 1/(N-1); \ldots; c_{j-1}, 1/(N-1); c_{j+1}, 1/(N-1); \ldots; c_N, 1/(N-1)]$. That is, c_{-j} is an equally weighted mixture of the $N-1$ prospects chosen in all tasks other than j. We can infer that the subject has the true preference $x'_j \succeq y'_j$ where x'_j and y'_j are the simple prospects formed by reducing $[x_j, 1/N; c_{-j}, (N-1)/N]$ and $[y_j, 1/N; c_{-j}, (N-1)/N]$, respectively. Thus, a decision task which ostensibly involves the prospects x_j and y_j actually elicits the subject's preference between x'_j and y'_j. The random lottery design elicits true preferences if and only if $x_j \gtrless y_j \Leftrightarrow x'_j \gtrless y'_j$. The proposition that this implication is true for all x_j, y_j and c_{-j} is equivalent to the independence axiom. Thus, if the reduction hypothesis is true, systematic violations of independence will cause biases in the random lottery design. This bias will take the form of *cross-task contamination*—that is, responses to a given task in a random lottery experiment are influenced by the nature of the other tasks.

Starmer and Sugden (1991) show that if the reduction hypothesis is true, the common consequence form of the Allais paradox[3] will not be observed in a random lottery experiment, irrespective of whether individuals' true preferences satisfy the independence axiom. They also report an experiment which tested for a common consequence effect in three treatments: a two-task random lottery design, a single choice design, and a *hypothetical choice* design (i.e., a design in which there are no financial incentives, subjects merely being asked to say what they would do, if faced with the relevant decision). A significant common consequence effect was found in all three treatments.

This result allows us to reject the reduction hypothesis in the strong form in which we have formulated it, but it does not establish that the random lottery design is immune to cross-task contamination. The reduction hypothesis represents the extreme case of such contamination, just as the isolation hypothesis represents the opposite extreme case in which there is no contamination at all. Between those extremes is a range of cases in which cross-task contamination occurs to varying degrees; Starmer and Sugden's results do not allow us to reject these possibilities.

For example, suppose there are two types of individual—those who behave according to the isolation hypothesis and those who behave according to the reduction hypothesis. If the subject pool contains a sufficiently large number of individuals of the first type, tests which aggregate across subjects will reject the reduction hypothesis; but if the second type of individual is also represented in the subject pool, the random lottery design is still subject to bias.

Perhaps a more realistic possibility is that, for each subject, there is some degree of cross-task contamination, but less than is implied by the reduction hypothesis. From a psychological point of view, the reduction hypothesis seems to require too much mental effort on the part of the subject; it seems more plausible to suppose that subjects tackle one task at a time than that they treat the whole experiment as a single decision problem. But when tackling any one task, a subject might still be influenced by some of the more general and easily recalled properties of the experiment as a whole. Thus, even if the extreme form of the reduction hypothesis is false, there may be *some* cross-task contamination of the kind suggested by Holt. This weaker hypothesis will be called the *contamination hypothesis*. The experiments reported in this paper were designed to test for such contamination effects.

2. Theory: incentive effects

Although our primary concern is to test for contamination effects, our experiments also allow us to investigate another challenge to the random lottery design. In a frequently quoted paper, Smith (1982) argues that for an economic experiment to be valid, the rewards must be sufficiently large to dominate the subjective costs and benefits to the subject of participating in the experiment. Thus, in experiments which investigate decision-making behaviour, incentives should be sufficiently large that subjects do not economize on mental effort by taking short cuts that they would not take when making 'real' economic decisions. Harrison (1994) uses this precept as the basis of a critique of the random lottery design.

Harrison points out that in a random lottery experiment, the apparent incentives offered by the face values of the options are diluted by the fact that each task has only a small probability of being for real. Thus, the *expected* money payout per subject per task is usually very small. Harrison argues that in such cases, the random lottery design is biased towards those responses that are most likely to result from error: for a given task, random lottery responses will contain more errors than single choice responses. If we are to test this claim, we need to combine the principal hypothesis, that the frequency of errors is negatively related to the strength of incentives, with some auxiliary hypothesis about the properties of true preferences and/or the nature of errors. We consider two such auxiliary hypotheses.

Some economists have suggested that true preferences satisfy the axioms of EUT, and that the violations of EUT found in experiments are a product of the weak incentives for correct reasoning offered by those experiments (e.g., Hirshleifer and Riley, 1992, pp. 33–41). On this view, systematic deviations from EUT result from subjects' using simplifying heuristics to economize on mental effort. The stronger the incentives associated with a decision task, the less such heuristics will be used. One implication of this hypothesis is that violations of EUT, such as the common consequence and common ratio effects, should be more pronounced in random lottery experiments than in single choice experiments.[4] Wilcox (1993) presents a version of this argument. He hypothesizes that the greater the dilution of incentives in a random lottery design, the less 'accurate' will be the heuristics used by subjects. As a first approximation, he assumes that fully accurate decisions maximize expected value. He reports an experiment whose results provide some support for his hypothesis when decision tasks require choices over compound prospects, but no support in the case of choices over simple prospects.[5]

Davis and Holt (1993, pp. 449–457) suggest the hypothesis that subjects' risk-aversion is positively related to the strength of incentives. As evidence, they refer to an experiment conducted by Battalio et al. (1990), which found that decisions were less risk-averse in a hypothetical choice treatment than in a random lottery treatment. Davis and Holt conjecture that, similarly, decisions may be less risk-averse in random lottery treatments than in single choice treatments. In other words, if true preferences are defined in terms of single choices, the random lottery design induces systematic errors: subjects understate their real aversion to risk.[6]

3. Strategy

Our strategy is to take as the maintained hypothesis that the random lottery design is unbiased. We test this hypothesis in situations in which we have a priori expectations that individuals' preferences violate the independence axiom in ways which, if the contamination hypothesis were true, would induce observable biases.

Two kinds of violation of independence are widely regarded as particularly robust: the *common consequence* and *common ratio* effects, discovered by Allais (1953). Let a, b be money consequences such that $a > b > 0$. Consider the prospects $R_1 = (a, \lambda; 0, 1 - \lambda)$, $R_2 = (a, \lambda p; b, 1 - p; 0, [1 - \lambda]p)$, $R_3 = (a, \lambda p; 0, 1 - \lambda p)$, $S_1 = S_2 = (b, 1)$, $S_3 = (b, p; 0, 1 - p)$, such that $0 < \lambda < 1$ and $0 < p \leq 1$. The independence axiom implies that $R_1 \gtrless S_1 \Leftrightarrow R_2 \gtrless S_2 \Leftrightarrow R_3 \gtrless S_3$. The common consequence effect is a tendency for preferences over the pair $\{R_3, S_3\}$ to be less risk averse than preferences over $\{R_2, S_2\}$; the common ratio effect is a tendency for preferences over $\{R_3, S_3\}$ to be less risk averse than preferences over $\{R_1, S_1\}$. Preferences which violate independence in a way that is consistent with these two tendencies will be called *Allais preferences*.

Our first two experiments are designed around these effects—the common consequence effect in the case of Experiment 1, the common ratio effect in the case of Experiment 2. In each of these experiments, given the auxiliary hypothesis that individuals have Allais preferences, we can make qualitative predictions about the direction of bias that would be implied by cross-task contamination. The null hypothesis is that there is no bias; the alternative hypothesis is that there is some bias in the predicted direction.

If we are to test the contamination hypothesis we have no option but to build experiments around some particular expected violations of independence. In view of their status in the literature of non-expected utility theory, the common consequence and common ratio effects seem obvious choices for the role of 'expected violation'. However, we cannot be completely confident that even these effects are to be found in true preferences—for two reasons.

First, almost all the evidence of common consequence and common ratio effects comes from hypothetical choice and random lottery experiments. Since the purpose of our research programme is to test whether the random lottery design is biased, and since some of the hypotheses we wish to test imply that the hypothetical choice design is biased too, we cannot treat the existing evidence as definitely establishing that individuals have Allais preferences. Second, at least in relation to the common consequence effect, the evidence from hypothetical choice and random lottery experiments is mixed. While some experiments have found that effect (e.g., Kahneman and Tversky, 1979; Starmer and Sugden, 1991), others have not found it at all, or have found it in some pairs of problems but not in others (e.g., Chew and Waller, 1986; Camerer, 1989; Starmer, 1992). In contrast, the evidence from hypothetical choice and random lottery experiments is generally consistent in confirming the existence of the common ratio effect (e.g., Kahneman and Tversky, 1979; Tversky and Kahneman, 1981; Starmer and Sugden, 1989; Battalio et al., 1990; Bernasconi, 1994). Because we cannot be sure that our subjects have Allais preferences, Experiments 1 and 2 include controls which allow us to test this auxiliary hypothesis.

The common consequence and common ratio effects are defined in terms of very specific classes of decision problem. By using random lottery designs with only two tasks, we are able to build Experiments 1 and 2 around those particular effects. In the third experiment, however, our aim is to test for bias in a many-task random lottery design, more typical of actual applications. In this case, we need an auxiliary hypothesis about violations of independence which applies to a wider class of problems. Here we follow Holt in using Machina's (1982) hypothesis of *fanning-out*. This hypothesis, which is explained in Section 5, imposes a particular pattern on violations of the independence axiom; among its implications are the common ratio and common consequence effects.

4. Experiments 1 and 2

Experiment 1 is built around five decision tasks. Each task requires a choice between two simple prospects. The pair of options in a task is either $\{R_i, S_i\}$ $(i = 1, 2, 3)$ or $\{A_i, B_i\}$ $(i = 1, 2)$:

$$\begin{array}{ll}
R_1 = (£10, 15/60; £6, 40/60; 0, 5/60) & S_1 = (£6, 60/60) \\
R_2 = (£10, 15/60; 0, 45/60) & S_2 = (£6, 20/60; 0, 40/60) \\
R_3 = (£10, 45/60; 0, 15/60) & S_3 = (£6, 60/60) \\
A_1 = (£6, 60/60) & B_1 = (£7, 5/60; -£5, 55/60) \\
A_2 = (0, 60/60) & B_2 = (£7, 5/60; -£5, 55/60)
\end{array}$$

Negative payoffs represent payments from the subject to the experimenters.

Each $\{R_i, S_i\}$ task requires a choice between a 'riskier' option R_i and a 'safer' option S_i. Each $\{A_i, B_i\}$ task is constructed so that B_i, while not stochastically dominated by A_i, would appeal only to a person with strongly risk-loving preferences. The intention was that subjects should think of these as genuine decision tasks, but that almost all of them would choose A_i.

Notice that the pair of tasks $\{R_1, S_1\}$, $\{R_2, S_2\}$ is an instance of the class of problem pairs in which the common consequence effect has been observed. (In terms of the notation introduced in Section 3, $a = £10$, $b = £6$, $\lambda = 3/4$, $p = 1/3$.) If subjects have Allais preferences, decisions will be less risk-averse in the case of $\{R_2, S_2\}$ than in the case of $\{R_1, S_1\}$. The other three tasks are linked to $\{R_1, S_1\}$ and $\{R_2, S_2\}$ by virtue of the facts that $R_1 = r(R_3, 1/3; A_1, 2/3)$, $S_1 = r(S_3, 1/3; A_1, 2/3)$, $R_2 = r(R_3, 1/3; A_2, 2/3)$, and $S_2 = r(S_3, 1/3; A_2, 2/3)$. Our experiment exploits these relationships to test for contamination effects.

Subjects were randomly divided into four groups.[7] Each subject faced 20 decision tasks involving choices between two (or in some cases three) prospects with money outcomes. At the start of the experiment, subjects in Groups 1.1 and 1.2 were told that the first 18 of these tasks were hypothetical. The payment they received for taking part in the experiment would depend entirely on the last two tasks; one of these tasks, to be selected by a random device at the end of the experiment, would be for real. After subjects had responded to the first 18 tasks, they were reminded that one of the two final tasks would be for real, and the probability of selection for each task (see below) was explained. Subjects in Groups 1.3 and 1.4 were told that the first 19 tasks were hypothetical and that the final task was for real; they were reminded of this before facing the final task.

The hypothetical choice tasks have no significance for our tests of the random lottery design. The first 18 of these tasks did not differ between the four groups; for Groups 1.3 and 1.4, the nineteenth task was a repeat of one of the first 18. Leaving aside these hypothetical choice tasks, Groups 1.1 and 1.2 faced a two-task random lottery design while Groups 1.3 and 1.4 faced a single choice design. The relevant tasks for the four groups are described below:

Group 1.1. Random lottery design with two tasks: Task 1 = $\{R_3, S_3\}$; Task 2 = $\{A_1, B_1\}$. Task 1 is for real with probability $1/3$; Task 2 is for real with probability $2/3$.

Group 1.2. Random lottery design with two tasks: Task 1 = $\{R_3, S_3\}$; Task 2 = $\{A_2, B_2\}$. Task 1 is for real with probability $1/3$; Task 2 is for real with probability $2/3$.

Group 1.3. Single choice design with the task $\{R_1, S_1\}$.

Group 1.4. Single choice design with the task $\{R_2, S_2\}$.

Notice that the task $\{R_3, S_3\}$ is common to Groups 1.1 and 1.2. If the random lottery design is unbiased, the expected proportion of subjects choosing R_3 will be the same in both groups. However, if the reduction hypothesis is true and if A options are always preferred to B options, Task 1 in Group 1.1 is equivalent to a choice between $r(R_3, 1/3; A_1, 2/3)$ and $r(S_3, 1/3; A_1, 2/3)$, i.e., to the choice $\{R_1, S_1\}$, while Task 1 in Group 1.2 is equivalent to a choice between $r(R_3, 1/3; A_2, 2/3)$ and $r(S_3, 1/3; A_2, 2/3)$, i.e., to $\{R_2, S_2\}$. This implies that the expected distribution of responses between the R and S options is the same in Group 1.1 as in Group 1.3, and the same in Group 1.2 as in Group 1.4. If subjects have Allais preferences, decisions will be more risk-averse in the case of $\{R_2, S_2\}$ than in the case of $\{R_1, S_1\}$. Thus, we can predict the direction of any contamination effect for Groups 1 and 2: the expected frequency of R_3 choices will be greater in Group 1.2 than in Group 1.1. Groups 1.3 and 1.4 provide a test for the presence of a common consequence effect in single choices.

Experiment 2 has a similar design, adapted to the common ratio effect instead of the common consequence effect. It uses the following three tasks:

$$R_4 = (£24, 80/100; 0, 20/100) \quad S_4 = (£15, 100/100)$$

$$R_5 = (£24, 20/100; 0, 80/100) \quad S_5 = (£15, 25/100; 0, 75/100)$$

$$A_3 = (0, 100/100) \quad B_3 = (£1, 40/100; -£2, 60/100)$$

Here, the pair of tasks $\{R_4, S_4\}$, $\{R_5, S_5\}$ is an instance of the class of problem pairs in which the common ratio effect has been observed. (In the notation introduced in Section 3: $a = £24, b = £15, \lambda = 4/5, p = 1/4$.) In addition, $R_5 = r(R_4, 1/4; A_3, 3/4)$ and $S_5 = r(S_4, 1/4; A_3, 3/4)$.

In Experiment 2, subjects were divided at random into three groups. Each subject faced 20 decision tasks. The first 19 of these for Groups 2.1 and 2.2, and the first 18 for Group 2.3, were hypothetical. The remaining, non-hypothetical tasks are described below.

Group 2.1. Single choice design with task $= \{R_4, S_4\}$.
Group 2.2. Single choice design with task $= \{R_5, S_5\}$.
Group 2.3. Random lottery design with two tasks: Task 1 $= \{R_4, S_4\}$; Task 2 $= \{A_3, B_3\}$. Task 1 is for real with probability 1/4; Task 2 is for real with probability 3/4.

Notice that $\{R_4, S_4\}$ is common to Group 2.1 (where it is faced as a single choice) and Group 2.3 (where it is faced in a random lottery treatment). Thus, if the random lottery design is unbiased, the expected proportion of subjects choosing R_4 will be the same in the two groups. However, if the reduction hypothesis is true and if A_3 is always preferred to B_3, Task 1 in Group 2.3 is equivalent to a choice between $r(R_4, 1/4; A_3, 3/4)$ and $r(S_4, 1/4; A_3, 3/4)$, i.e., to the choice $\{R_5, S_5\}$. This implies that the expected distribution of responses between R and S is the same in Group 2.3 as in Group 2.2. If subjects have Allais preferences, any contamination effect will make the expected frequency of R_4 choices greater in Group 2.3 than in Group 2.1. Groups 2.1 and 2.2 provide a test for the presence of a common ratio effect in single choices.

We now turn to the mechanics of the experiments. 201 subjects took part in Experiment 1, and 150 new subjects took part in Experiment 2. Subjects were recruited on the campus of the University of East Anglia. Decision tasks were presented as visual displays on a computer monitor, and responses were entered at a keyboard. A typical display (of Task 1 for Group 1.1) is shown in figure 1. When a decision task was played out for real, the subject drew a disc from a bag of 60 (or, in Experiment 2, 100) numbered discs, and the number of the disc drawn determined the payoff. In the display, each option is represented by a box. The entries in the box are payoffs. The numbers along the top of each box refer to the discs associated with each payoff, and the numbers along the bottom show the probability of each payoff (in units of 1/60 or 1/100). The widths of the subdivisions of the boxes

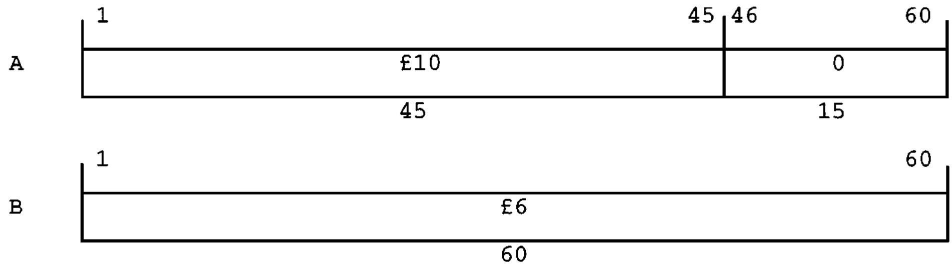

Figure 1. A typical task display.

are proportional to the relevant probabilities. For each task, the location of options as 'top' and 'bottom' was randomized, independently for each subject. The left-right orientation of payoffs in the boxes was also randomized, subject to the constraint that the two options were always comonotonic.[8] For subjects in Groups 1.1, 1.2 and 2.3, the order of Tasks 1 and 2 was randomized. To ensure that these subjects were able to act on the reduction hypothesis, if they chose to do so, the software required them to look at both tasks at least once before responding to either. The selection of the task to be for real was implemented by rolling a six-sided die (Experiment 1) or by drawing a card from a pack (Experiment 2).

The results are shown in Tables 1 and 2. The first thing to notice is that, as expected, *A* options were almost always chosen in preference to *B* options. Each of the six subjects

Table 1. Results of Experiment 1.

		Number (percentage) of subjects choosing			
		Task 1		Task 2	
Group	*n*	*R*	*S*	*A*	*B*
1.1	57	28 (49.1)	29 (50.9)	57 (100.0)	0 (0.0)
1.2	62	31 (50.0)	31 (50.0)	57 (91.9)	5 (8.1)
1.3	38	25 (65.8)	13 (34.2)	N/A	N/A
1.4	44	24 (54.5)	20 (45.5)	N/A	N/A

Table 2. Results of Experiment 2.

		Number (percentage) of subjects choosing			
		Task 1		Task 2	
Group	*n*	*R*	*S*	*A*	*B*
2.1	51	21 (41.2)	30 (58.8)	N/A	N/A
2.2	46	24 (52.2)	22 (47.8)	N/A	N/A
2.3	53	20 (37.7)	33 (62.3)	52 (98.1)	1 (1.9)

who chose a B option in his Task 2 also chose the R option in Task 1. This evidence is consistent with the subject pool's containing a small minority of strongly risk-loving individuals. We cannot eliminate B-choosing subjects from our analysis without biasing our between-group comparisons. However, it seems reasonable to assume that any subject who was so risk-loving as to choose B in his Task 2 would still have chosen R in Task 1 if he had been constrained to choose A in Task 2. On this assumption, we can test for contamination effects in Experiment 1 by comparing the Task 1 responses of all subjects in Groups 1.1 and 1.2; for Experiment 2, the corresponding comparison is between the Task 1 responses of Groups 2.1 and 2.3.

In the case of Experiment 1, the proportion of subjects choosing R is virtually the same in the two groups (49.1% in Group 1.1, 50.0% in Group 1.2). In the case of Experiment 2, the proportions choosing R are very similar (41.2% in Group 2.1, 37.7% in Group 2.3); the difference is not statistically significant. In neither case, then, is there any evidence of any contamination effect.

However, these comparisons are appropriate tests of the contamination hypothesis only if subjects' true preferences violate the independence axiom. Our experiments allow direct tests of this property of true preferences within our subject pools. For Experiment 1, we can compare Groups 1.3 and 1.4. There is no significant difference between the proportions of R choices in the two groups (65.8% in Group 1.3, 54.5% in Group 1.4; $z = 1.04$). For Experiment 2, the relevant comparison is between Groups 2.1 and 2.2. Here too, the proportions of R choices in the two groups (41.2% in Group 2.1, 52.2% in Group 2.2; $z = 1.08$) are not significantly different. Thus, in neither case can we reject the null hypothesis that true preferences satisfy the independence axiom.

The implication is that the comparisons between Groups 1.1 and 1.2, and between Groups 2.1 and 2.3, are not appropriate tests of the contamination hypothesis. However, since Experiment 2 presented the same choice problem $\{R_4, S_4\}$ in both random lottery and single choice treatments, we *can* use the data from that experiment to test the hypothesis that random lottery responses are less risk averse than single choice responses. That hypothesis implies that the expected frequency of R choices is greater in Group 2.3 than in Group 2.1. Our data do not support that hypothesis.

5. Experiment 3

The designs of Experiments 1 and 2 were premised on the expectation that strong common consequence and common ratio effects would be found in single choices; but we failed to find these effects. In the case of the common consequence effect, the results generated by our single choice treatment are not out of line with the findings of other experiments which have used hypothetical choice and random lottery designs. As noted in Section 3, previous evidence concerning the common consequence effect is mixed. Why some experiments have found this effect while others have not remains an unsolved problem; but we see no particular reason to attribute the absence of a common consequence effect in Experiment 1 to the single choice design.

In the case of the common ratio effect, the evidence from previous hypothetical choice and random lottery experiments is much less ambiguous. In this respect, our failure to find a significant common ratio effect is more surprising. One possible explanation is that, although individuals in our subject pool have preferences of the kind that generate the common ratio effect, our statistical tests were not powerful enough to pick this up. Notice that single choice designs require between-subjects tests, while most random lottery experiments produce within-subject data. Inevitably there is more stochastic variation in between-subjects data, making it harder to detect underlying patterns. However, there is another possibility: that the common ratio effect is not, in fact, a property of *true* preferences, even though it is found in hypothetical choice and random lottery experiments. That interpretation would be consistent with the hypothesis that the stronger incentives are, the less deviation there is from EUT.

In designing Experiment 3, we had two main objectives. First, as in the case of Experiments 1 and 2, we wanted to test the contamination hypothesis in a context in which we could expect the independence axiom to be violated. For Experiment 3, however, we chose a somewhat different approach. In Experiments 1 and 2, the random lottery treatments involved only two tasks. In practical applications of the random lottery design, there are usually many tasks, and we wished to test the contamination hypothesis in such a setting. There are some reasons for expecting the extent of any bias in the random lottery design to depend on the number of tasks. On the one hand, it might be argued that, the more tasks there are in a random lottery experiment, the more likely subjects are to use the simplifying heuristic of treating each task in isolation. On the other hand, the more tasks there are, the more incentives are diluted; thus if bias is a product of dilution, its extent will increase with the number of tasks.

Our predictions about the direction of contamination effects in the many-task design are derived from Machina's (1982) fanning-out hypothesis. In Machina's theoretical framework, preferences in a sufficiently small neighbourhood of probability space around any given prospect p can be represented as expected utility maximizing, relative to a *local* utility function defined for p. The fanning-out hypothesis is that, for any prospects p, p' where p' (first-order) stochastically dominates p, the local utility function for p' is more risk-averse than that for p. We recognize that this hypothesis is not completely successful in organising all the data that have been generated by experiments to date; but it would seem that any general theory of preferences that is to explain that evidence must have some fanning-out tendency (see Camerer, 1995). Given the objective of predicting the direction of potential bias in random lottery experiments, Machina's theory proves to be particularly tractable.

Our second objective was to test whether the common ratio effect is eliminated when a single choice design is used in place of a random lottery design. As noted above, the contrast between the results of Experiment 2 and the mass of evidence from random lottery experiments suggests that violations of independence *might* be less frequent in single choices; but to reach firm conclusions, we need an experimental design in which single choice and random lottery treatments are applied to the same tasks and to the same subject pool. In such an experiment, the random lottery treatment should be as similar as possible to those which, in previous experiments, have induced common ratio effects. It is particularly important to replicate the degree to which the random lottery designs used in those previous experiments have diluted incentives. The two-task random lottery designs used in Experiments 1 and 2 are clearly inappropriate for this purpose. In Experiment 3 we used a more typical 20-task design.

Experiment 3 was built around the two tasks $\{R_6, S_6\}$, $\{R_7, S_7\}$, where the relevant prospects are:

$$R_6 = (£15, 80/100; 0, 20/100) \quad S_6 = (£10, 100/100)$$
$$R_7 = (£15, 20/100; 0, 80/100) \quad S_7 = (£10, 25/100; 0, 75/100).$$

This pair of tasks is typical of those in which the common ratio effect has been found, and is very similar to the pair of tasks around which Experiment 2 was built. (In the notation introduced in Section 3: $a = £15, b = £10, \lambda = 4/5, p = 1/4$.) In this context, the common ratio effect is a tendency for decisions to be less risk-averse in the case of $\{R_7, S_7\}$ than in the case of $\{R_6, S_6\}$.

The experiment also used two sets of 18 additional decision tasks, a *low value* set and a *high value* set. There is a one-to-one relation between tasks in the two sets. For every high value task $\{x, y\}$ and the corresponding low value task $\{x', y'\}$, x stochastically dominates x' and y stochastically dominates y'. In the high value set, the expected values of the options range from £7.20 to £12.50, with a mean of £10.63; in the low value set, the range is from £2.36 to £5.00, with a mean of £3.54. The values of the possible payoffs are similar in the two sets, with a maximum of £25 in each case; differences in expected value between the two sets are mainly due to differences in the probabilities associated with given payoffs. Thus, the payoffs and expected values of the high value options are similar to those of R_6 and S_6, while the low value options are similar to R_7 and S_7.

The experiment involved 202 subjects, none of whom had taken part in the previous two experiments, randomly divided into four groups. Each group faced 20 tasks. The set of tasks, and the incentive system applied to particular tasks, varied across groups in the following way:

Group 3.1.
$\{R_6, S_6\}$: for real
$\{R_7, S_7\}$: hypothetical choice
18 high value tasks: hypothetical choice
Group 3.2.
$\{R_6, S_6\}$: hypothetical choice
$\{R_7, S_7\}$: for real
18 high value tasks: hypothetical choice
Group 3.3. Random lottery design with 20 tasks: $\{R_6, S_6\}$, $\{R_7, S_7\}$, and 18 high value tasks. Each task has 1/20 probability of being for real.
Group 3.4. Random lottery design with 20 tasks: $\{R_6, S_6\}$, $\{R_7, S_7\}$, and 18 low value tasks. Each task has 1/20 probability of being for real.

Notice that each of the two common ratio tasks is faced in four different treatments: as a single choice (i.e., as the task which is for real in a treatment in which one task is for real and all the other tasks are hypothetical), as a hypothetical choice, in a random lottery design along with 18 high value tasks, and in a random lottery design along with 18 low value tasks. This allows us to carry out two different kinds of test. First, we can test for common ratio effects in hypothetical choice, random lottery and single choice treatments. Second, by comparing responses to the common ratio tasks between the two random lottery groups, we can test for a contamination effect.

Consider a subject in Group 3.3 or Group 3.4, facing one of the common ratio tasks, say $\{R_6, S_6\}$. The reduction hypothesis implies that this task will elicit the subject's true preference between $r(R_6, 1/20; m, 19/20)$ and $r(S_6, 1/20; m, 19/20)$, where m is an equally weighted mixture of the options chosen in the other 19 tasks. Recall that for the 18 non-common-ratio tasks, each option in each of the high value tasks faced by Group 3.3 stochastically dominates a corresponding option in the low value task faced by Group 3.4. The nineteenth option, i.e., $\{R_7, S_7\}$, is common to both groups. Thus, we can expect that the m of a typical Group 3.3 subject will stochastically dominate the m of a typical Group 3.4 subject. If the fanning-out hypothesis is true, preferences in the region of m will tend to be less risk-averse for Group 3.4 subjects than for Group 3.3 subjects. Thus, any contamination effect will tend to make Group 3.4's responses to the common ratio task less risk-averse than those of Group 3.3.

The mechanics of Experiment 3 were very similar to those of Experiments 1 and 2. The main difference was that the order of the complete set of 20 tasks was randomized, independently for each subject. Thus, for subjects in Groups 3.1 and 3.2, the position of the single choice task in the series of 20 tasks was random. At the start of the experiment, each of these subjects was told the position of her single choice task (e.g., that it was the thirteenth in the series); immediately before facing this task, the subject was reminded that it was for real. The reason for this modification was to ensure that any differences between random lottery and single choice responses to the common ratio problems could not be attributed to differences in the order in which tasks were faced.

The software allowed subjects in all groups to backtrack at any point in the experiment, going back to previous tasks and changing their responses if they wished. After they had made all 20 responses they were reminded of this option. In this way, and as in Experiments 1 and 2, we gave subjects the opportunity to treat the whole experiment as a single decision problem if they so wished.[9]

The main results are shown in Table 3. First, we test for a contamination effect. Given the auxiliary assumption of fanning-out preferences, the contamination hypothesis implies that, for each of the two common ratio tasks, Group 3.4 will be less risk averse than Group 3.3. In fact, for each task, there is a *lower* proportion of R choices in Group 3.4 (the differences are not significant): there is no evidence of any contamination.

Table 3. Results of Experiment 3.

		Number (percentage) of subjects choosing			
Group	n	R_6	S_6	R_7	S_7
3.1	49	14 (28.6)	35 (71.4)	29 (59.2)	20 (40.8)
3.2	56	22 (39.3)	34 (60.7)	25 (44.6)	31 (55.4)
3.3	52	13 (25.0)	39 (75.0)	31 (59.6)	21 (40.4)
3.4	45	8 (17.8)	37 (82.2)	21 (46.7)	24 (53.3)

Could our failure to find contamination be due to the failure of the fanning-out hypothesis? We can test this possibility by comparing responses to high value and low value tasks in the random lottery groups. Eight of these pairs of tasks have the special property that the relevant options x, y, x', y' are different probability distributions over the same three money consequences; if these prospects are plotted in a Marschak-Machina triangle diagram,[10] the lines xy and $x'y'$ are parallel. In each case, the fanning-out hypothesis implies that true preferences over $\{x', y'\}$ will be less risk-averse than true preferences over $\{x, y\}$. We can test for fanning-out by comparing Group 3.3 responses to each $\{x, y\}$ task with Group 3.4 responses to the corresponding $\{x', y'\}$ task. Since both tasks were faced in random lottery treatments, we must of course allow for the possibility of contamination effects. But notice that in Group 3.3, the principal source of potential contamination is from other high value tasks, while in Group 3.4 it is from other low value tasks. Thus if true preferences have the fanning-out property, then whatever the extent of contamination, we can expect Group 3.4 responses to each $\{x', y'\}$ to be less risk averse than Group 3.3 responses to $\{x, y\}$.

In seven of the eight pairs of tasks, the difference between the responses of the two groups is in the direction predicted by the fanning-out hypothesis (in the eighth case, the distributions of choices are virtually identical); this difference is significant at the 5% level in five cases. Averaging over all eight pairs, the riskier option is chosen in 24.8% of cases in Group 3.3 and in 42.2% of cases in Group 3.4. It seems, then, that our subjects' preferences did have the fanning-out property, but that this property did not generate cross-task contamination.

We now turn to the question of whether the common ratio effect is observed in each of the four treatments. To test for this effect in single choices, we compare Group 3.1's responses to $\{R_6, S_6\}$ with Group 3.2's responses to $\{R_7, S_7\}$. The proportion of R_7 choices is significantly greater than the proportion of R_6 choices ($z = 1.70$), although only just so: the critical value of z for a one-tail test is 1.65. On this basis we can reject the null hypothesis of no common ratio effect for the single choice treatment. That hypothesis can also be rejected for the hypothetical choice treatment (comparison between Group 3.2's responses to $\{R_6, S_6\}$ and Group 3.1's responses to $\{R_7, S_7\}$; $z = 2.04$), for the random lottery treatment with 18 high value tasks (Group 3.3; $z = 3.57$), and for the random lottery treatment with 18 low value tasks (Group 3.4; $z = 2.93$).[11]

Finally, we test the hypothesis that the random lottery treatment induces less risk-averse responses than does the single choice treatment. In the case of $\{R_6, S_6\}$, the proportion of R_6 choices is slightly *less* in the two random lottery treatments (21.6%) than in the single choice treatment (28.6%). In the case of $\{R_7, S_7\}$, the difference in the proportion of R_7 choices is in the right direction (53.6% for the random lottery design, 44.6% for single choices), but not significant ($z = 1.07$).[12]

The net effect of these differences is to make the common ratio effect more pronounced in the random lottery treatments than in the single choice treatment. However, it does not seem plausible to explain this fact in terms of the stronger incentives in the single choice treatment, as the common ratio effect is also more pronounced in the random lottery treatments than in the hypothetical choice treatment—where there are no financial incentives at all.

6. Conclusions

The three experiments reported in this paper can be seen as belonging to a set of five related experiments; the other two are those reported by Starmer and Sugden (1991) and by Beattie and Loomes (1997). These experiments involved a total of over 850 subjects. Each experiment included both single choice and random lottery treatments, and offered very substantial incentives—expected payoffs of up to £19.20, equivalent to around five hours labour at a typical unskilled wage rate—for very straightforward decision tasks. They are by far the most sustained attempt to date to test for potential biases in the random lottery incentive system. In summarizing our present results, it is useful to put them in this wider context.

The main purpose of the three experiments reported in this paper was to test for cross-task contamination effects in the random lottery design. One of the experiments (Experiment 2) was designed so that, if contamination effects were at work, there would be systematic differences between responses to single choice and random lottery treatments. Two other experiments (Experiments 1 and 3) were designed so that contamination effects would show up as systematic differences between two random lottery treatments. Since the contamination hypothesis predicts bias only when subjects' true preferences violate the independence axiom, we designed the experiments so that the violations necessary to induce such biases were those that are most commonly predicted: the common consequence and common ratio effects, and the general property of fanning-out preferences. Surprisingly, in Experiments 1 and 2 we failed to find significant violations of independence in the single choice treatments we used as controls. However, in the case of Experiment 3, there was clear evidence of fanning-out, but no evidence that this induced contamination. Our main conclusion, then, is that there is no evidence of cross-task contamination in the random lottery design.

Two of our present experiments (Experiments 2 and 3) allow direct comparisons between single choice and random lottery responses to the same decision problems. Starmer and Sugden's and Beattie and Loomes's experiments also allow such comparisons. All but one of the relevant decision problems in these four experiments is a choice between two simple prospects. In each of these cases, no significant difference between the two treatments is found. Further, when one looks at the whole body of data for such tasks, the observed differences in the distributions of choices between treatments seem to show no consistent pattern. In particular, there is no evidence of any general tendency for random lottery responses to be less risk-averse than single choices; and there is no evidence of any general tendency for violations of expected utility theory to be more frequent in random lottery treatments than in single choice treatments. After allowing for normal sampling variation, the hypothesis that there is no difference between responses to random lottery and single choice experiments seems adequate to organize the data—that is, the data concerning choice among simple prospects.

The qualification about simple prospects is significant. Beattie and Loomes find a significant difference between random lottery and single choice responses to one of the four decision tasks they study; this task differs from the others in that it requires a choice between compound prospects. Here random lottery responses are less risk averse than single choices.[13] Thus, the *general* validity of the random lottery incentive system remains an open question. What can be said on the basis of our findings is that this system does appear to be unbiased when applied to choices among simple prospects.

Acknowledgments

The research reported in this paper was carried out as part of the Economic and Social Research Council's programmes on Economic Beliefs and Behaviour and on Risk and Human Behaviour (award nos. L 122 251 024 and L 211 252 053). An earlier version of the paper was presented at the Sixth Amsterdam Workshop on Experimental Economics in August 1997 and at the ESRC Risk and Human Behaviour conference in York in September 1997. We are grateful to participants at those conferences, and in particular to Jane Beattie, Colin Camerer, Graham Loomes, Arthur Schram, and an anonymous referee, for advice. We also thank Jan Anderson for help in organizing the experiments.

Notes

1. There are some variants on the system. For example, in some experiments, a subset of *subjects* is selected at random after a task has been completed, and only those subjects are rewarded. This method raises similar issues to those discussed in this paper.
2. Camerer (1989) reports a random lottery experiment in which, after the random mechanism had determined which task was for real, each subject was allowed to change his decision for that task. Out of 80 subjects, only two opted to change their decisions. Camerer interprets this as evidence that single choices are very similar to random lottery choices. However, this result might also be explained by status quo bias, or by subjects not wanting to acknowledge having made errors.
3. The common consequence effect is defined in Section 3.
4. An experiment carried out by Conlisk (1989) is sometimes quoted as evidence that violations of EUT are less frequent in single choice than in random lottery designs. Conlisk investigated the common consequence effect using a single choice design. In each of the two relevant tasks, almost all subjects (26 out of 27 in one case, 24 out of 26 in the other) chose the riskier option. Clearly, this distribution of responses between riskier and safer choices is far too asymmetric for the experiment to be a satisfactory test for systematic deviations from EUT.
5. In Wilcox's experiment, a tendency in the direction of expected value maximization can be distinguished from a tendency towards lower risk aversion. Our experiments cannot distinguish between these tendencies, as in our decision tasks, the riskier options always have the higher expected values.
6. In Starmer and Sugden's (1991) experiment, single choice responses were somewhat more risk-averse than random lottery responses, although the difference was not statistically significant. Davis and Holt (1993, pp. 449–457) interpret this evidence as giving some support for their conjecture.
7. Each subject was assigned to one of Groups 1, 2, 3, 4 with probability 0.3, 0.3, 0.2, 0.2. We were principally interested in differences between Groups 1 and 2; the other two groups were controls.
8. That is, there is a ranking of states of the world, the same for both options, such that larger payoffs occur in the higher ranked states. If prospects are not comonotonic, apparent violations of independence can be due to the juxtaposition effects predicted by regret theory.
9. The number of times each subject used the "backtrack" key was recorded. In all four groups, about a third of subjects used this option. An analysis of the data suggests that this option was most commonly used after all tasks had been faced once; subjects in Groups 1 and 2 tended to backtrack to the single choice task, while subjects in Groups 3 and 4 tended to go back to the beginning and review all their decisions. We do not know how far subjects changed their original decisions.
10. This diagram represents all probability distributions over three given consequences as a unit triangle in probability space. The horizontal axis measures the probability of the least-preferred of the three consequences while the vertical axis measures the probability of the most-preferred consequence.

11. Beattie and Loomes (1977) report a similar finding. They find a common ratio effect, significant at the 1% level, in all three of their treatments: single choice, random lottery, and hypothetical choice.
12. There does seem to be a tendency for responses to be less risk-averse in the hypothetical choice treatment than in the random lottery treatment. In addition to the data shown in Table 3, there are 15 high value decision tasks in which one option is unambiguously riskier than the other. Aggregating across these 15 tasks, the riskier option was chosen in 40.1% of cases in the hypothetical choice treatment (Groups 1 and 2), compared with 29.2% of cases in the random lottery treatment (Group 3). Each of the other three high value tasks involved choices between a stochastically dominating and a stochastically dominated option. The dominated option was chosen in 8.6% of cases in the hypothetical choice treatment, compared with 3.2% of cases in the random lottery treatment. This latter result suggests that incentives may reduce the frequency of errors that result from carelessness.
13. And, contrary to Wilcox's (1993) results, further from expected value maximization.

References

Allais, M. (1953). "Le Comportement de L'homme Rationnel Devant le Risque: Critique des Postulats et Axiomes de L'école Américaine." *Econometrica*. 21, 503–546.

Battalio, R.C., Kagel, J.H., and Jiranyakul, K. (1990). "Testing Between Alternative Models of Choice Under Uncertainty: Some Initial Results." *Journal of Risk and Uncertainty*. 3, 25–50.

Beattie, J. and Loomes, G. (1997). "The Impact of Incentives Upon Risky Choice Experiments." *Journal of Risk and Uncertainty*. 14, 149–162.

Bernasconi, M. (1994). "Nonlinear Preferences and Two-Stage Lotteries." *Economic Journal*. 104, 54–70.

Camerer, C. (1989). "An Experimental Test of Several Generalized Expected Utility Theories." *Journal of Risk and Uncertainty*. 2, 61–104.

Camerer, C. (1995). "Individual Decision Making." In John Kagel and Alvin Roth (eds.). *Handbook of Experimental Economics*. Princeton, NJ: Princeton University Press, chap. 8.

Conlisk, J. (1989). "Three Variants on the Allais Example." *American Economic Review*. 79, 392–407.

Davis, D.D. and Holt, C.A. (1993). *Experimental Economics*. Princeton, NJ: Princeton University Press.

Harrison, G.W. (1994). "Expected Utility Theory and the Experimentalists." *Empirical Economics*. 19, 223–253.

Hirshleifer, J. and Riley, J.G. (1992). *The Analytics of Uncertainty and Information*. Cambridge: Cambridge University Press.

Holt, C.A. (1986). "Preference Reversals and the Independence Axiom." *American Economic Review*. 76, 508–515.

Kahneman, D. and Tversky, A. (1979). "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica*. 47, 263–291.

Machina, M. (1982). "Expected Utility Analysis Without the Independence Axiom." *Econometrica*. 50, 277–323.

Smith, V.L. (1982). "Microeconomic Systems as an Experimental Science." *American Economic Review*. 72, 923–955.

Starmer, C. (1992). "Testing New Theories of Choice Under Uncertainty Using the Common Consequence Effect." *Review of Economic Studies*. 59, 813–830.

Starmer, C. and Sugden, R. (1989). "Probability and Juxtaposition Effects: An Experimental Investigation of the Common Ratio Effect." *Journal of Risk and Uncertainty*. 2, 159–178.

Starmer, C. and Sugden, R. (1991). "Does the Random-Lottery Incentive System Elicit True Preferences? An Experimental Investigation." *American Economic Review*. 81, 971–978.

Tversky, A. and Kahneman, D. (1981). "The Framing of Decisions and the Psychology of Choice." *Science*. 211, 453–458.

Wilcox, N.T. (1993). "Lottery Choice: Incentives, Complexity and Decision Time." *Economic Journal*. 103, 1397–1417.