Journal of Behavioral Decision Making, *J. Behav. Dec. Making*, **30**: 976–986 (2017)
Published online 16 May 2017 in Wiley Online Library (wileyonlinelibrary.com) **DOI:** 10.1002/bdm.2017

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

MARK SCHNEIDER[1]* and MIKHAEL SHOR[2]
[1]*Economic Science Institute, Chapman University, Orange, CA USA*
[2]*Department of Economics, University of Connecticut, Storrs, CT USA*

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 large shifts in preference patterns across tasks, both within and between subjects. In particular, we find that both the consistency and distribution of responses differ 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. Copyright © 2017 John Wiley & Sons, Ltd.

KEY WORDS common ratio effect; preference reversals; dual processes; happiness ratings

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. As the effect was introduced by Allais (1953) and popularized by Kahneman and Tversky (1979), it has been replicated in numerous studies (e.g., Ballinger & Wilcox, 1997; Barron & Erev, 2003; Baucells & Heukamp, 2010; Loomes & Sugden, 1998)[1] and has served as a motivating example for many models of choice under risk including prospect theory (Kahneman & Tversky, 1979), regret theory (Loomes & Sugden, 1982), similarity theory (Leland, 1994; Rubinstein, 1988), and salience theory (Bordalo, Gennaioli, & Shleifer, 2012).

Despite the widespread observation of the common ratio effect when experimental subjects provide *choices* between lotteries, the effect has surprisingly not been investigated for its robustness in other 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, 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., Bell, 1982; Bordalo et al., 2012; Leland, 1994; Loomes & Sugden, 1982; Rubinstein, 1988), which predict that preferences can depend on joint versus separate evaluation of choice alternatives, but do not provide predictions for how choices might vary across response modes; and (iii) dual process models (Mukherjee, 2010; Schneider & Coulter, 2015), which predict that tasks that 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 monetarily value each alternative, and happiness ratings, which asks subjects for a subjective assessment for each alternative. We find that the dual process evaluability framework (DPEF) of Schneider and Coulter provides the most complete account of our results and predicts the modal preference patterns across tasks.

We proceed by first reviewing the common ratio effect and a number of its leading explanations. 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 theories to explain our results.

THE ALLAIS COMMON RATIO EFFECT

Virtually every alternative to expected utility theory—the standard model of rational decision making under risk—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:

*Correspondence to: Mark Schneider, Economic Science Institute, Chapman University, One University Drive, Orange, CA 92866, USA. E-mail schneide@chapman.edu
†The authors thank the editor and two anonymous referees for insightful comments that helped to greatly improve the paper. All errors are our own.

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

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′	Same pattern in pricing and happiness tasks	
Dual system model	BA′	AA′	BA′
Dual process evaluability	BA′	AA′	BB′

Problem 1

Option A Receive \$4000 with probability .8 (and \$0 with probability .2)

Option B Receive \$3000 with certainty

Problem 2

Option A′ Receive \$4000 with probability .20 (and \$0 with probability .8)

Option B′ Receive \$3000 with probability .25 (and \$0 with probability .75)

The options in Problem 2 are obtained by mixing each option in Problem 1 with a .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. As mixing two lotteries with the same common lottery should not (per expected utility theory) change a person's preference ranking, the only strict preference patterns consistent with the 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 & Tversky, 1979; Tversky & Kahneman, 1992), regret aversion (Bell, 1982; Loomes & Sugden, 1982), reliance on similarity judgments (Leland, 1994; Rubinstein, 1988) or salience perceptions (Bordalo et al., 2012), and dual system models (DSMs; Mukherjee, 2010; Schneider & Coulter, 2015). 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. The predictions of these theories across the tasks in our experiment are summarized in Table 1.

Under the probability weighting explanation, people systematically underweight high probabilities and overweight low probabilities (Kahneman & Tversky, 1979).[2] 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 .80 and (ii) the 20% chance of \$4000 is overweighted such that the perceived probability of .20 is closer to .25.

An emotion-based account of the common ratio effect is given by regret theory. A decision maker may anticipate regretting the choice of an 80% chance of \$4000 if she or he receives \$0 when she or he could have obtained \$3000 with certainty. However, in Problem 2, there is no certain money left on the table, and so regret plays less of a role in this choice.

Rubinstein (1988) and Leland (1994) offer explanations of the common ratio effect on the basis of similarity judgments. For instance, Rubinstein argues that a decision maker essentially ignores similar attributes across alternatives and bases her or his 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 that performs better on the less similar dimension (Option B). In Problem 2, however, the decision maker views probabilities of .20 and .25 to be more similar than the difference between \$4000 and \$3000, and so the decision maker chooses the option that performs better on the payoff dimension (Option A′).

Related to the idea of underweighting attributes with similar values, Bordalo et al. (2012) propose a salience-based model of decision making in which people focus on attributes with large differences in 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′.

A class of models that 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 (DSM) of choice under risk, and the more recent Dual Process Evaluability Framework (DPEF) of Schneider and Coulter (2015). In Mukherjee's DSM, 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 that the affective system has a concave value function for gains and 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 DSM predicts a preference for Option B in Problem 1. In Problem 2, the DSM predicts the choice of A′ regardless of the weight on

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

DOI: 10.1002/bdm

the affective system because 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). As 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 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 the existence of two valuation processes that a decision maker may rely on: valuation by feeling and valuation by calculation (Hsee & 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 easier to evaluate than incremental differences (e.g., small changes in the degree of risk). 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, risk-neutral calculation is relied on 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 .20 and .25) in which case evaluation is more difficult and DPEF predicts valuation by calculation to dominate. As 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 because 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 do not make clear predictions for how individual evaluation of alternatives will differ between pricing and rating tasks, at least not without additional *ad hoc* assumptions. For instance, salience theory (Bordalo et al., 2012) explains reversals between choice and pricing tasks based on the presentation mode, not based on the response mode. Bordalo et al. comment: "In our model, choosing and pricing are the same operation" (p. 1273). In order to explain preference reversals between choice and pricing tasks, the authors make the assumption that if a salience-based agent "is asked to price a lottery in isolation, this approach suggests that he evaluates it together with the alternative of not having the lottery, namely, having zero for sure" (p. 1271). While Bordalo et al. (2012) do not consider ratings, if one makes the analogous assumption that a lottery rated in isolation is compared with having zero for sure, then the salience-based model predicts the same preference ranking for pricing and for rating.

The main feature of similarity theory, salience theory, and regret theory is that they are comparative models. Thus, it seems plausible and even natural within these frameworks that they predict reversals between joint and separate evaluations, as changing the presentation mode (joint vs. separate) changes what is being compared (e.g., comparing the two lotteries directly or comparing each lottery to having \$0 for sure). As in the case of the salience models, the similarity and regret models also require additional ad hoc assumptions not present in the basic principles of these frameworks in order to explain differences in pricing and rating tasks. For this reason, we indicate in Table 1 that these approaches make the same prediction for rating and pricing tasks under the same evaluation mode, as there is no a priori implication in these frameworks that suggests that preferences in these cases will differ.

Regarding the other predictions in Table 1, probability weighting models such as prospect theory or rank-dependent utility theory (Quiggin, 1982) make the same prediction independent of the response mode and presentation mode and so cannot explain preference reversals between choice and pricing tasks (Bordalo et al., 2012) or between these tasks and rating tasks. Dual process models seem to more clearly predict differences between pricing and happiness ratings to the extent that these response modes elicit different psychological processes (with pricing being more cognitive or calculation based and happiness ratings being more affective). The differences in prediction between the DSM and the DPEF are not in which process governs happiness rating or pricing but rather what behavior is predicted by the affective process. Both dual process models predict the calculation-based processes to maximize expected value, but the DSM predicts affective processes to yield the common ratio effect, whereas the DPEF predicts affective processes to produce consistent risk aversion. More precisely, Mukherjee's DSM predicts a shift in behavior toward expected value maximization for tasks that 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 (as the affective system weights Option A by .5, not by .8 in the DSM) and the choice of A′ in Problem 2 (as the affective system weights both options by .5).

DOI: 10.1002/bdm

Under DPEF, behavior is predicted to shift toward expected value maximization in tasks that 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 that 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 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.

EXPERIMENTAL DESIGN

Participants

A convenience sample of 158 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 3-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.

Design

The experiment involved three basic tasks for each subject: (i) a choice task in which subjects choose between options A and B and between options A′ and B′; (ii) a pricing task in which subjects stated the minimum price at which they would sell each of the four options, evaluated in isolation; and (iii) a happiness rating task in which subjects rated each of the four lotteries in isolation on a happiness scale 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. We believe that the most interesting (and most diagnostic) 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.

We also employed 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 5-point scale. A 5-point emoticon scale was also used by Shampanier, Mazar, and Ariely (2007) to gauge subjects' feelings about the value of free products. However, the 5-point scale is prone to overestimating the proportion of "indifference" responses owing 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. Screenshots of experimental tasks are provided in Appendix A.

Procedure

Each participant completed a choice task, a happiness task, and a pricing task. The order of tasks 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 (CRT) 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

Table 2. Predictions of the dual process evaluability framework across tasks

Task	Valuation process	Predicted choices
Happiness ratings	Valuation by feeling	Risk aversion (BA′)
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′)

DOI: 10.1002/bdm

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 without indifference	28	7	42	2	N/A	N/A	79	44.3
Choice with 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, the proportion of subjects in each task who exhibited preference patterns RA, EV, or two ties.

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.

RESULTS BETWEEN SUBJECTS

The distributions of response patterns for all 158 subjects across tasks are displayed in Table 3. Subjects are categorized by the four strict preference 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 the rational choice theory for each task, either (A and A′), or (B and B′), or indifference for both pairs of options. Indifference for one pair of alternatives and a strict preference ranking for the other are technically inconsistent with the 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 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, ties likely reflect insufficient precision in measuring preferences rather than true indifferences as that scale has only five points of discretization. In this respect, note that the large number of indifferences observed under the coarse rating scale was 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 5-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.

Our main result is the large shifts in subject-level preferences in the common ratio effect across tasks. Regardless of the precise ratings and prices assigned by subjects, we observe a very consistent shift in subject-level preferences with modal choices replicating the Allais paradox, modal happiness ratings producing consistent risk aversion, and modal prices consistent with expected value maximization. We include the precise mean and median ratings and valuations for each of the four alternatives (A, B, A′, and B′) as a general description of the data in Table 4.[3] However, we emphasize that the main sources of support for our conclusions are the large and systematic shifts in subject-level responses, not the aggregate values in Table 4.

For the data summarized in Table 4, 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,

Table 4. Summary statistics for happiness and pricing tasks

	Option A $4000, .8	Option B $3000, 1	Option A′ $4000, .20	Option B $3000, .25
	Happiness (coarse)			
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)			
Median rating	1836	2000	1500	1515
Mean rating	1799	1956	1451	1507
Standard deviation	158	113	270	256
	Pricing[a]			
Median price	3200	3000	800	750
Mean price	2925	3000	891	881
Standard deviation	788	0	699	520

[a]Mean prices and standard deviations of prices are rounded to the nearest dollar.

[3]Table 4 displays the pricing data for the 103 subjects in our experiment who did not value the lottery more than its maximum possible outcome or less than its worst possible outcome, consistent with the internality axiom (Gneezy, List, & Wu, 2006), and who assigned a price greater than zero to all four lotteries, viewing these as basic criteria for quality responses to the open-ended pricing task. Our results are robust to including the additional 55 subjects who violated at least one of these criteria in at least one case, and we observe qualitatively similar responses to those in Table 4 (but with lower mean prices and larger price standard deviations) when including them.

DOI: 10.1002/bdm

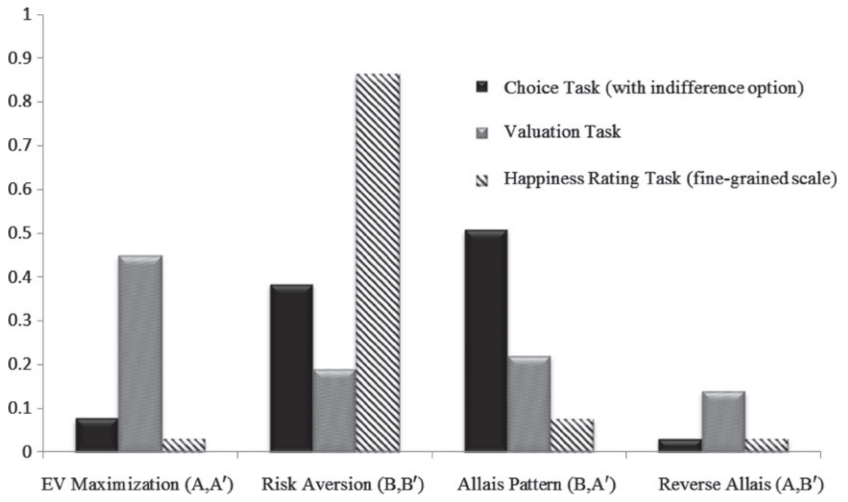


Figure 1. Distribution of strict preference patterns observed in the experiment

79% of all respondents to the fine-grained happiness scale rated B′ higher than A′. While the preference ranking over all four options implied by median happiness ratings is the same as the preference ranking implied by mean happiness ratings, there is a difference in implied preference rankings in the pricing task. In particular, the median subject priced option A higher than B, but the mean price assigned to A is less than \$3000. An exploratory analysis of the data suggests outliers are partially responsible for this difference. Removing the bottom 10% and top 10% of observations from the data in Table 4 reverses this ranking with a mean of \$3018 assigned to A and \$3000 assigned to B. In addition, it is still the case that pricing A higher than B and A′ higher than B′ is the modal response pattern in the pricing task. There is also a clear shift in responses under the pricing task, with 12 of 158 subjects displaying the expected value maximizing preference ranking in the choice task, 2 of 158 subjects revealing this ranking in either of the happiness tasks, and 45 subjects displaying this preference ranking in the pricing task.

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 and B′). The modal pattern (50.7%) in the choice task was the Allais common ratio pattern (B and 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 and A′).

In Figure 1, differences between choices and prices are highly significant both for the modal preference pattern for choices and for the modal preference pattern for prices (both $p < .001$, two-tailed Z difference in proportions test). Comparisons between choices and happiness ratings are also highly significant both for the modal pattern for choices and for the modal pattern for happiness ratings (both $p < .001$, two-tailed Z difference in proportions test). In addition, comparisons between prices and happiness ratings are highly significant both for the modal pattern for prices and for the modal pattern for happiness ratings (both $p < .001$, two-tailed Z difference in proportions test).

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. The DPEF of Schneider and Coulter (2015) predicts that valuation by feeling predominates over happiness rating tasks and that valuation by calculation predominates over pricing tasks. This implies a specific pattern of response mode reversals. Specifically, DPEF predicts risk-averse preferences for the happiness tasks, expected value maximization for the pricing tasks, and the Allais common ratio pattern for the choice task.

Table 5 displays modal response patterns across pairs of response modes. 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. In each case, the modal response pattern captured at least 35% of all preference patterns. Further, 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,[4] and no one response pattern dominated by a large margin. The modal

[4]There are nine response patterns (AA′; BB′; AB′; BA′; A, tie; B, tie; A′, tie; B′, tie; tie, tie) that may be observed for each of the three tasks.

DOI: 10.1002/bdm

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 (Pricing)	0.378	28	74
Happiness vs. Pricing	Problem 2	B′ (Happiness), A′ (Pricing)	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.

response pattern, however, was the one predicted by DPEF, with 11 subjects who 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, as an exploratory analysis, 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 .430, $p < .001$; correlation between prices for A′ and B′ is .452, $p < .001$) as are happiness ratings for each pair of alternatives (correlation between fine happiness scale ratings for A and B is .440, $p < .001$; correlation between fine happiness scale ratings for A′ and B′ is .9408, $p < .001$), 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 .077, p = .515; for B, .087, p = .464; for A′, −0.188, p = .109; for B′, −0.010, p = .935). This 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 responses are provided by the same subjects for the same alternatives in the same survey.

Data from cognitive reflection questions

To examine what may account for differences between subjects, the three common ratio tasks were separated by filler questions including the three-question CRT (Frederick, 2005). These questions are shown in the screen shots in Appendix A along with an additional filler question that was used (the "Count-the-F's" question in Rubinstein, 2013). The average score on the 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 and its relation to the consistency of subject responses.

For purposes of evaluation, the subjects were divided into high and low CRT groups. Subjects with CRT scores of 0 or 1 were assigned to the low CRT group. All others, with scores of 2 or 3, were assigned to the high CRT group. Table 6 displays the consistency of responses both within and across tasks for the two CRT groups. Choices are consistent *within* a task when subjects respond either AA′, or BB′, or a tie for both pairs of alternatives. Responses are consistent *across* tasks, for a given pair of tasks, when 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 preference reversals across response modes. Results of this comparison show that none of the six comparisons in Table 6 are statistically significant, suggesting that the preference reversals are not driven by a subject's degree of cognitive reflection.

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 suggests 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

Table 6. Consistency of responses for high and low cognitive reflection subjects[a]

Proportion consistent within tasks[a]	Choice	Happiness (fine scale)	Pricing
High CRT	0.432	0.921	0.481
Low CRT	0.416	0.806	0.377
Proportion consistent across tasks[b]	Choice vs. Pricing	Choice vs. Happiness	Pricing vs. Happiness
High CRT	0.185	0.395	0.053
Low CRT	0.208	0.333	0.111

The high cognitive reflection test (CRT) group includes all participants who scored a 2 or 3 (N = 81 for Choice, Pricing, and Choice vs. Pricing; N = 38 for fine happiness, choice vs. happiness, and pricing vs. happiness) on the CRT. The low CRT group includes all subjects who scored a 0 or 1 (N = 77 for Choice, Pricing, and Choice vs. Pricing; N = 36 for fine happiness, choice vs. happiness, and pricing vs. happiness) on the CRT.
[a]Responses are consistent *within* a task if they are either AA′, or BB′, or a tie for both pairs.
[b]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.

DOI: 10.1002/bdm

appear to be supportive of this hypothesis. In this section, we consider two other factors that 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 median response patterns in the pricing task.

A second factor that may provide some insight into the underlying process is the response time, both within and across tasks, because feeling-based processes generally operate more quickly than calculation-based processes. Our hypothesis was that response times would be shortest for happiness ratings and longest for the pricing task, consistent with feeling-based processes operating in the former and more calculation-based processes operating in the latter. Response times were recorded to the nearest second in the online survey. Table 7 displays the median and average response times for each task. 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. Each happiness task had a distribution of response times that was significantly faster than the corresponding pricing task ($p < .001$, two-tailed Wilcoxon signed-rank test).

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 (in Choice 1) and 14.84 seconds (in Choice 2) for subjects choosing the expected value maximizing options and were 10.81 seconds (in Choice 1) and 11.33 seconds (in Choice 2) 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. However, 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.

For the choice task, DPEF predicts that response times will be shorter for Problem 1 than for Problem 2, as Problem 2 is predicted to involve more calculation-based processing. This prediction is supported by the data with response times in the choice task being significantly quicker for Problem 1 ($p < .001$, two-tailed Wilcoxon signed-rank test). In addition, DPEF also predicts that choice task response times should not be different from the happiness response times to Problem 1 (if both tasks involve feeling), and should not be different from the pricing response times to Problem 2 (if both tasks involve calculation). We do not find strong support for this prediction. While choice task response times in Problem 1 were not significantly different from the happiness response times to option A, the choice task response times were significantly different from the happiness response times to option B ($p < .001$, two-tailed Wilcoxon signed-rank test). Moreover, the choice task response times in Problem 2 were significantly faster than each of the pricing response times ($p < .001$, two-tailed Wilcoxon signed-rank test). These results are roughly consistent with there being a continuum of processes ranging from more automatic to more effortful, in which response times are faster for happiness than for choice tasks, and are faster for choice tasks than for pricing tasks.

Table 7. Response times across tasks (seconds)[a,b]

Task	A ($4000, .8)	B ($3000, 1)	A′ ($4000, .2)	B′ ($3000, .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				
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 1 second, but it inflates the means and standard deviations.
[b]The choice response times are for the respondents who chose the corresponding option. The happiness rating and pricing response times are computed across all respondents for each option. This table pools the response times for both happiness tasks and the response times for both choice tasks.

DOI: 10.1002/bdm

Discussion

In our test of the Allais common ratio effect in choice, pricing, and happiness rating tasks, we observed large and systematic shifts in subject-level preferences across tasks with modal choices replicating the Allais paradox, modal prices producing consistent expected value maximization, and modal happiness ratings producing consistent risk aversion. This distribution of response patterns (summarized in Figure 1) presents a new paradox for theories of choice under risk. As many normative and descriptive models of decision making are procedure invariant, including expected utility theory, rank-dependent utility theory, 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 theory, similarity theory, and salience theory do not make clear predictions for how behavior may differ across individual evaluation tasks with different response modes. In particular, these models do not distinguish between rating and pricing tasks in terms of their core principles. We note that one could imagine an *ad hoc* salience-based explanation of the response mode differences we observe, although this explanation has no formal connection to the salience theory of Bordalo et al. (2012). In particular, one may posit that pricing tasks make the monetary dimension salient (similar to the scale-compatibility bias proposed by Tversky, Slovic, & Kahneman, 1990), and that happiness tasks make the probability dimension salient. If the salient dimension is overweighted in a given task, this would produce a shift toward expected value maximization in pricing tasks and toward risk aversion in happiness tasks, consistent with the modal responses in our experiment. But this explanation requires another explanation for why happiness tasks make probabilities salient.

One might think that Mukherjee's (2010) DSM 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 be largely explained by the 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. The distribution of response times to the happiness task was significantly faster than the corresponding response times to the pricing task. In addition, the distribution of response times to Problem 1 was significantly faster than that of the response times to Problem 2 in the choice task, consistent with DPEF. 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 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) is susceptible to whether the options are presented in an Allais-type format, or a Savage matrix (Savage, 1954). This behavior is consistent with perceptual-based (i.e., similarity and salience) 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 that there are (at least) three qualitatively different decision-making heuristics that a decision maker may apply to a given task. In particular, a decision maker may choose an option that "looks better" (e.g., if salient comparisons make one alternative more visually appealing than the other), or choose an option that "feels better" (e.g., selecting the option that elicits a more positive or less aversive affective response), or choose the option that is "calculated as better" (e.g., which can be justified by logical reasoning or calculation). If the "looks better" heuristic is predominantly used in the choice task (as it permits visual comparisons between alternatives), if the "feels better" heuristic is predominantly used in the happiness task (as it may involve affect), and if the "calculated as better" heuristic is predominantly used in the pricing task (as it may involve calculation), that would explain both the response mode reversals for the common ratio effect in our experiment and the framing reversals for the common ratio effect observed by Harless (1992) and Harman and Gonzalez (2015). These three heuristics do not require a multiple-systems perspective, although they are consistent with Kahneman's (2003) framework, which distinguishes between three systems: perception, intuition ("System 1"), and reasoning ("System 2"). One could imagine that the perceptual system recommends the alternative that "looks better," System 1 prefers the option that "feels better," and System 2 prefers the option that is justified as better through a logical reasoning process. This is essentially a hybrid explanation in which the similarity and salience judgments operate in choice, but feeling and calculation operate when

DOI: 10.1002/bdm

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*(443), 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*(3), 215–233.

Baucells, M., & Heukamp, F. H. (2010). Common ratio using delay. *Theory and Decision*, *68*(1-2), 149–158.

Bell, D. (1982). Regret in decision making under uncertainty. *Operations research*, *30*(5), 961–981.

Blavatskyy, P. R. (2010). Reverse common ratio effect. *Journal of Risk and Uncertainty*, *40*(3), 219–241.

Bordalo, P., Gennaioli, N., & 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*(4), 25–42.

Friedman, H. H., & Amoo, T. (1999). Rating the rating scales. *Journal of Marketing Management*, *9*(3), 114–123.

Gneezy, U., List, J. A., & Wu, G. (2006). The uncertainty effect: When a risky prospect is valued less than its worst possible outcome. *The Quarterly Journal of Economics*, *121*(4), 1283–1309.

Harless, D. W. (1992). Actions versus prospects: The effect of problem representation on regret. *The American Economic Review*, *82*(3), 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*(4), 369–381.

Hsee, C. K., & Rottenstreich, Y. (2004). Music, pandas, and muggers: On the affective psychology of value. *Journal of Experimental Psychology: General*, *133*(1), 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*, *93*(5), 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.

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*(260), 581–598.

Mukherjee, K. (2010). A dual system model of preferences under risk. *Psychological Review*, *117*(1), 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*(1), 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.

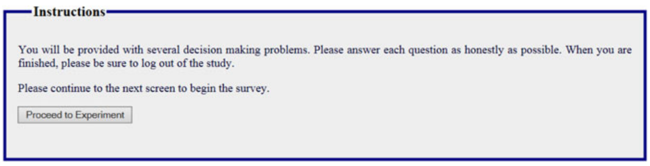
Schneider, M., & Coulter, R. A. (2015). A dual process evaluability framework for decision anomalies. *Journal of Economic Psychology*, *51*, 183–198.

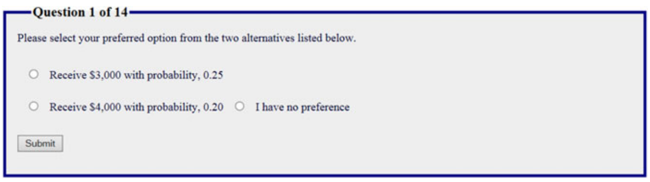
Shampanier, K., Mazar, N., & Ariely, D. (2007). Zero as a special price: The true value of free products. *Marketing Science*, *26*(6), 742–757.

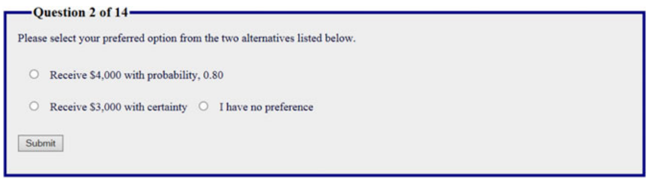
Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323.

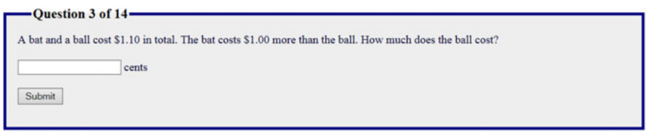
Tversky, A., Slovic, P., & Kahneman, D. (1990). The causes of preference reversal. *The American Economic Review*, *80*(1), 204–217.

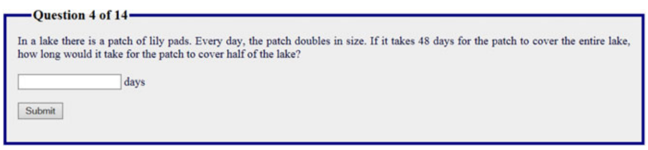
APPENDIX : SCREEN SHOTS FOR EXPERIMENT



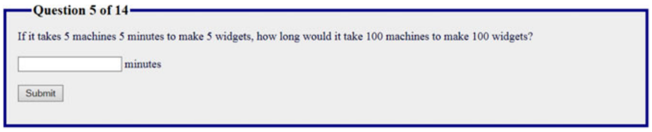


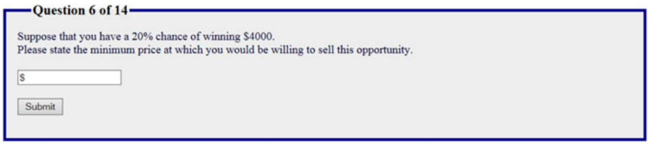


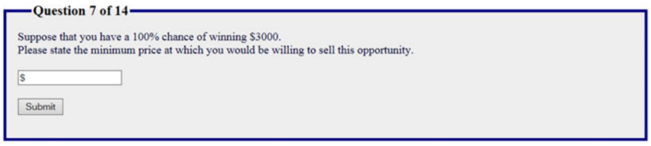




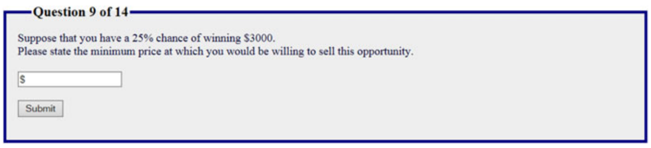
DOI: 10.1002/bdm

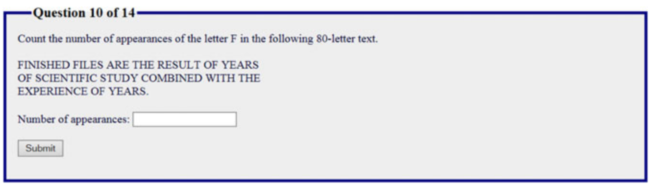

Question 5 of 14
If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?
minutes
Submit

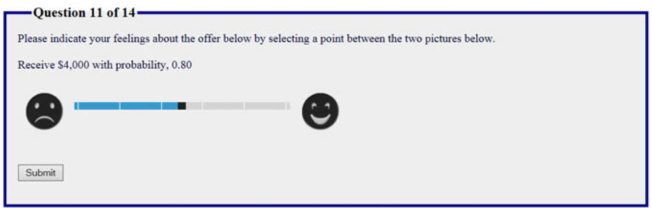

Question 6 of 14
Suppose that you have a 20% chance of winning $4000.
Please state the minimum price at which you would be willing to sell this opportunity.
$
Submit

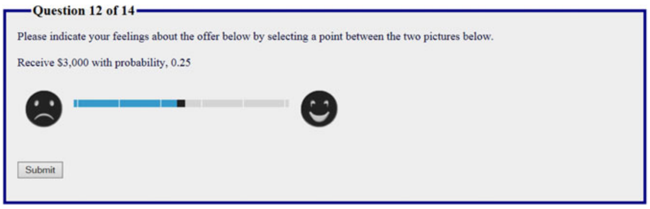

Question 7 of 14
Suppose that you have a 100% chance of winning $3000.
Please state the minimum price at which you would be willing to sell this opportunity.
$
Submit

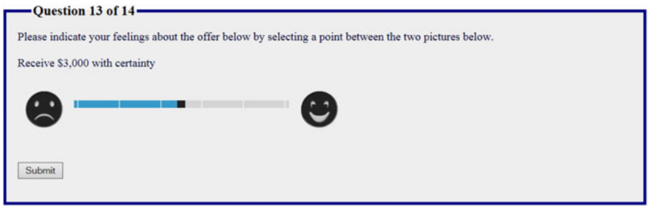

Question 8 of 14
Suppose that you have an 80% chance of winning $4000.
Please state the minimum price at which you would be willing to sell this opportunity.
$
Submit

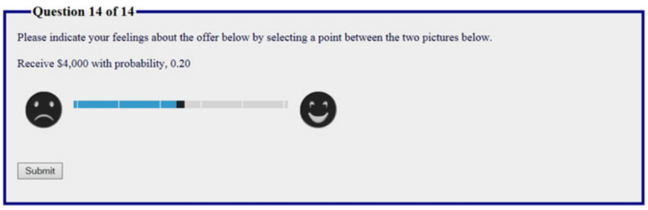

Question 9 of 14
Suppose that you have a 25% chance of winning $3000.
Please state the minimum price at which you would be willing to sell this opportunity.
$
Submit


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

Authors' biographies:

Mark Schneider is a postdoctoral fellow at the Economic Science Institute at Chapman University. His research areas include dual process models of decision making and the role of salience perception in choices involving risk, time, or uncertainty.

Mikhael Shor is an associate professor of economics at the University of Connecticut. His behavioral research areas include choice overload and the use of scientific data in decision making.

Authors' addresses:

Mark Schneider, Economic Science Institute, Chapman University, Orange, CA USA.

Mikhael Shor, Department of Economics, University of Connecticut, Storrs, CT USA.