University of Connecticut

Department of Economics Working Paper Series

The Common Ratio Effect in Choice, Pricing, and Happiness Tasks

by

Mark Schneider
Chapman University

Mikhael Shor
University of Connecticut

Working Paper 2016-29
May 2016

365 Fairfield Way, Unit 1063
Storrs, CT 06269-1063
Phone: (860) 486-3022
Fax: (860) 486-4463
http://www.econ.uconn.edu/

The Common Ratio Effect in Choice, Pricing, and Happiness Tasks

Mark Schneider[1] Mikhael Shor[2]

May 2016

Abstract

The Allais common ratio effect is one of the most robust violations of rational decision making under risk. In this paper, we conduct a novel test of the common ratio effect in which we elicit preferences for the common ratio choice alternatives in choice, pricing, and happiness rating tasks. We find that both the consistency and distribution of responses differs systematically across tasks, with modal choices replicating the Allais preference pattern, modal happiness ratings exhibiting consistent risk aversion, and modal prices maximizing expected value. We discuss the predictions of various cognitive explanations of the common ratio effect in the context of our experiment. We find that a dual process framework provides the most complete account of our results. Surprisingly, we also find that although the Allais pattern was the modal behavior in the choice task, *none* of the 158 respondents in our experiment exhibited the Allais pattern simultaneously in choice, happiness, and pricing tasks. Our results constitute a new paradox for the leading theories of choice under risk.

Key Words: *Common Ratio Effect; Preference Reversals; Dual Processes; Happiness Ratings*

[1] Economic Science Institute, Chapman University. One University Drive, Orange, CA 92866.
Phone: (860) 268-7791. schneide@chapman.edu

[2] Department of Economics, University of Connecticut, 321 Oak Hall. 365 Fairfield Way, Unit 1063.
Storrs, CT 06269-1063, mike.shor@uconn.edu

1 Introduction

The Allais common ratio effect (Allais, 1953) is widely regarded as one of the most robust empirical violations of rational decision making under risk. Since the effect was introduced by Allais (1953) and popularized by Kahneman and Tversky (1979), it has been replicated in numerous studies[3] (e.g. Ballinger & Wilcox, 1997; Loomes & Sugden, 1998; Barron & Erev, 2003; Baucells & Heukamp, 2010) and has served as a motivating example for many descriptive models of choice under risk including prospect theory (Kahneman & Tversky, 1979), regret theory (Loomes & Sugden, 1982), similarity theory (Rubinstein, 1988; Leland, 1994), and salience theory (Bordalo et al., 2012).

Despite the widespread observation of the common ratio effect when experimental subjects are provided *choices* between lotteries, the effect has surprisingly not been investigated for its robustness in other standard response modes such as monetary valuation (pricing) tasks or happiness rating tasks. In this paper, we test for the common ratio effect in choice, happiness rating, and pricing tasks, motivated by the possibility that different response modes may help us to better understand the processes that generate the common ratio effect.

In the context of our study, the alternative explanations of the common ratio effect fall into three classes of models: (i) procedure-invariant models (e.g., Kahneman & Tversky, 1979; Tversky & Kahneman, 1992) which predict the same choice pattern across response modes; (ii) comparative models (e.g., Loomes & Sugden, 1982; Bell, 1982; Rubinstein, 1988; Leland, 1994; Bordalo et al., 2012) which predict that preferences can depend on joint versus separate evaluation of choice alternatives, but not on different types of 'separate evaluation' tasks such as rating versus pricing; and (iii) dual process models (Mukherjee, 2010; Schneider & Coulter, 2015) which predict that tasks which systematically elicit different processes will produce systematically different preferences. We distinguish among these different explanations by eliciting preferences through three tasks. In addition to the traditional joint evaluation of choices, we examine two distinct measures of separate evaluation of alternatives: pricing, which asks subjects to value each alternative, and happiness ratings, which ask for a subjective assessment. We find that the dual process evaluability framework (DPEF) of Schneider & Coulter provides the most complete account of our results and predicts the modal preference patterns across tasks.

[3] But see Blavatskyy (2010) which observed that the common ratio effect reverses under some parameter values.

However, even this approach cannot account for the large number of indifference responses across tasks which is inconsistent with all models of the common ratio effect discussed here.

We proceed by first reviewing the common ratio effect and a number of explanations for the effect. We then introduce our experiment, present and discuss both the aggregate and within-subject results, and conclude with a discussion of the ability of different explanations to explain our results.

2 The Allais Common Ratio Effect

Virtually every alternative to expected utility theory, the standard model of rational decision making under risk, that has been developed since Prospect Theory (Kahneman & Tversky, 1979) provides an explanation for the common ratio effect. Indeed, the effect poses a minimum standard for alternative theories of decision making. Consider the most famous version of the effect due to Kahneman and Tversky (1979) in which a decision maker is given two choice problems:

Problem 1

Option A) Receive $4,000 with probability 0.8 (and $0 with probability 0.2)

Option B) Receive $3,000 with certainty

Problem 2

Option A') Receive $4,000 with probability 0.20 (and $0 with probability 0.8)

Option B') Receive $3,000 with probability 0.25 (and $0 with probability 0.75)

The options in Problem 2 are obtained by mixing each option in Problem 1 with a 0.75 probability of receiving $0. That is, the probability of receiving a positive sum in Problem 2 is exactly one fourth of the probability in Problem 1. Since mixing two lotteries with the same common lottery should not change a person's preference ranking, the only strict preference patterns consistent with rational choice theory are (A and A') and (B and B'). Yet, (B and A') is a robust modal choice pattern (Allais, 1953, Kahneman & Tversky, 1979).

A number of qualitatively different psychological explanations of the common ratio effect have been advanced including non-linear probability weighting (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992), regret aversion (Loomes & Sugden 1982; Bell, 1982), reliance on similarity judgments (Rubinstein 1988, Leland, 1994) or salience perceptions (Bordalo et al., 2012), and dual system models (Mukherjee 2010; Schneider & Coulter, 2015). The predictions of these theories across the tasks in our experiment (when the theories are calibrated to explain the modal responses from Kahneman and Tversky, 1979) are summarized in Table 1. We briefly review how each of these approaches predicts the preferences for B and A' in the choice task and then discuss the predictions of each approach in the tasks in our study.

Under the probability weighting explanation, people systematically underweight high probabilities and overweight low probabilities (Kahneman and Tversky, 1979).[4] This approach explains the common ratio effect if (i) the 80% chance of $4000 is underweighted such that the perceived (distorted) probability is less than 0.80, and (ii), the 20% chance of $4000 is over-weighted such that the perceived probability of 0.20 is close to the perceived probability of 0.25.

An emotion-based account of the common ratio effect is given by regret theory. Loosely speaking, a decision maker may anticipate regretting the choice of an 80% chance of $4000 if she receives $0 when she could have obtained $3000 with certainty. However in Problem 2, there is no certain money left on the table by choosing the 20% chance of $4000, and so regret plays less of a role in this choice, allowing the decision maker to take on greater risk, choosing A'. Regret approaches depend on comparing alternatives separately or jointly, not on whether the task involves pricing or rating. When evaluating options individually, regret plays no role, suggesting that relative evaluations should be consistent across different presentations of separate evaluation of alternatives. It thus seems that these approaches predict the *same* preference pattern to be observed in both individual evaluation tasks (happiness and pricing), although it is not clear a priori, which of the four preference patterns this should be.

Rubinstein (1988) and Leland (1994) offer explanations of the common ratio effect based on similarity judgments. For instance, Rubinstein argues that a decision maker essentially ignores similar attributes across alternatives and bases her choices on the less similar attribute dimension. In Problem 1, the decision maker may view $3000 and $4000 to be more similar than the difference between an 80% chance and a 100% chance of winning. The decision maker then chooses the option which performs better on the less similar probability dimension (Option B).

[4] More precisely, under cumulative prospect theory (Tversky and Kahneman, 1992), the probability weighting function exhibits a property called subproportionality, in which the ratio of probability weights decrease when both are scaled down by a common factor.

In Problem 2, however, the decision maker views probabilities of 0.20 and 0.25 to be more similar than the difference between 4000 and 3000 and so the decision maker chooses the option which performs better on the payoff dimension (Option A'). Like regret-based theories, similarity judgments cannot explain any differences between different treatments of separate evaluations.

Whereas the similarity explanation assumes people underweight or ignore small differences in attribute values, Bordalo et al. (2012) proposed a salience-based model of decision making in which people overweight or focus on large differences in attribute values. In Problem 1, the fact that one option is risky and the other option is certain is salient in the mind of the decision maker, producing a choice for the certain Option B. In Problem 2, the difference between 4000 and 3000 is more salient than the 5% difference in the probability of winning, producing a choice for Option A'. This explanation is closely related to the similarity-based account, although Bordalo et al. provide a more precise approach to modeling the perception of differences than the earlier similarity models. The salience model of Bordalo et al. (2012) assumes that when lotteries are evaluated in isolation, the decision maker compares each lottery to receiving $0 with certainty. Under the same approach used by Bordalo et al. (2012) to explain the preference reversal phenomenon (Lichtenstein & Slovic, 1971), the salience model predicts the pattern AA' to be observed when all lotteries are evaluated in isolation.

An additional class of models which explains the common ratio effect by a very different means is the class of dual system theories of affect and cognition. In particular, we consider Mukherjee's (2010) dual system model of choice under risk, and the more recent dual process evaluability framework of Schneider and Coulter (2015). In Mukherjee's dual system model, the value of a lottery is determined by a weighted average of the values assigned to the lottery by an affective system and a deliberative system. Mukherjee assumes the affective system has a concave value function for gains and that, rather than weighting probabilities, it assigns a weight of $1/n$ to each outcome, where n is the number of possible outcomes in the given lottery. Mukherjee also assumes that the deliberative system maximizes expected monetary value. If the weight on the affective system is sufficiently high, the dual system model predicts a preference for Option B in Problem 1 (since the affective system weights Option B by 1, and weights Option A by ½). In Problem 2, the dual system model predicts the choice of A' regardless of the weight on the affective system since the affective system is assumed to transform these choices into a decision between a 50% chance of 4000 (and 0 otherwise) or a 50% chance of 3000 (and 0 otherwise). Since Option A' has a higher expected value than B', both the affective system and the deliberative system are assumed to value Option A' higher than B', thereby explaining the common ratio effect.

The dual process evaluability framework (DPEF) of Schneider and Coulter (2015) integrates two streams of literature in judgment and decision making—the literature on evaluability theory and the literature on dual process models. It assumes there to be two valuation processes that a decision maker may rely on: valuation by feeling and valuation by calculation (Hsee and Rottenstreich, 2004). Valuation by calculation maximizes expected value while valuation by feeling is risk-averse, consistent with Mukherjee's assumption that the value function of the affective system is concave. In contrast to Mukherjee, DPEF assumes that choices are typically governed by one particular valuation process (feeling or calculation), and that the relative dominance of these processes depends on properties of the choice set (whether alternatives differ categorically or incrementally) and on the response mode (whether the task makes evaluation of the alternatives easy or difficult). Following Hsee and Zhang (2010), Schneider and Coulter assume that categorical differences (e.g., risk vs. no risk) are relatively easy to evaluate, whereas incremental differences (e.g., small changes in the degree of risk) are more difficult to evaluate. They link evaluability theory to dual process theory by assuming that when evaluation is easy, people systematically rely on valuation by feeling, but when evaluation is difficult, calculation is necessary to make the decision.

For a choice task, the alternatives in Problem 1 differ categorically in that Option A involves risk but Option B does not. Hence, DPEF predicts that evaluation is easy and that the choice is governed by valuation by feeling which is risk-averse, leading to the selection of the safer Option B. However, in Problem 2, the alternatives differ incrementally (probabilities are 0.20 and 0.25) in which case evaluation is more difficult and DPEF predicts valuation by calculation to dominate. Since valuation by calculation maximizes expected value, DPEF predicts the choice of the option with the higher expected value (Option A') in Problem 2.

Testing among the various explanations for the common ratio effect in a standard choice task is not diagnostic since all of these explanations make the same predictions. Our approach is to test among the theories across response modes. Explanations based on probability weighting are context-independent, predicting the same choices regardless of how the problem is framed or how preferences are elicited. Models based on regret aversion, similarity judgments, or salience perceptions are comparative in nature. While they allow for differences between separate and joint evaluation, they cannot explain differences between separate evaluation methods. In contrast, dual system approaches to decision making predict that preferences can vary systematically with tasks that elicit more affective versus deliberative processing.

With regard to dual process explanations of the common ratio effect, Mukherjee's dual system model (DSM) predicts a shift in behavior toward expected value maximization for tasks which systematically involve logical or calculation-based processes. Hence, if pricing tasks involve more 'calculation' than choice tasks, the DSM predicts more expected value maximizing behavior if the alternatives are priced in isolation, as compared with choice tasks. If the response mode systematically elicits more affective or emotional processing, the DSM predicts the common ratio effect to be observed. For instance, if an emoticon or "happiness" scale induces more affective processing than a choice task, the DSM predicts the choice of B, in Problem 1 (since the affective system weights Option A by 0.5, not by 0.8 in the DSM) and the choice of A' in Problem 2 (since the affective system weights both options by 0.5).

Under DPEF, behavior is predicted to shift toward expected value maximization in tasks which involve more calculation (similar to the prediction of the DSM), leading to the preferences of A and A' in the pricing task. However, in tasks which elicit more feeling-based processing, DPEF predicts consistent risk-averse behavior, and thus predicts preferences of B and B' in the happiness task. DPEF is the only theory of those considered that predicts different consistent choices in each of the separate evaluation response modes (as each induces a specific processing frame). A summary of the dominant process and choice pattern predicted by DPEF for each task is provided in Table 2.

3 Experimental Design

In this section we summarize the sample, design, and procedures used in our experiment. A display of the tasks from the experiment is provided in Appendix A.

3.1 Participants

A convenience sample of one-hundred fifty-eight undergraduate students at a large public New England University participated in an online survey. Participants were recruited through a daily e-mail bulletin sent to all undergraduate students in which they were asked to participate in a decision making study requiring less than 15 minutes of their time. The sample consisted of all students who responded during a three-week period at the end of the Spring 2014 semester. Three participants were randomly selected to receive a $25 gift card to the university store.

3.2 Design

The experiment involved three basic tasks for each subject: (1) a choice task in which subjects choose between options A and B and between options A' and B', (2) a pricing task in which subjects stated the minimum price at which they would sell each of the four options (when evaluated in isolation), and (3) a happiness rating task in which subjects rated each of the four lotteries in isolation on a happiness scale. The Allais gambles can be tested in happiness scales by asking subjects to indicate their feelings toward each lottery by selecting a point on a scale with endpoints of very sad and very happy emoticons. A point is selected for each of the four lotteries, where each point corresponds to a number reflecting that point's proximity to the happy emoticon (higher numbers correspond to more positive feelings). We can then rank the points assigned to the lotteries to observe which lotteries made people 'happier'.

One might view the fact that the choice task involved joint evaluation of alternatives but the pricing and happiness tasks involved separate evaluation as a confound, but we view the choice task as a control – to confirm that we observe the standard common ratio pattern where it is usually observed, and we believe that the most interesting comparisons are between behavior in the pricing and happiness tasks where alternatives are evaluated in isolation.

It is possible for subjects to be indifferent between two options in a pricing task (if the options are assigned the same price) or in a happiness task (if the options are assigned the same rating), but not in a choice task where subjects can only select one of the two options. We therefore employed two variants of the choice task – one without an indifference option, and one in which subjects could express indifference between the two options.

There were also two variants of the happiness task—one with a coarse rating scale and the other with a fine-grained scale. Emoticon scales avoid words that can anchor or bias ratings (Friedman & Amoo, 1999). The coarse-grained scale was a five-point scale. A five-point emoticon scale was also used by Shampanier, Mazar, and Ariely (2007) to gauge subjects' feelings about the value of free products. However, the coarse-grained scale is prone to overestimating the proportion of ‘indifference’ responses due to generating a potentially large number of ties in the ratings for two options. To reduce the number of ties, we also employed a fine-grained scale in which participants could slide a bar with the same endpoints of a very sad and very happy emoticon used in the coarse scale to express their rating of each alternative. The bar’s location was captured using a discretization of 2000 points.

3.3 Procedure

Each participant completed a choice task, a happiness task, and a pricing task. The order of tasks (choice, pricing, and happiness ratings) was randomized and filler questions were used between tasks. The filler questions between the first pair of tasks were the three questions in the cognitive reflection test of Frederick (2005). The filler question between the second pair of tasks was the ‘count-the-F’s’ question studied by Rubinstein (2013). Respondents were randomly assigned either the choice task with an indifference option or the choice task with no indifference option. Respondents were also randomly assigned to either the coarse-grained happiness task or the fine-grained happiness task. The order in which the alternatives appeared on the screen was also randomized in the choice task. Finally, within each task, the order of the two problems (for the choice task), and the order of the four alternatives (for both the pricing and happiness tasks) were randomized. For each task, response time was recorded to the nearest second.

In the instructions, subjects were informed, “You will be provided with several decision making problems. Please answer each question as honestly as possible.” For the choice task, subjects were instructed, “Please select your preferred option from the two alternatives listed below.” In the happiness task, subjects were instructed, “Please indicate your feelings about the offer below by selecting a point between the two pictures below,” where the pictures were images of sad and happy emoticons. In the pricing task, subjects were provided with each of the four alternatives in isolation and instructed “Suppose you have an 80% chance of winning $4,000” (with analogous text used for the other lotteries). Participants were then asked to state the minimum price at which they would sell this opportunity.

4 Results Between Subjects

The distribution of response patterns for all 158 subjects across tasks are displayed in Table 3. Subjects are categorized by the four response patterns or by the number of ties in evaluations. The distributions are strikingly different, with the modal strict preference patterns replicating the Allais common ratio effect in the choice task but revealing consistent risk aversion in the happiness task and consistent expected value maximization in the pricing task. The rightmost column in Table 3 labeled "% Consistent" shows the percent of subjects whose responses were consistent with rational choice theory for each task, either (A, A'), (B, B'), or indifference for both pairs of options. Indifference for one pair of alternatives and a strict preference ranking for the other is technically inconsistent with rational choice theory. The pricing task and the coarse happiness task each displayed a large number of indifferences as shown in the "tie" columns in Table 3. If the ties are counted as indifferences, then over 40% of responses in the coarse happiness task and over half of the responses in the pricing task were inconsistent with expected utility theory. However, for the coarse scale in particular, the ties likely reflect insufficient precision in measuring preferences rather than true indifferences since that scale has only five points of discretization. In this respect, note that the large number of indifferences observed under the coarse rating scale were not resolved randomly under the fine-grained scale, but rather shifted almost entirely in favor of the less risky alternatives (B and B'). This suggests that that ties in coarse happiness ratings represent not indifference but instead differences too small to be picked up by a five-point scale. Remarkably, over 85% of subjects in the fine-grained happiness task exhibited consistent preferences, nearly twice as high a percentage as in the choice task.

The average and median evaluations for each of the four alternatives (A, B, A', B') in the happiness tasks and the pricing task are displayed in Table 4.[5] Recall that the coarse happiness scale has only five possible ratings, whereas the fine happiness scale has a discretization of 2000 points. The median valuation for each alternative in the pricing task is equal to that alternative's expected value. In addition, while the median responses to both happiness tasks may suggest that respondents did not see alternatives A' and B' as very different, 79% of all respondents to the fine-grained happiness scale rated B' higher than A'.

[5] 54 subjects either assigned a value to Option B (3000 with certainty) that was not equal to 3000 or assigned a price of $0 to one or more options. Our results are robust to these responses: Considering the remaining 104 subjects, we observe very similar distributions of responses to those in Table 3 and very similar mean and median responses to those in Table 4 (but with higher mean prices and smaller price standard deviations).

4.1 Distribution of Strict Preference Patterns

We refer to a preference pattern as *strict* if it does not include a tie between either pair of options. Figure 1 displays the distribution of strict preference patterns (as a proportion of all strict preferences) for the happiness task with the fine-grained scale, the choice task with an indifference option, and the monetary valuation (pricing) task. As can be seen from Figure 1, the distribution of response patterns differs remarkably across different tasks. The overwhelming pattern (86.3% of all strict preferences) in the happiness task was in favor of the risk-averse alternatives (B, B'). The modal pattern (50.7%) in the choice task was the Allais common ratio pattern (B, A') observed by Kahneman and Tversky (1979). In the pricing task, the modal response pattern (45%) corresponded to the alternatives that maximize expected value (A, A').

In Figure 1, differences between choices and prices are highly significant for both the modal preference pattern for choices and for the modal preference pattern for prices (both $p < 0.001$, two-tailed Z difference in proportions test). Comparisons between choices and happiness ratings are also highly significant for both the modal pattern for choices and for the modal pattern for happiness ratings (both $p < 0.001$, two-tailed Z difference in proportions test). In addition, comparisons between prices and happiness ratings are highly significant for both the modal pattern for prices and for the modal pattern for happiness ratings (both $p < 0.001$, two-tailed Z difference in proportions test).

5 Within-Subject Differences across Response Modes

The design of the experiment enables us to also make inferences within subjects across choice, pricing, and happiness tasks. To the extent that respondents try to be consistent with their earlier responses when evaluating the same set of alternatives, our design biases results against the hypothesis that choices, prices, and happiness ratings differ since each respondent evaluated the same alternatives in each type of task in the same survey. Yet strong reversals in preferences were nevertheless observed within subjects.

The dual process evaluability framework (DPEF) of Schneider & Coulter (2015) hypothesizes that valuation by feeling predominates over happiness rating tasks and that valuation by calculation predominates over pricing tasks. This implies a novel pattern of response mode reversals. In particular, these hypotheses are consistent with risk-averse preferences for the happiness tasks (since both tasks are predicted to be governed by feeling), expected value maximization for the pricing tasks (since both tasks are predicted to be governed by calculation), and the Allais common ratio pattern for the choice task (since Problem 1, which involves certainty, elicits more affective processing than Problem 2).

The modal response patterns across response modes observed in the experiment are displayed in Table 5. In each case, the modal response pattern was the one predicted by DPEF. Table 5 also includes the proportion of respondents exhibiting each modal response pattern out of all preference patterns for a given pair of tasks. We see that in each case, the modal response pattern captured at least 35% of all preference patterns, and that four of the six modal response patterns involved preference reversals *within subjects*.

We also briefly consider within-subject responses across all three tasks simultaneously. This allows for 729 different response patterns when ties are considered, and no one response pattern accounted for the bulk of subjects. The modal response pattern, however, was the one predicted by DPEF, with 11 subjects who each replicated the Allais pattern in the choice task, maximized expected value in the pricing task, and exhibited consistent risk aversion in the happiness task. In contrast, consistent risk-aversion across all three tasks was observed by only three subjects. Surprisingly, *none* of the 158 respondents in our experiment displayed the Allais preference pattern simultaneously in choice, pricing, and happiness tasks.

Finally, we examined the correlations within and between response modes. Both prices for each pair of alternatives are highly correlated (correlation coefficient between prices for A and B is 0.430, $p < 0.001$; correlation between prices for A' and B', is 0.452, $p < 0.001$) as are happiness ratings for each pair of alternatives (correlation between fine happiness scale ratings for A and B, is 0.440, $p < 0.001$; correlation between fine happiness scale ratings for A' and B' is 0.9408, $p < 0.001$), suggesting that people who value one option higher value other options higher, and similarly for happiness ratings. However, prices and happiness ratings, even for the same alternative, appear uncorrelated (the correlation between prices and happiness ratings for A is 0.077, $p = 0.515$; the correlation for B is 0.087, $p = 0.464$; the correlation for A' is -0.188, $p = 0.109$; the correlation for B' is -0.010, $p = 0.935$). This interesting result is consistent with the idea that subjects approach the two tasks very differently. That is, happiness is not merely a proxy for price (or vice versa), even though the responses are provided by the same subjects for the same alternatives in the same survey.

5.1 Data from Cognitive Reflection Questions

To examine what may account for differences between subjects, the three common ratio tasks were separated by filler questions involving the three-question cognitive reflection test (CRT, Frederick, 2005) and the 'Count-the-F's' question in Rubinstein (2013). All four of these questions are shown in the screen shots in Appendix A. As all four questions require cognitive reflection to answer correctly, we used the responses to these questions to construct a four-question cognitive reflection index which we denote CRT*. The average score on this four-question test was 2.10 with a standard deviation of 1.36. The average score on the original three-question CRT was 1.55 with a standard deviation of 1.16. As the CRT is designed to measure a person's natural tendency to use intuitive versus rational processes (Frederick, 2005), we use this measure to understand the relative importance of natural tendencies and the prompting implied in different response modes.

Table 6 displays the consistency of responses both within and across tasks for the high CRT* subjects (those scoring 3 or 4 out of 4, N = 69 for Choice and Pricing; N = 30 for fine happiness scale) and for the low CRT* subjects (those scoring 0 or 1 out of 4, N = 54 for Choice and Pricing; N = 27 for fine happiness scale). Choices are consistent *within* a task if they are either AA', BB', or a tie for both pairs of alternatives. Responses are consistent *across* tasks, for a given pair of tasks, if the same preference pattern was revealed in both tasks. While both high and low CRT* subjects displayed moderate to high levels of consistency within a given task, both groups also exhibited substantial response mode effects, displaying different preference patterns for identical alternatives across tasks. Essentially, high and low CRT* subjects showed similar behavior in the experiment. In fact, none of the six comparisons in Table 6 between high and low CRT* subjects is statistically significant.

6 How do we know if the Process is Really Feeling or Calculation?

We observe that the distribution of response patterns differs systematically across tasks for the same set of alternatives evaluated by the same subjects in the same survey. This seems to strongly suggest that different decision processes are engaged across tasks. We do not claim that our results confirm the underlying processes are feeling or calculation-based, but our results appear to be supportive of this hypothesis. In this section, we consider two other factors which may be used to infer shifts in feeling-based processing versus calculation-based processing: expected value calculations and response time.

By simply counting the number of subjects who priced all four lotteries at their expected values, we can observe whether at least some respondents were unambiguously 'calculating' in the pricing task. In this regard, 22 respondents priced all four lotteries at exactly their expected values and this was both the modal and the median response pattern in the pricing task.

A second factor which may provide some insight into the underlying process is the response time, both within and across tasks, since feeling-based processes operate more quickly than calculation-based processes. Response times were recorded to the nearest second in the online survey. Table 7 displays the median and average response times for each task. Notice that the average response times were all between 7.5 and 10.5 seconds for rating each alternative in the happiness task and were all between 18 and 22 seconds for valuing each alternative in the pricing task. This is consistent with the hypothesis that a common process was used in all happiness tasks and that a common process was used in all pricing tasks, but that different processes were used for happiness and pricing tasks. Indeed, it is striking that the average response time for the pricing task was approximately twice as long as the average response time for the happiness task when evaluating each of the four alternatives. The choice task revealed more heterogeneity in response times, but in a systematic way: the average response times for respondents who chose the expected value maximizing options were longer than for respondents who chose the risk-averse options. In particular, average response times were 17.65 seconds and 14.84 seconds for subjects choosing the expected value maximizing options and were 10.81 seconds and 11.33 seconds for subjects choosing the risk-averse options, which are closer to the average response times for the happiness tasks. The response time data (to the extent that it reflects subjects' decision-making processes) is roughly consistent with the DPEF hypotheses about the relative dominance of feeling versus calculation across tasks. Note however that the mean and median response times for the expected value maximizing choice in Problem 2 were closer to the mean and median response times in the happiness rating task than in the pricing task, contrary to the prediction that this choice involved calculation-based processes for a majority of subjects. In addition, the standard deviations in response times were fairly large, reflecting a large degree of heterogeneity in response times. Thus, while the modal responses appear consistent with DPEF, the timing data suggest that at least some of these responses may be for different reasons than DPEF hypothesizes.

7 Discussion

The distribution of response patterns summarized in Figure 1 presents a new paradox for theories of choice under risk. Since many normative and descriptive models of decision making are procedure invariant, including expected utility theory, rank-dependent utility theory (Quiggin 1982), and cumulative prospect theory (Tversky & Kahneman, 1992), these models predict that a given decision maker will have the same preference ordering revealed under each of the three response modes. As can be seen from Figure 1, the prediction of the same preference pattern across tasks is strongly rejected. Moreover, although the probability weighting explanation as formalized by rank dependent utility and cumulative prospect theory is the dominant explanation of the common ratio effect in the literature, none of our 158 subjects exhibited the Allais pattern across all three tasks simultaneously, contrary to the predictions of probability weighting and all other absolute evaluation models.

The predictions of regret aversion, similarity judgments, and salience perceptions are also not consistent with our results, since these approaches predict that the common ratio effect does not depend on pricing versus happiness rating but rather on joint versus separate evaluation. Both pricing and happiness rating tasks involved separate evaluation, but produced very different response patterns.

One might think that Mukherjee's (2010) dual system model can explain our observed behavior across tasks, but this is not the case. The DSM *is* consistent with our finding of greater expected value maximization in pricing tasks. This observation naturally follows if the calculation-based, deliberative system is more influential in pricing tasks than in choice. However, as noted in Table 1, the DSM predicts the Allais pattern in the happiness tasks where instead we observed consistent risk-averse behavior.

Our findings can, however, be largely explained by the dual process evaluability framework (DPEF) of Schneider and Coulter (2015). DPEF predicts risk aversion in the happiness task, expected value maximization in the pricing task, and the Allais common ratio pattern in the choice task. These are the modal response patterns we observed, both between and within subjects. However, the evidence in support of DPEF based on response times appears mixed. Mean and median response times to the happiness task were noticeably faster than the corresponding response times to the pricing task. However, the mean and median response times to the risk-averse and expected value maximizing choices in Problem 2 did not vary widely for the choice task, which does not provide a strong indication that different processes were used in that task between those subjects who selected A' and B'.

Taking a broader of perspective of the common ratio effect, it is quite likely that multiple factors determine the effect. An alternative approach to testing explanations of the common ratio effect would be to change the 'frame' of the decision, rather than changing the response mode. Experimental studies (Harless 1992; Harman & Gonzalez, 2015; Incekara-Hafalir & Stecher, 2012) have found that the common ratio effect (and the related common consequence effect) are susceptible to whether the options are presented in an Allais-type format, or a Savage matrix (Savage, 1954). This behavior is consistent with perceptual (i.e., similarity and salience) based explanations, but not with the other explanations discussed here. Taking both the response mode and framing variations of the common ratio effect into account, it seems that none of the currently available alternatives provides a complete explanation of the common ratio effect.

One explanation that accounts for both the response mode and frame-dependencies of the common ratio effect is based on Kahneman's (2003) framework which distinguishes between three systems: perception, System 1, and System 2. One could imagine that the perceptual system recommends the choice alternative that 'looks better,' System 1 prefers the option which 'feels better,' and System 2 prefers the option which is justified as better through a logical reasoning process. If the perceptual system dominates in the choice task (since it is comparative), if System 1 dominates in the happiness task (since it may involve affect), and System 2 dominates in the pricing task (since it involves calculation), then this framework directly explains both the response mode and framing versions of the common ratio effect. This is essentially a hybrid explanation in which the similarity and salience judgments operate in choice, but feeling and calculation operate when evaluating options in isolation (in which case comparison between alternatives is more difficult). However appealing this explanation may be, it is admittedly speculative and post-hoc. Future work is needed to elucidate the relationship between response mode and framing effects in decision making.

References

Allais, M. (1953). Le comportement de l'homme rationnel devant le risqué: Critique des postulats et axioms de l'ecole Americaine. *Econometrica* 21 (4) 503-546.

Ballinger, P., & Wilcox, N. (1997). Decisions, error and heterogeneity" *Economic Journal.* 107, 1090-1105.

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

Baucells, M., & Heukamp, F.H. (2010). Common ratio using delay. *Theory and Decision*. 68, 149-158.

Blavatskyy, P.R. (2010). Reverse Common Ratio Effect. *Journal of Risk and Uncertainty*. 40, 219-241.

Bordalo, P., N. Gennaioli, & Shleifer, A. (2012). Salience theory of choice under risk. *The Quarterly Journal of Economics* 127 (3) 1243-1285.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19,* 25-42.

Friedman, H.H., & Amoo, T. (1999). Rating the rating scales. *J. Marketing Mgmt*. 9(3) 114-123.

Harless, D. W. (1992). Actions versus prospects: The effect of problem representation on regret. *The American Economic Review*, 634-649.

Harman, J.L., & Gonzalez, C. (2015). Allais from experience: Choice consistency, rare events, and common consequences in repeated decisions. *Journal of Behavioral Decision Making, 28,* 369-381.

Hsee, C.K., & Rottenstreich Y. (2004). Music, pandas, and muggers: On the affective psychology of value. *Journal of Experimental Psychology General, 133,* 23-30.

Hsee, C. K., & Zhang, J. (2010). General evaluability theory. *Perspectives on Psychological Science*, *5*(4), 343-355.

Incekara-Hafalir, E. & Stecher, J.D. (2012). An experimental test of theories of behavior in Allais-type tasks. Manuscript, Carnegie Mellon University.

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American economic review*, 1449-1475.

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

Leland, J. W. (1994). Generalized similarity judgments: An alternative explanation for choice anomalies. *Journal of Risk and Uncertainty*, 9(2), 151-172.

Lichtenstein, S., & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology* 89(1) 46-55

Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal* 92 805-824.

Loomes, G., & Sugden, R. (1998). Testing different stochastic specifications of risky choice. *Economica.* 65, 581-598.

Mukherjee, K. (2010). A dual system model of preferences under risk. *Psychological Review,* 117, 243-255.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior and Organization*. 3(4) 323-343.

Rubinstein, A. (1988). Similarity and decision-making under risk (Is there a utility theory resolution to the Allais paradox?). *Journal of Economic Theory* 46 145-153.

Rubinstein, A. (2013). Response time and decision making: An experimental study. *Judgment and Decision Making,* 8(5), 540-551.

Savage, L. J. (1954). *The Foundations of Statistics*. New York, Wiley.

Schneider, M. & Coulter, R.A. (2015). A dual process evaluability framework for decision anomalies. *Journal of Economic Psychology*, 51, 183-198.

Shampanier, K., Mazar, N., & Ariely, D. (2007). Zero as a special price: The true value of free products. *Marketing Science*. 26(6) 742-757.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5 297-323.

Table 1. Predictions of Cognitive Explanations of the Common Ratio Effect across Tasks

Explanation	Choice Task Prediction	Pricing Task Prediction	Happiness Prediction
Probability Weighting	BA′	BA′	BA′
Regret Aversion	BA′	Same pattern in pricing and happiness tasks	
Similarity Judgments	BA′	Same pattern in pricing and happiness tasks	
Salience Perception	BA′	AA′	AA′
Dual System Model	BA′	AA′	BA′
Dual Process Evaluability	BA′	AA′	BB′

Table 2. Predictions of the Dual Process Evaluability Framework (DPEF) across Tasks

Task	Valuation Process	Predicted Choices
Happiness Ratings	Valuation by feeling	Risk Aversion (BB')
Choice Task, Problem 1	Valuation by feeling	Risk Aversion (B)
Choice Task, Problem 2	Valuation by calculation	Expected Value Maximization (A')
Pricing Task	Valuation by calculation	Expected Value Maximization (AA')

Table 3. Distribution of responses

Task	RA (B,B’)	EV (A,A’)	CR (B,A’)	RCR (A,B’)	One tie	Two ties	N	% Consistent
Choice w/o Indifference	28	7	42	2	N/A	N/A	79	44.3%
Choice w/Indifference	25	5	33	2	12	2	79	40.5%
Happiness (coarse)	18	0	2	0	33	31	84	58.3%
Happiness (fine)	57	2	5	2	3	5	74	86.4%
Pricing	19	45	22	14	54	4	158	43.0%

RA = risk aversion (B, B′); EV = expected value maximization (A, A′); CR = common ratio pattern (B, A′); RCR = reverse common ratio pattern (A, B′); N = number of subjects assigned to each task; % Consistent is the proportion of subjects in each task who exhibited preference patterns RA, EV, or two ties.

.

Table 4. Responses across tasks for all subjects

Happiness (coarse)	Option A 4000, 0.8	Option B 3000, 1	Option A’ 4000, 0.20	Option B’ 3000, 0.25
Median rating	4	5	3	3
Mean rating	4.43	4.90	3.02	3.33
Standard Deviation	0.57	0.48	1.08	0.97
Happiness (fine)	**4000, 0.8**	**3000, 1**	**4000, 0.20**	**3000, 0.25**
Median rating	1836	2000	1500	1515
Mean rating	1799	1956	1451	1507
Standard Deviation	158	113	270	256
Pricing	**4000, 0.8**	**3000, 1**	**4000, 0.20**	**3000, 0.25**
Median price	3200	3000	800	750
Mean price	2728	2887	889	874
Standard Deviation	1400	981	826	663

Table 5. Within-Subject Modal Response Patterns[a]

Response Modes	Choice Set	Modal Response	Proportion [b]	N	Total
Choice vs. Pricing	Problem 1	B (Choice), A (Pricing)	0.411	65	158
Choice vs. Pricing	Problem 2	A'(Choice), A'(Pricing)	0.373	59	158
Choice vs. Happiness	Problem 1	B (Choice), B (Happiness)	0.810	60	74
Choice vs. Happiness	Problem 2	A'(Choice), B'(Happiness)	0.378	28	74
Happiness vs. Pricing	Problem 1	B (Happiness), A (Valuation)	0.378	28	74
Happiness vs. Pricing	Problem 2	B'(Happiness), A' (Valuation)	0.351	26	74

[a] The happiness response mode corresponds to the happiness task with the fine grained scale.

[b] This column displays the proportion of respondents exhibiting the modal response pattern (N) out of all response patterns (Total) for a given pair of tasks and a given choice set.

Table 6. Consistency of Responses for High and Low Cognitive Reflection Subjects*

% Consistent within tasks**	Choice	Happiness (fine scale)	Pricing
High CRT*	0.435	0.900	0.493
Low CRT*	0.389	0.778	0.352

% Consistent across Tasks***	Choice vs. Pricing	Choice vs. Happiness	Pricing vs. Happiness
High CRT*	0.174	0.367	0.033
Low CRT*	0.185	0.296	0.074

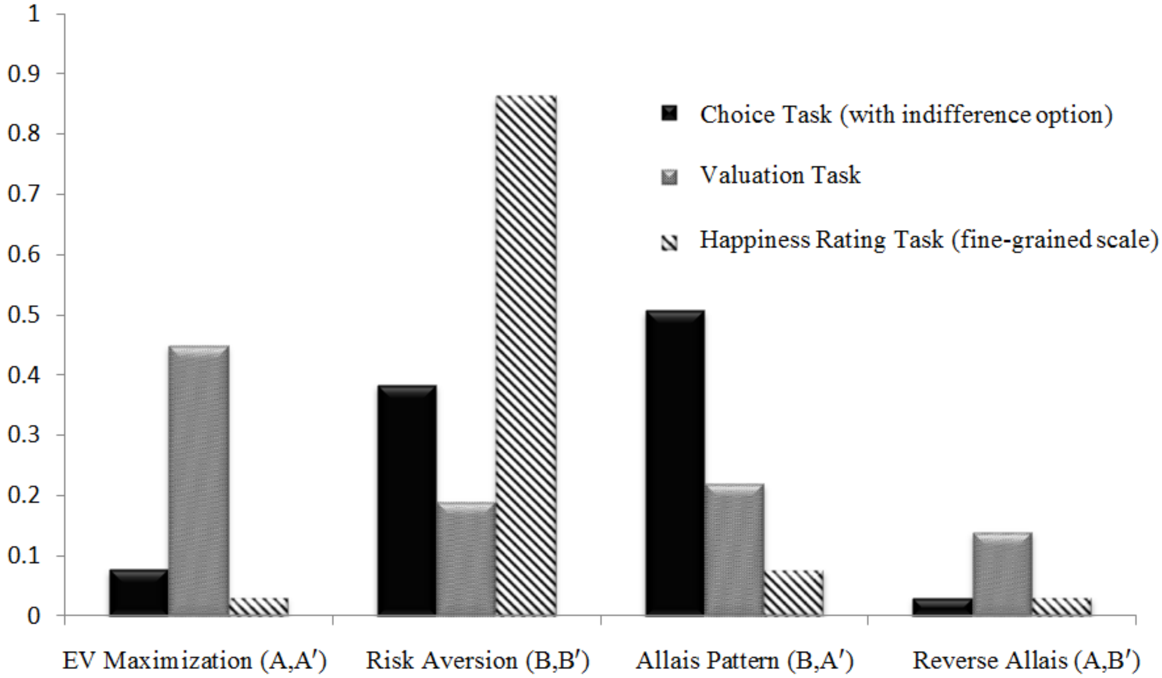
The High CRT group includes all subjects who scored a 3 or 4 (N = 69 for Choice and Pricing; N = 30 for fine happiness scale) out of the four cognitive reflection questions (the three CRT questions and the 'Count the F's' problem). The Low CRT* group includes all subjects who scored a 0 or 1 (N = 54 for Choice and Pricing; N = 27 for fine happiness scale) in total on the same four questions. **Responses are consistent *within* a task if they are either AA', BB', or a tie for both pairs of alternatives. ***Responses are consistent *across* tasks, if for a given subject and a given pair of tasks, the same preference pattern is revealed in both tasks.

Table 7. Response Times Across Tasks (seconds)[a]

Task	A (4000, 0.8)	B (3000, 1)	A' (4000, 0.2)	B' (3000, 0.25)
Happiness				
Median	7	6	7	7
Mean	10.42	7.53	9.57	9.92
Standard Deviation	9.86	5.17	7.06	8.35
Choice[b]				
Median	17	8	10	9
Mean	17.65	10.81	14.84	11.33
Standard Deviation	10.97	8.97	13.88	5.40
Pricing				
Median	16	16	16	14
Mean	20.63	21.77	21.82	18.83
Standard Deviation	14.23	15.76	14.93	14.42

[a] Response times were recorded to the nearest second. Response times greater than 1 minute were truncated to 1 minute to reduce the influence of outliers without skewing the results. Their inclusion does not change any of the medians by more than one second, but it inflates the means and standard deviations. [b] The choice response times are for the respondents who chose the corresponding alternative. The happiness rating and pricing response times are computed across all respondents for each alternative. This table pools the response times for both happiness tasks as well as the response times for both choice tasks.

Figure 1. Distribution of Strict Preference Patterns observed in the Experiment.



Appendix A: Screen Shots for Experiment

Instructions

You will be provided with several decision making problems. Please answer each question as honestly as possible. When you are finished, please be sure to log out of the study.

Please continue to the next screen to begin the survey.

Proceed to Experiment

Question 1 of 14

Please select your preferred option from the two alternatives listed below.

○ Receive $3,000 with probability, 0.25

○ Receive $4,000 with probability, 0.20 ○ I have no preference

Submit

Question 2 of 14

Please select your preferred option from the two alternatives listed below.

○ Receive $4,000 with probability, 0.80

○ Receive $3,000 with certainty ○ I have no preference

Submit

Question 3 of 14

A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?

________ cents

Submit

Question 4 of 14

In a lake there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

[ ] days

Submit

Question 5 of 14

If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?

[ ] minutes

Submit

Question 6 of 14

Suppose that you have a 20% chance of winning $4000.
Please state the minimum price at which you would be willing to sell this opportunity.

$

Submit

Question 7 of 14

Suppose that you have a 100% chance of winning $3000.
Please state the minimum price at which you would be willing to sell this opportunity.

$

Submit

Question 8 of 14

Suppose that you have an 80% chance of winning $4000.
Please state the minimum price at which you would be willing to sell this opportunity.

$

Submit

Question 9 of 14

Suppose that you have a 25% chance of winning $3000.
Please state the minimum price at which you would be willing to sell this opportunity.

$

Submit

Question 10 of 14

Count the number of appearances of the letter F in the following 80-letter text.

FINISHED FILES ARE THE RESULT OF YEARS
OF SCIENTIFIC STUDY COMBINED WITH THE
EXPERIENCE OF YEARS.

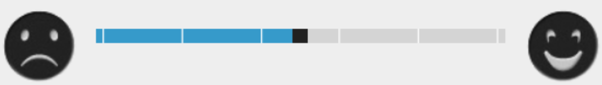
Number of appearances:

Submit

Question 11 of 14

Please indicate your feelings about the offer below by selecting a point between the two pictures below.

Receive $4,000 with probability, 0.80

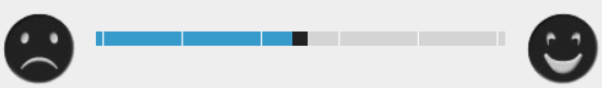


Submit

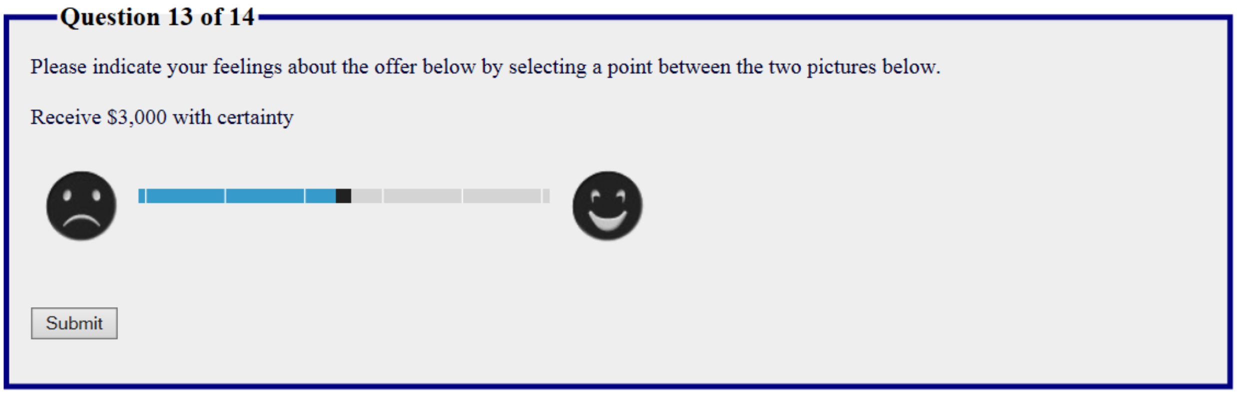
Question 12 of 14

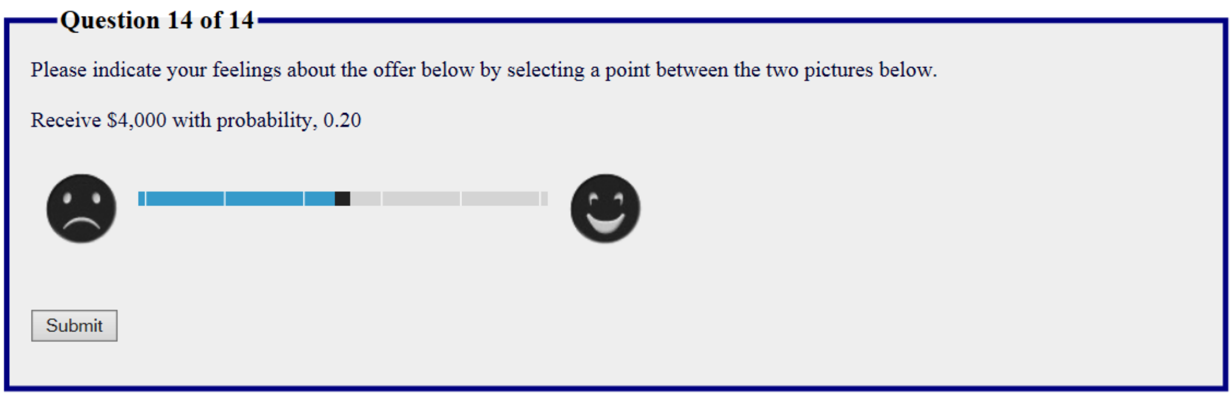
Please indicate your feelings about the offer below by selecting a point between the two pictures below.

Receive $3,000 with probability, 0.25



Submit


Question 13 of 14
Please indicate your feelings about the offer below by selecting a point between the two pictures below.
Receive $3,000 with certainty
Submit


Question 14 of 14
Please indicate your feelings about the offer below by selecting a point between the two pictures below.
Receive $4,000 with probability, 0.20
Submit