Individual and Group Decision Making Under Risk: An Experimental Study of Bayesian Updating and Violations of First-order Stochastic Dominance

Gary Charness, University of California, Santa Barbara

Edi Karni, Johns Hopkins University

Dan Levin*, Ohio State University

Proof and reprints to be sent to:

Edi Karni, Department of Economics, Johns Hopkins University, Baltimore, MD 21218, E-mail address: karni@jhu.edu

*We thank two anonymous referees, Nick Burger, Asen Ivanov, Jim Peck, Tim Salmon, and in particular John Kagel for valuable comments. Charness thanks the Department of Economics at UCSB for funding, Karni thanks the Department of Economics at Johns Hopkins University for funding, and Levin thanks the National Science Foundation for funding.

Abstract: In this paper, we report the results of experiments designed to test whether individuals and groups abide by the axioms of monotonicity, with respect to first-order stochastic dominance and Bayesian updating, when making decisions in the face of risk. The results indicate a significant number of violations of both principles. The violation rate when groups make decisions is substantially lower, and decreasing with group size, than when solitary individuals make decisions, suggesting that social interaction or consultation improves the decision-making process. Greater transparency for the decision task tends to reduce the violation rate, suggesting that these violations are due to errors of judgment rather than reflecting the structure of the preference relations. However, we do find some cases where the less difficult decision leads to a higher rate of decision error, both for individuals and groups.

Keywords: Decision making under risk, group decisions, Bayesian updating, first-order stochastic dominance.

JEL category numbers: D80, C91, C92

Among the tenets of rational decision making under risk, monotonicity with respect to first-order stochastic dominance (that is, preference for a better chance of winning a larger sum of money) is the most compelling.[1] Violations of monotonicity with respect to first-order stochastic dominance may, therefore, be regarded as decision-making errors rather than a genuine expression of preferences.[2] In situations involving decision making under risk, the application of Bayes' rule incorporates the impact of new information on the risks associated with alternative courses of action seems compelling, but it is not always intuitive. Consequently, uninitiated decision makers are more likely to err when faced with situations requiring a re-evaluation of risks in the light of new information.

Most experimental studies of individual decision making in the face of risk examine the patterns of choice displayed by subjects acting in isolation. The detection and classification of systematic errors committed by individuals *acting in isolation* in an unfamiliar laboratory environment might improve the understanding of the psychology of individual decision makers. However, the significance of such errors for understanding economic behavior is far from obvious. Seldom, in real life, does an individual find himself in a situation in which he must make an important decision in isolation. Typically, when facing decisions of important consequences, individuals seek advice from family members, friends, and colleagues and, if necessary, consult experts. In other words, most decisions are made in a "social context." It is natural to suppose that, ceteris paribus, the less familiar and/or more complicated the decision problem, the more inclined the decision maker is to consult others before acting. We hypothesize that seeking advice and/or merely deliberating the implications of choosing alternative courses of action helps improve the understanding of the decision problem and makes errors less likely. Consequently, the study of group decision making should enhance the relevance of the experimental results for the analysis of economic behavior.

This paper presents an experimental study of the hypothesis that the simpler the problem and the less

[1] Preference relations having exepcted utility representation are monotonic with respect to first-order stochastic dominance if and only if the utility function is increasing monotonically in income or in wealth.

[2] Kahneman and Tversky (1972) report such violations. However, their subjects changed their minds once the nature of the alternative involved was made apparent.

likely it is to involve emotions, the more likely a decision maker acting in isolation is to reach the correct decision when one exists. In addition, ceteris paribus, decisions made by groups of decision makers are more likely to be correct than decisions made by individual decision makers in isolation, and the larger the group, the more likely the decision is to be correct.

The paper contains two main findings. First, when making decisions in isolation, a substantial number of individuals choose first-order stochastically dominated alternatives. Second, when social interaction is allowed, the number of such violations tends to decrease with the size of the group.

The experimental design, which builds on Charness and Levin (2005), requires subjects to choose between risky prospects under four distinct treatments of different levels of complexity in terms of the potential presence of affect, Bayesian updating, and compounding lotteries. To discern the effect of social interaction, we replicate some of the individual treatments with groups comprising two and three individuals.

Our study differs from Charness and Levin (2005) in at least three important respects. First, we examine choices involving alternatives that may be ranked according to first-order stochastic dominance. Second, to identify the factors responsible for errors, we consider a series of simplifications of the decision problem that allow us to test the hypothesis that greater transparency tends to reduce the observed number of violations. Third, and perhaps most importantly, we investigate the effect of group decision making on the prevalence of violations of first-order stochastic dominance.

Whether groups (or individuals receiving advice) are better at avoiding making systematic errors is an empirical question. While it might seem intuitive that a broader base of opinions would tend to mitigate the effects of misunderstanding and incompetence, thereby leaving less room for error, the verdict of existing literature is not clear cut. Kerr, MacCoun, and Kramer (1996) survey the psychological literature on the relative susceptibility of individuals and groups to systematic judgmental biases, finding that there is no simple answer to the question of performance. In fact, one determinant appears to be whether a decision problem requires a certain flash of insight that, once provided, makes matters transparent; this is known as "truth wins."[3] In this case, it seems intuitive that a team should perform at least as well as any individual member. However, some judgmental decisions do not have this flavor, so that group dynamics may either improve or impair the decision-making process.

In all of the situations considered here, there is a demonstrably correct course of action. It is, therefore, tempting to consider these decisions to be clear enough so that an individual who understands the problem can readily convince the other group members to choose the correct alternative. However, while applying Bayes' rule or correctly analyzing compound lotteries may seem simple to some, subjects in the laboratory and decision makers the field may have difficulty with these tasks and may well not understand the underlying mechanics and principles. By varying the degree of complexity in our design, we offer evidence of how simple a decision must be to generate relative error rates consistent with a truth-wins paradigm.

Recent experimental studies indicate the tendency of groups to outperform individuals in situations involving decisions and games. Blinder and Morgan (2006) ran two laboratory experiments–one a statistical urn problem, the other a monetary experiment–to test the hypothesis that groups make decisions more slowly than individuals. In both experiments, groups were just as quick as individuals to reach decisions and outperformed individuals on average. Kocher and Sutter (2005) find that groups learn more quickly than individuals in a beauty contest (guessing game). Cooper and Kagel (2005) study the behavior of individuals and two-player teams in signaling games; teams consistently play more strategically than individuals do and in some cases, the improvement exceeds the truth-wins (order-statistic) norm. Our study indicates that larger groups tend to commit fewer errors. This is true for every variation in the decision task, whether simple or complex, that we test.

With one exception, we find that reducing complexity tends to reduce errors. Only the elimination of the need to perform Bayesian updating (while still requiring subjects to contend with a compound lottery) results in a higher error rate for both individuals and pairs. We offer a tentative explanation for this anomaly, based on the *representativeness bias* (see Kahneman and Tversky (1972)). This finding needs to be replicated

[3] See Cooper and Kagel (2006) for a discussion of this issue.

before concluding that this is indeed a phenomenon to contend with and not merely a fluke.

The remainder of the paper is organized as follows. In the next section we describe the hypotheses and experimental design. In Section 2 we summarize the findings. We offer some discussion of our results in Section 3, and the conclusions appear in Section 4.

1 The Hypotheses and the Experiment

1.1 Hypotheses

Our basic premise is that all the tasks presented to the participants in our experiment are choice problems that have right and wrong answers. Specifically, choices violating monotonicity with respect to first-order stochastic dominance are bad, representing decision-making errors. Furthermore, these errors are more likely to occur when the choice is accompanied by affect, when the stochastic order is less transparent and when the choices are made in isolation.

To test these hypotheses, we began by confronting the subjects with a two-stage process. In the first stage, the subjects acquired information, and in the second stage they were asked to choose between two lotteries (risky prospects). If utilized correctly – that is, by the application of Bayes' rule – the information acquired in the first stage permits the subject to order the lotteries that figure in the second stage, by first order stochastic dominance.

The presence of the information-acquisition stage simultaneously introduces several factors that may contribute to decision making errors. First, if the information acquired is considered to be "good news" or "bad news" it might be accompanied by an emotional response (affect) blurring the difference between the subsequent alternatives and biasing the choice in a predictable direction, resulting in violations of monotonicity with respect to first order stochastic dominance. Second, to the uninitiated, the need to update the beliefs using Bayes' rule, is confusing and blurs the stochastic order of the risky prospects encountered in the second stage.

In the experiment described below we start by confronting the subjects with tasks that confound all these issues. We then disentangle the issues by first removing affect and then, successively, increasing the transparency of the task. Finally, to test for the effect of social interaction, we introduce group decisions and observe and compare the choices made by groups consisting of two and three individuals to those made by solitary decision makers.

1.2 Experimental design

We conducted web-based sessions on the UCSB campus, with students recruited by e-mail from the same general student population (but with different students) as in Charness and Levin (2005); approximately 65% of the participants were female. Sessions lasted about 45 on average. Participants met in the lab and were given a handout explaining the experimental set-up; detailed, hands-on instructions were provided on the Web site, and participants were required to correctly answer questions testing comprehension.

In our design, there are two equally likely states of the world, Up and Down, and two lotteries (Left and Right) consisting of two different sides of the screen, from which the individual can draw face-down 'cards' that may be either black or white. There is always a mix of colors on the Left side, while the Right side has cards of only one color. Throughout the experiment,[4] only black cards have value, that is, black cards are assigned positive payoff, while white cards pay nothing. In the state Up, the Left side has four black cards and two white cards, while the Right side has six black cards. In the state Down, the Left urn has two black

[4] Except in the *BCD* treatment, described below, where the paying color is not determined until after the first draw.

cards and four white cards, while the Right urn has six white cards.

	Left Side	Right Side
Up ($p = .5$)	●●●●○○	●●●●●●
Down ($p = .5$)	●●○○○○	○○○○○○

Our design includes four different treatments. We refer to the first, most complex, treatment as $ABCD$, where A signifies the potential presence of *affect*, B the need for *Bayesian updating*, C the presence of *compounding* effect, and D the ranking by first-order stochastic *dominance*. In this treatment, subjects are asked to select one of six cards on either the left or right side of the screen. The card is then revealed and replaced, with the face-down cards on the screen then 'shuffled'. The subject then selects a second card, knowing that the state (Up or Down) remains the same as for the first card, and this card is then revealed; people were paid one unit of experimental currency for each black card drawn.

We are primarily interested in the choice behavior when the first draw is from the Left urn, as previous evidence indicates that the error rate is very low in all cases when the first draw is from the Right urn. To ensure receiving relevant observations for each individual, during the first 20 (out of 60) periods in the session we required the first draw to be from the Left side in the odd-numbered periods and from the Right side in the even-numbered periods.

After drawing a black card from the Left side, a second draw from the Right side stochastically dominates a second draw from the left side. Therefore, the subject should switch to the Right side for the second and final draw of the period. (Updating his prior according to Bayes' rule, the subject's probability of success, that is, winning a unit of the experimental currency, from the second draw are: 5/9 if he draws again from the Left side, and 6/9 if his second draw is from the Right side). By the same logic, after having drawn a white card from the Left side, the decision maker should stay with the Left side for the second draw of the period. (Updating his prior by Bayes' rule, the subject's probability of success in the second draw are 4/9 if he draws again from the Left side, and 3/9 if he draws from the Right side).[5]

[5] In the case of an initial black draw, there is a 2/3 chance that the state is Up. If the state is indeed Up, then by drawing

A subject's choice for the second draw may be biased by his emotional response to the observed outcome. A black card in the first draw from the Left represents success (and an award), and may induce a sensation of optimism reinforcing the Left side choice. In other words, having just won a prize drawing from the Left, the subject may consider this to be his lucky side, and be reluctant to switch. On the other hand, drawing a white card first may instill a sensation of disappointment (failure) with the first Left draw and "push" the subject to switch sides for the second draw. To the extent that such heuristics exist and influence decisions, they may exacerbate the violations of monotonicity with respect to first-order stochastic dominance.

In order to mitigate the potential influence of such sensations, we implemented a treatment, dubbed BCD (the *Affect* being removed), in which the first draw from the Left was not associated with a payoff. Moreover, to counteract the possible presence of a more subtle emotional response that might accompany success or failure when observing the first draw, we did not tell the subject whether it is the black or white card that has a positive payoff in the second draw until after the first card was drawn. This was feasible due to the symmetry of the distribution of black and white cards across the Up and Down states. People were required to make their initial draws from the Left side, observing the color of the card before replacement; there was no payment for the first draw. At the bottom of the screen, the subject was then informed as to the color that would pay on the second draw.

In our third treatment, CD, we eliminate the need for Bayesian updating while preserving the compound-lottery structure involving combining continuation probabilities after different initial outcomes. In this treatment prizes were awarded for drawing black cards during 80 periods. As before, after a first black draw from the Left the correct Bayesian posterior probability of Up is 2/3 and that of Down is 1/3. In order to eliminate the need for Bayesian updating, we simplified the procedure so that there is only one draw in a period. We offer subjects a choice between the Left and Right side, informing them that the probability of

from the Left side, the subject has a 2/3 chance of drawing another black card. If the state is Down (probability 1/3), then the subject has a 1/3 chance of drawing a black card. Thus, the expected value is 2/3*2/3 + 1/3*1/3 = 5/9. However, by choosing from the Right side for the second draw, the subject has a 2/3 (6/9) chance of drawing a black ball. The calculations are similar for the case of an initial draw of a white card.

state Up is 2/3 (the same as in the BCD treatment after a would-be successful first draw); this matches the case when a would-be-successful color is drawn first in the BCD treatment. Note that to compute overall probabilities a subject must analyze, for the Left side, a compound lottery involving different states of Nature.

The fourth treatment, D, focuses on dominance. In this treatment, done on paper, subjects were offered a reduced lottery instead of the compound lottery mentioned in the preceding paragraph. Subjects were presented with a choice of choosing the Left or Right side after being informed that we would roll a nine-sided die (actually ten-sided, but if 0 came up we rolled again); if a subject chose Left, a prize of $2 would be awarded if numbers 1-5 came up, while if the decision maker chose Right, this prize would be awarded if numbers 1-6 came up. We handed out a sheet of paper with this choice at the end of BCD or CD sessions, rather than repeating the choice multiple times. This procedure is equivalent to the case (in the more complex treatments) where the first draw has been made from the Left side and was successful, as there is the same 5/9 probability of success by drawing again from the Left side and a 6/9 chance of success by instead drawing from the Right side.

The procedure that we chose to follow in these experiments, that is, starting with the $ABCD$ treatment rather than the simpler ones, reflects our view that decision making under risk and under uncertainty, typically involves all, or some, of these factors. One makes a choice and experiences an emotional response to the outcome; this outcome also generates information that necessitates updating the priors, and the calculations may also involve compound lotteries.

1.3 Individual and group decisions

The second part of our study examines the social-interaction effects on decisions. Our interest in this issue stems from the obvious fact that most real-life decisions take place in environments in which individuals can, and do, consult and seek advice before making decision, thereby benefiting from the experience and expertise of others. We hypothesize that the opportunity to exchange ideas and opinions improves the decision making ability of individuals. To test this hypothesis, we administered some of the treatments with groups instead of individuals acting as decision making units. Specifically, we repeated treatments $BCD/CD/D$ and CD/D with pairs of subjects making joint decisions, and treatments $BCD/CD/D$ with groups of three subjects making decisions.[6] The subjects in each group were permitted to speak to each other (quietly) and could reach decisions in any manner; while experimenter intervention (e.g., a coin flip) was available if needed. In practice, no group ever failed to reach a decision on its own.

To grasp our hypotheses, consider the case of the BCD treatment. In a population of subjects an unknown proportion, say x, of the subjects are Bayesians (that is, subjects who know and are able to apply Bayes' rule). We hypothesize that, without fail, Bayesians choose the dominating option. The rest of the subjects, whose proportion in the population is $(1-x)$, employ other means to assess the alternative risks and make their choices. Among these subjects, a certain proportion, say $(1-\alpha_1)$, may actually "guess" the right choice and pick the dominating alternative, without using Bayesian formula. Thus, $(1-\alpha_1)(1-x)$ is the proportion of non-Bayesian subjects in the population that, by chance or by intuition, made the right choice. Clearly, $\alpha_1(1-x)$ is the proportion of non-Bayesian subjects that made the wrong selection.

We are reluctant to speculate on what might be reasonable values for x or α_1. In particular, if the non-Bayesian participants are as likely to make the right decision as to make the wrong decision, then $\alpha_1 = 0.5$. However, it is conceivable that, even if they do not know how to apply Bayesian rule, subjects may nevertheless intuitively choose correctly more often than not. Thus, we hypothesize that $\alpha_1 < 0.5$.

We propose here an hypothesis that would allow us to estimate the values of x and α_1, and test them. To illustrate how this is done, suppose that Bayesian decision makers do not err. Then observing the proportion of wrong choices, say k_1, in the individual experiments, we learned that:

[6] We have not investigated group effects for the ABCD treatment, as we are more interested in simpler environments and trying to identify where the truth-wins paradigm breaks down.

$$\alpha_1 (1 - x) = k_1 \tag{1}$$

Turning next to decisions by groups, we hypothesize that whenever matched with non-Bayesians, the Bayesians are able to convince them of the right way of looking at the (posterior) alternatives and, consequently, choose the right alternative - this would be the case if the truth-wins norm applies. We also hypothesize that, even if Bayesians are not represented in a group, the mere deliberation of the alternatives among members of a group, tends to increase the number of right choices. Formally, let k_2 be the proportion of wrong choices by groups consisting of pairs of individuals. Then,

$$\alpha_2 (1 - x)^2 = k_2, \tag{2}$$

where it might seem a natural presumption that $\alpha_2 \leq \alpha_1$; we test this presumption later. That is, k_2 is the proportion of non-Bayesians that were randomly matched with other non-Bayesians and who jointly chose the wrong answer.

If we alternatively insist that $\alpha_2 = \alpha_1 = \alpha$, these two equations allow us to estimate (solve) simultaneously for x and α. Call these estimates x^* and α^*. But if in fact, $\alpha_2 < \alpha_1 \leq 0.5$, then x^* overestimates the true value of x, and α^* overestimates α_1 and α_2, where the overestimation of α_2 is relatively larger than that of α_1.[7]

Observe next that in groups of three subjects, the proportion of wrong choices should satisfy $k_3 =$

[7] To see this, note that $\alpha_i = \alpha, i = 1, 2$ implies that

$$\frac{k_2}{(1-x)^2} = \frac{k_1}{1 - x}$$

thus $x^* = 1 - \frac{k_2}{k_1}$. If $\alpha_1 > \alpha_2$ then, by the same reasoning, $\hat{x} < 1 - \frac{k_2}{k_1}$, where $\hat{x}$ is the true value of x. Thus, $\alpha_i^* = k_i / (1 - x^*)^i > k_i / (1 - \hat{x})^i = \alpha_i, i = 1, 2$. Hence,

$$\frac{\alpha_2^*}{\alpha_1^*} = \frac{k_2}{k_1 (1 - x^*)} > \frac{k_2}{k_1 (1 - \hat{x})} = \frac{\alpha_2}{\alpha_1}.$$

$\alpha_3 (1-x)^3$, where x is the true proportion of Bayesians in the population and $\alpha_3 \leq \alpha^* \leq \alpha_2$. Thus, our hypothesis does not indicate unambiguously the relationship between the magnitude of k_3 and $\alpha^* (1-x^*)^3$. To see why, note that

$$\alpha_2 (1-x)^2 = k_2 = \alpha^* (1-x^*)^2 . \tag{3}$$

Thus, solving α^* from equation (3), gives

$$\alpha^* (1-x^*)^3 = \frac{\alpha_2 (1-x)^2}{(1-x^*)^2} (1-x^*)^3 = \alpha_2 (1-x)^2 (1-x^*) \gtreqless \alpha_3 (1-x)^3 = k_3, \tag{4}$$

where we used the inequality $\alpha_3 \leq \alpha_2$ and that x^* overestimates x. However, $k_3 \leq \alpha^* (1-x^*)^3$ implies that $\alpha_3 < \alpha_2$. Hence, the last inequality, that non-Bayesian trios make better guesses than non-Bayesian pairs, is a testable hypothesis.

In the CD and D treatments, the need for Bayesian updating is removed. This is, it can be modeled along the same lines as above assuming that $x = 0$. Note, however, that because the treatments are less complex, it is not necessarily true that the proportions of subjects that make the wrong selection remain the same across treatments. It seems reasonable to suppose that this proportion declines when the stochastic order becomes more transparent. Formally, let α_i^{CD} and $\alpha_i^D, i = 1,2,3$ denote the proportion of subjects that make the wrong selection in the CD and D treatments, respectively, rather than in the BCD treatment. We hypothesize that $\alpha_i > \alpha_i^{CD} > \alpha_i^D$, $i = 1,2,3$.

We summarize our treatments with individual decision makers in Table 1, and those with groups in Table 2:

Table 1: Treatment Summary - Individuals

Treatment	1st draw restriction	Payment	Participants	Periods
$ABCD$[8]	In first 20 periods	Black cards from both draws	57	60
BCD	Left draw only	2nd draw only, color TBD	57*	80
CD	None	Black cards	48	80
D	None	Black card	104	1

*One person played solo for the 30 periods of the 3-person BCD treatment.

Table 2: Treatment Summary - Groups

Treatment	1st draw restriction	Payment	Groups	Periods
$BCD-$ 2 person	Left draw only	2nd draw only, color TBD	30	30
$BCD-$ 3 person	Left draw only	2nd draw only, color TBD	21	30
$CD-$ 2 person	None	Black cards	36	30
$D-$ 2 person	None	Black card	66	1
$D-$ 3 person	None	Black card	21	1

Participants were paid \$0.30 for each successful draw in the $ABCD$, BCD, and CD treatments. In order to pay reasonably similar amounts across treatments, we had 80 periods in the individual BCD and CD treatments rather than the 60 periods in the $ABCD$ treatment (80 decisions counted, compared to 120) and also included the D treatment at the end of the session. No person could participate in more than one session or treatment (with the exception of the BCD/D and CD/D sessions). Group decisions took longer, so we reduced the number of periods to 30 and increased the show-up fee (to \$8 from \$5) in these sessions.[9]

[8] As mentioned earlier, we only consider $ABCD$ decisions made after a first draw from the Left urn.

[9] We chose to maintain the same marginal benefit for choosing the right color, rather than keeping the same show-up fee and changing the marginal incentives. While it is possible (although unlikely) that the slightly larger show-up fee caused participants to take the task less seriously, this would only serve to lessen the difference between the error rates in individual and group

On average, our participants earned approximately $16-17 for a one-hour session.

2 Results and Analysis

2.1 Individual decisions

We present our findings regarding individual decision making according to the treatments. The aggregate result are summarized in Table 3 below.

Table 3 - The violation rates of First-order stochastic dominance, Individual Decisions

Treatment	Following Success	Following Failure	Aggregate Error %
ABCD	37.5% (225/600)	40.0% (234/585)	38.7% (459/1185)
BCD	18.8% (420/2234)	39.9% (908/2276)	29.4% (1328/4510)
CD	30.2% (1158/3840)	–	–
D	8.7% (9/104)	–	–

Individual treatment ABCD: For this treatment we document two types of first-order stochastic dominance violations: 1) Violations that follow success (namely, a draw of a black card) and receiving a reward in the first round and 2) violations that follow failure (namely, a draw of a white card) and not receiving a reward in the first round. Recall that monotonicity with respect to first order stochastic dominance implies that success should have resulted in the subject switching to the Right lottery for the second draw. Violations in this instance mean that the subject didn't switch. Similarly, after a first draw of a white card (representing a failure), subjects should have stayed with the Left lottery for the second draw.

decisions.

Violations in this case mean that the subjects did switch. The aggregate violation rate of the first type, that is, following success is 37.5% and the aggregate violation rate of the second type, that is, following failure is 40.0%; there is no significant difference between these rates. The combined aggregate error rate is 38.7%.[10]

Recall that the implementation of the *ABCD* treatment involved the forced and voluntary choice of the side of the first draw. Specifically, the first 20 periods are forced first draws (alternating Left and Right) and the last 40 periods are voluntary first draws. The results of the forced and voluntary treatments displayed separately, are as follows: The error rate with the forced-first-draw is 34.5%, (101/293), following success and 40.4% (111/277) following failure; the error rate with the voluntary-first-draw is 40.4% (124/307) following success and 39.9% (123/308) following failure.

Perhaps the most important, and striking, aspect of these findings is the high rate of violations of monotonicity with respect to first-order stochastic dominance. A second aspect of these findings is the remarkable similarity of the error rates following success and failure. This suggests that success induces a tendency to erroneously "stay the course" while failure induces a tendency to, again erroneously, change course. These tendencies may reflect a bias due to affect. To test for the effect of affect, we remove it in the next treatment.

Individual treatment BCD: Charness and Levin (2005) observed a large reduction in aggregate error rates, from 47.0% to 28.2%, when the affect is eliminated; the reduction was particularly dramatic in the case of a successful (same color as the one paying on the second draw) first draw. There the reduction is from 36.8% to 13.5%. In both of these cases the reductions were strongly statistically significant.

In our experiment, the overall reduction in error rates after removing the affect is smaller, yet still large (from 38.7% to 29.4%). The difference between these two rates is marginally significant in a conservative

[10] This is considerably lower than the aggregate error rate in the *ABCD* treatment of Charness and Levin (2005). In that experiment, however, the amount paid for a successful draw differed for the two urns rendering the decision somewhat more complex. Due to this differential payment, not all Bayesian errors in Charness and Levin (2005) reflected violations of first order stochastic dominance.

ranksum test using each individual's overall rate as one observation ($Z = 1.34, p = 0.090$, one-tailed test). Remarkably, this reduction is due entirely to the very large reduction following success (from 37.5% to 18.8%).[11] There is no noticeable reduction following failure. These distinct tendencies are puzzling and seem to merit further study.[12] Moreover, this rather large drop in the error rate following success is responsible for some other anomalies in the findings, to be discussed below. Finally, the binomial test shows that the error rate following "failure" is significantly higher than the error rate following "success" ($Z = 4.81, p = 0.000$), since this was higher for 44 individuals and lower for 9 individuals (and the same for the other four people in this treatment).

Individual treatment CD: In going from the BCD to the CD treatment we eliminated the need for Bayesian updating. That is, subjects are told that the probability of Up is 2/3. It's "as if" they made a successful first draw and now face the second decision. However, in this treatment there is no actual first draw, and the one draw that remains is equivalent to the second draw in treatments $ABCD$ and BCD.

It is well documented in the psychological and experimental economics literature that subjects are not good Bayesians.[13] This suggests that eliminating the need for Bayesian updating would have reduce the error rate. In fact, however, the error rate *increased* significantly from that in the BCD treatment (from 18.8% to 30.2%; $Z = -2.40, p = 0.017$, in a two-tailed ranksum test).

[11] The reduction is statistically significant, with $Z = -2.92$, and $p = 0.004$, two-tailed test.

[12] While we cannot be certain why the effect of affect was smaller here than in Charness and Levin (2005), we note that the corresponding ABCD and BCD treatments in that experiment were slightly different, in that people received 7/6 unit for successful draws from the Right side (compared to 1 unit for a successful draw from the Left side). This added a layer of complexity that appears to have affected behavior (note the reduction in the overall ABCD error rate from 48.3% to 38.7%). Perhaps affect is less determinative with simpler problems. In any case, the reduction after success is similar in both cases; the puzzle is the lack of reduction in the error rate after failure, compared to the modest reduction found in Charness and Levin (2005).

[13] See, for example Tversky and Kahneman (1971, 1973), Kahneman and Tversky (1972), and Grether (1980, 1992). More recent studies (e.g., Ouwersloot, Nijkamp & Rietveld, 1998; Zizzo, Stolarz-Fantino, Wen & Fantino, 2000) also provide strong evidence that experimental subjects are often not even close to being 'perfect Bayesians'.

Individual treatment D: The last treatment simplifies the task further by eliminating the complication due to compounding. This enables us to observe, directly, how decision makers chose between two lotteries with probabilities of 5/9 and 6/9, respectively, of winning the same prize. Only nine individuals out of 104 (8.7%) chose Left.[14] This number is small enough to suggest that the violations are due to errors and misunderstandings, and don't actually reflect the participants' preferences. Checking each individual who made choices in D (as well as decisions in BCD or CD), the D error rate is lower than the BCD error rate for 35 subjects and is higher for nine subjects; the D error rate is lower than the CD error rate for 35 subjects and is never higher for anyone. The binomial test finds $Z = 3.18$ ($p = 0.001$) and $Z = 5.92$ ($p = 0.000$) for the respective differences, both significant.

2.2 Group decisions

We turn next to the evidence concerning group decisions. As mentioned earlier, the members of each randomly-formed group consulted with each other before making a group choice. Each group was left to its own devices as to the decision process[15]

The main findings are summarized in Table 4, below. Our findings indicate a general and significant drift towards a reduction of the error rates as the group size increase

[14] One person admitted he hadn't read the instructions. A colleague also suggests that some subjects may simply have been "jerking our chain" by deliberately choosing the dominated strategy.

[15] We told subjects that if they were deadlocked, we would flip a coin to determine their choice. However, this was never required.

Table 4- The violation rates of First-order stochastic dominance, Group Decisions

Treatment	Following Success	Following Failure	Aggregate Error %
$BCD-1$	18.8% (420/2234)	39.9% (908/2276)	29.4% (1328/4510)
$BCD-2$	15.4% (67/434)	32.6% (152/466)	24.3% (219/900)
$BCD-3$	7.5% (23/307)	10.8% (35/323)	9.2% (58/630)
$CD-1$	30.2% (1158/3840)	–	–
$CD-2$	23.0% (248/1080)	–	–
$D-1$	8.7% (9/104)	–	–
$D-2$	3.0% (2/66)	–	–
$D-3$	0.0% (0/21)	–	–

Group treatment BCD: With groups of two subjects, the error rate under the BCD treatment is lower than with individual decision-makers after both success and failure, and the aggregate error rate is reduced by over five percentage points, to 24.3%. However, this difference is not statistically significant.

The reduction is considerably more dramatic with groups of three subjects. In this case, the error rates following both success and failure, are less than half the corresponding rates with two-subject groups, and the aggregate error rate is only 9.2%, less than that for either individual or paired decision-makers; ranksum tests give $Z = 3.57$ ($p = 0.000$) and $Z = 2.96$ ($p = 0.003$) for the respective differences, both statistically significant. While the improvement is considerable for decisions following failure and following success, the effect is strongest following an unsuccessful first draw.

We note that the error rate after failure in the first draw is higher than that after a success in each of the four cases in Tables 3 and 4. A simple binomial test indicates that this will happen by chance with $p = 0.062$, on the margin of statistical significance.

Group treatment CD: This treatment indicates that groups are less prone to violations of first-order stochastic dominance than are individuals; in this case, the error rate for two-subject groups is 23.0%. As with all comparisons across individual and group decision-making entities, this rate is lower than the corresponding 30.2% rate for individuals. The difference between these two rates is marginally significant in a ranksum test using each individual's overall rate as one observation ($Z = 1.37, p = 0.085$, one-tailed test). We have no evidence concerning the possible effect of increasing the group size from two to three subjects, although the evidence from the BCD treatment suggests that hypothesis that such an increase will reduce the error rate significantly.

As with individual decision-makers, the error rate for pairs is lower in the BCD treatment than in the CD treatment (15.4% versus 23.0%). However, the difference between these two rates is at most marginally significant in a ranksum test using each pair's overall rate as one observation ($Z = 1.42, p = 0.156$, two-tailed test[16]); of course, this is a rather conservative test.

Group treatment D: In this treatment, group-decision making nearly eliminated the errors found in the individual case. With groups consisting of two subjects, the error rate was down to 3.0%; more telling, perhaps, is the 0% error rate observed for groups consisting of three subjects. Checking each pair who made choices in D as well as decisions in BCD or CD, the D error rate is lower than the BCD error rate for 12 pairs and is higher for one pair, while the D error rate is lower than the CD error rate for 24 pairs and is higher for one pair. The binomial test finds $Z = 3.05$ ($p = 0.001$) and $Z = 4.60$ ($p = 0.000$) for the respective differences from randomness. Checking each trio who made choices in D as well as decisions in BCD, the D error rate is significantly lower (at $p = 0.001$) than the BCD error rate for 13 groups and was never higher. The binomial test finds this one-sidedness is significant ($Z = 3.61$, $p = 0.000$).

These very low error rates for the groups tend to confirm our claim that the observed errors are due to mistakes rather than genuine expressions of the participants' preferences.

We perform a simple regression to illustrate in a more comprehensive manner the significance of the

[16] We cannot use a one-tailed test, as we have no *ex ante* prediction that error rates will increase going from BCD to CD.

difference in error rates across treatments (to keep the comparison fair, we only include successful draws in BCD), as well as across group size. We use each decision-making entity's overall error rate in the BCD or CD treatment (whichever applies) as one observation and count the D choice (if any) as another observation, with standard errors clustered by individual. We regress this error rate against various dummies:

Table 5- Regression of individual error rates across conditions

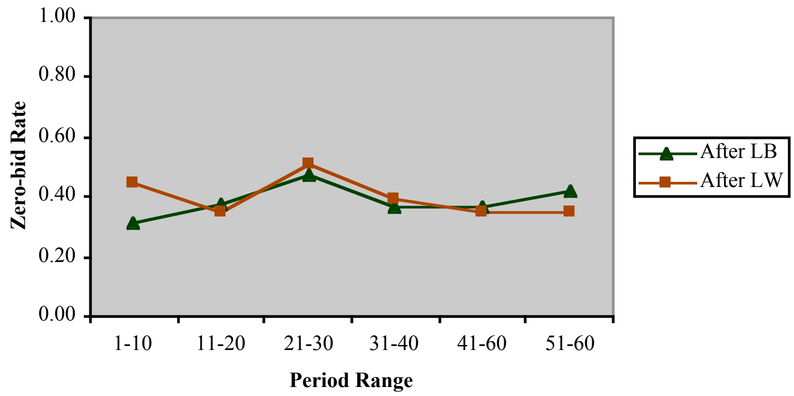
	Coefficient	Std. Error	Z-statistic	$P > \|Z\|$
CD	.1228	.0343	3.58	0.000
D	−.0984	.0267	−3.68	0.000
Pair	−.0637	.0277	−2.30	0.021
Trio	−.0975	.0428	−2.28	0.023
Constant	.1832	.0254	7.20	0.000

Number of obs. = 384; Adjusted $R^2 = 0.1416$

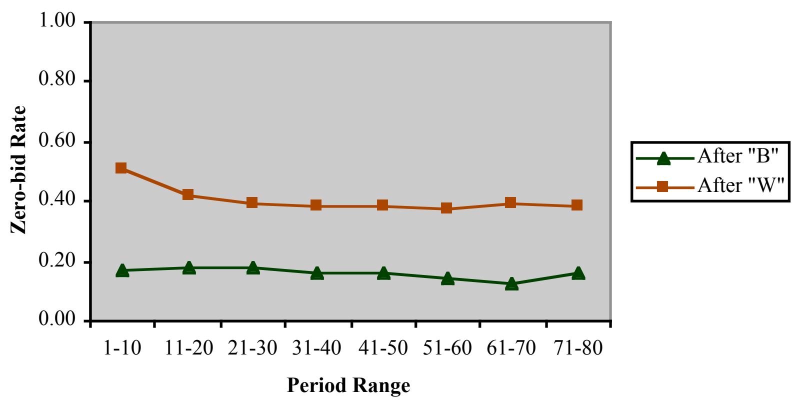
We see that in relation to the BCD baseline, the error rate in the CD treatment is significantly higher and the error rate in the D treatment is significantly lower. The error rate with larger groups (compared to the individual baseline) is significantly lower, with the negative coefficient for the trio dummy higher than the coefficient for the pair dummy.

Since we have differing number of periods for group and individual decisions, this could potentially cause a problem in our comparison However, choices are rather stable over time, as can be seen in Figures 1B-6B below. Furthermore, if the error rate is decreasing over time, having more periods in the group treatments would only serve to increase the difference between the error rates for individual and group decisions.

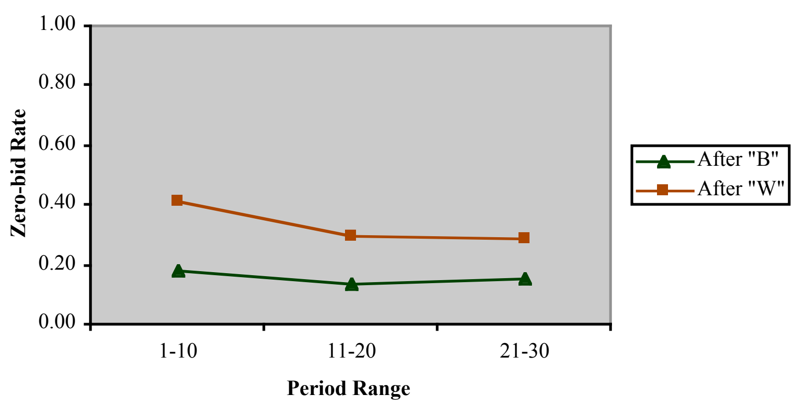
Error Rate over Time in ABCD



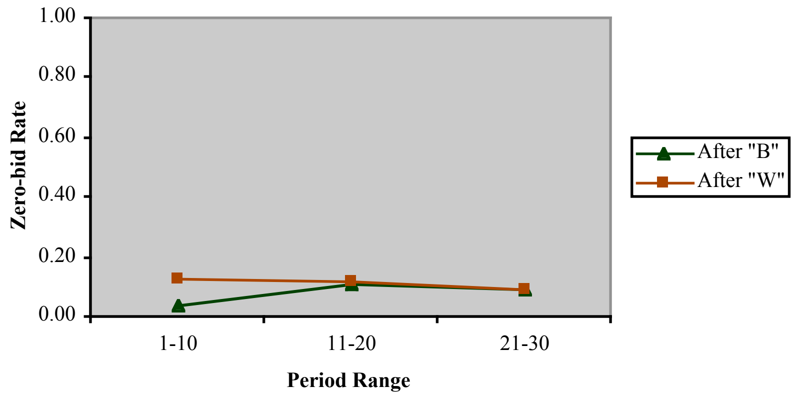
Error Rate over Time in BCD-1



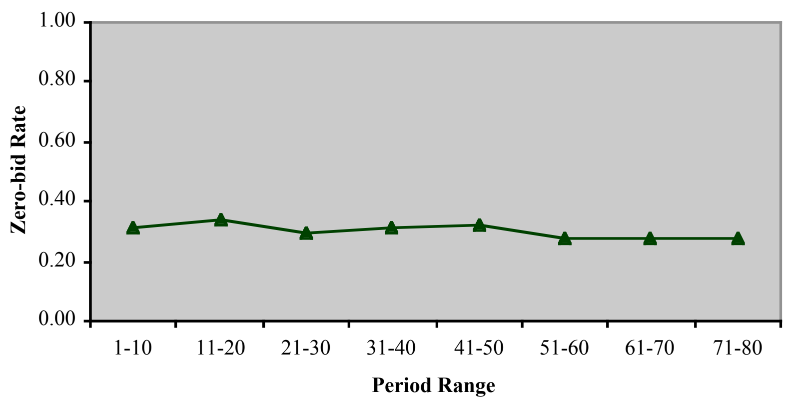
Error Rate over Time in BCD-2



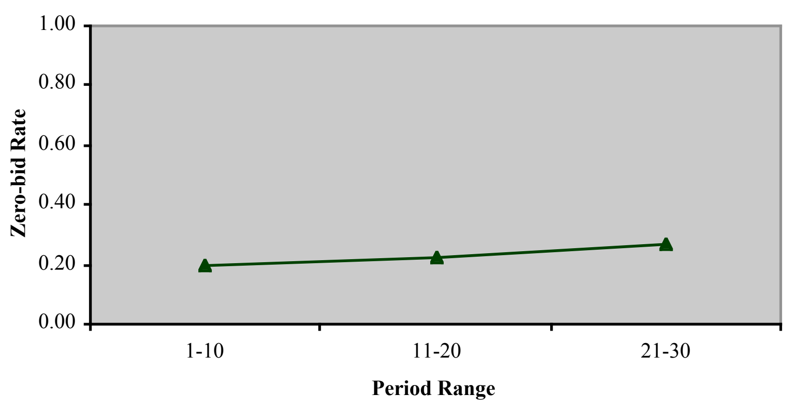
Error Rate over Time in BCD-3



Error Rate over Time in CD-1



Error Rate over Time in CD-2



3 Discussion

We turn next to an interpretation of these findings based on the model of section 2.3. Consider the treatment BCD. The aggregate figures in our data, our findings indicate that $k_1 = 0.294$ and $k_2 = 0.243$. If we presume that $\alpha_2 = \alpha_1 = \alpha$, calculations show that $x^* = 0.173$ and $\alpha^* = 0.356$.[17] Note, however, that x^* overestimates the value of x, the proportion of Bayesian decision makers in the subject population, and α^* overestimates α_1 and α_2, where the overestimate of α_2 is relatively larger than that of α_1. This lends support to the hypothesis that $\alpha_1 < 0.5$. Moreover, we argued that $\alpha^* (1 - x^*)^3 \geq k_3$ implies that $\alpha_3 < \alpha_2$. But,

$$k_3 = 0.092 < 0.201 = 0.356\,(1 - 0.173)^3\,.$$

Hence, our hypothesis that $\alpha_3 < \alpha_2$ is also supported by our findings. There are a number of different statistical tests that could be used to test for the degree of significance. First, we note that this is equivalent to testing how certain we are that our observed value of 0.092 for k_3 (in Table 4) indicates a true value smaller than 0.201. One approach considers the error rate for each trio making decisions. Using a standard t-test on the 21 trios, we can reject a true mean of 0.201 ($p = 0.001$, one-tailed test). However, the distributional assumption of this test may be questionable; in fact, while the mean error rate was 0.092, the median error rate for trios in BCD was only 0.033 (one error out of 30 chances), with seven groups with error rates above 0.092 and 14 groups with error rates below it. Thus, we instead use bootstrapping to construct a 95-percent confidence interval around the observed mean (0.041 to 0.149). Clearly our calculated value for k_3, 0.201 falls well outside this range.

Consider next the CD and D treatments. Observe that $\alpha_1^{CD} = 0.302 > \alpha_2^{CD} = 0.230$ and $\alpha_1^{D} = 0.087 > \alpha_2^{D} = 0.030 > \alpha_3^{D} = 0$. Moreover, $\alpha_i^{CD} > \alpha_i^{D}, i = 1, 2$. All of these inequalities are consistent with our hypothesis. Since these respective differences consist of five directional tests all going in the predicted direction, the likelihood that this would occur by chance is $p = 0.031$.

Finally, calculating α_1 and α_2 using the error rate following success in the BCD treatment, we obtain

[17] This is the solution to the equations: $\alpha\,(1 - x) = 0.294$ and $\alpha\,(1 - x)^2 = 0.243$.

$\alpha_1 = 0.228$ and $\alpha_2 = 0.226$.[18] Since x^* is an overestimate, the estimates of α_1 and α_2 also overestimates the true values. Thus, $\alpha_i < \alpha_i^{CD}$, $i = 1, 2$. This is not what we expected and it goes against the hypothesis stated earlier. It is related to the puzzle, pointed out before, that the error rates following success under the BCD treatment exceed those under the CD treatment.

The increased the aggregate error rates, for both individuals and pairs, associated with the simplification from BCD to CD warrants further examination. If this is turns out to be a real phenomenon rather than a fluke, one possible explanation might be the influence of the *representativeness bias*, which occurs when subject pay too much attention to the sample (recent) information, wrongly believing that it represents the true distribution and so effectively neglecting the (given) base rate. For example, after a success (failure) the probability of Up may seem really close to 1 (0). If this bias is sufficiently strong and dominates the effect of the simplification of the task that occurs when moving from BCD to CD, eliminating it could explain the higher error rate in the CD compared to that of the BCD treatment, as giving the correct posterior may help to silence the bias.[19] This argument also assumes that the representativeness bias dominates a "conservatism" bias (see Edwards (1982)) that occurs when the subject ignores, or "under-updates," the prior distribution given new evidence, instead holding on to the base rate. There is considerable evidence that subjects underweight base rates in this way,[20] and our results after success are consistent with this notion. We noted earlier that the representativeness bias should also reduce the error rate after an initial failure (since the probability of Up seems like 0 there is no point in trying the Right urn); in fact, the only

[18] The restriction to error rates following success is for the purpose of comparison with the other treatments. The estimated values are based on the solution of the equations $\alpha_1 (1 - 0.174) = 0.188$ and $\alpha_2 (1 - 0.174)^2 = 0.154$, where $x^* = 0.174$ and the error rates, 0.188 and 0.154 are from Tables 3 and 4, respectively.

[19] Using the methodology of modern experimental economics, Grether (1980) finds evidence of the representativeness bias in a task involving identifying the urn from which a ball was drawn. El-Gamal and Grether (1995) show that people are more prone to the representativeness bias than the conservatism bias. Gangulay, Kagel and Moser (2000) discuss a similar effect. This may help explain why we don't see a big drop going from treatment BCD to treatment CD, but the observed *increase* is still surprising.

[20] See for example Kahneman and Tversky (1972, 1973) and Ganguly, Kagel, and Moser (2000). Interestingly, the latter study finds that a more abstract context (such as ours) tends to *reduce* this bias.

comparison we have across treatments, $ABCD$ versusBCD in Table 3, shows dramatic improvement from removing the affect after an initial success, but no improvement after an initial failure. Perhaps people react asymmetrically to good news and bad news.[21]

Note that the beneficial effect of increasing group size is present, with improvement from 39.9% to 32.6% to 10.8% for individuals, pairs, and trios, respectively (Tables 3 and 4). In fact, we observe, in every case, a monotonic relationship between group size and the violation rate. Yet our results are not completely consistent with a truth-wins norm, as it appears that α does vary across conditions of complexity and group size. In principle, this norm predicts that we should see more improvement in the error rates when going from individuals to pairs than when going from pairs to trios; while this is true in the D treatment, it is not at all true in the BCD treatment. Thus, perhaps the less complex tasks are more like flashes of insight and are less deliberative, while deliberation plays a greater role in more complex environments.

4 Conclusions

The two most striking findings of our study are: First, there is a substantial number of individuals who, when making decisions in isolation, violate the most basic tenet of rational decision making by choosing first-order stochastically dominated alternatives. Second, the incidence of choice of dominated alternatives is negatively correlated with the size of the group; in fact, every comparison across group size favors the larger group. Thus, social interaction, whether through the presence of experts or consultation, improves the decision-making process in the laboratory.

Typically, important individual decisions are made in social environments, enabling the decision makers to discuss their options and benefit from the experience and expertise of others. Therefore, the two conclusions above lend support to the claim that the experimental evidence indicating *systematic deviations from the courses of action prescribed by normative models of decision making under risk and uncertainty, such as*

[21] Charness and Levin (2005) also find this asymmetric improvement from removing affect.

expected utility theory, are due, in part, to the artificial isolation imposed by the experimental setting. These violations tend to be less pronounced when social interaction is allowed. Our findings also suggest that, in addition to the presence of experts (Bayesian decision makers, in our particular context) the mere opportunity to discuss the alternatives and to clarify their meaning is important. To the extent that this also prevails in the field, perhaps a group decision-making environment may be more representative of non-laboratory conditions than one where individuals operate on their own.

Third, the presence of affect, the need to apply Bayesian updating and the complexity of the alternatives tend to contribute to erroneous judgments and poor decisions. This is in line with earlier results obtain by Charness and Levin (2005). Other evidence consistent with this finding is reported in Kagel and Levin (1986) and Charness and Levin (2006).

The results of this paper also raise some issues that merit further investigation. First, the systematic higher rates of violations of first-order stochastic dominance following a failure compared to those following success is puzzling and further examination of this tendency is warranted. Second, the significant increase in the violation rate following a successful draw, when we have dispensed with the need to perform Bayesian updating, is troubling; further examination of this issue seems worthwhile. Third, we examine joint decision making by requiring that the group jointly reach a decision. An alternative test of the social-interaction effect should allow subjects to discuss their decisions and than make their choice individually. Finally, our study focused on violations of monotonicity with respect to first-order stochastic dominance. It would be interesting to see if a similar tendency for reductions in violations of the independence axiom or the sure thing principle obtain when social interaction is allowed.

References

Blinder, Alan and John Morgan. (2006). "Are Two Heads Better than One? An Experimental Analysis of Group versus Individual Decision Making," *Journal of Money, Credit, and Banking* (forthcoming).

Charness, Gary and Dan Levin. (2005). "When Optimal Choices Feel Wrong: A Laboratory Study of Bayesian Updating, Complexity, and Psychological Affect," *American Economic Review*, 95, 1300-1309.

Charness, Gary, and Dan Levin. (2006). "The Origin of the Winner's Curse: A Laboratory Study," WP UCSB.

Cooper, David, and John Kagel. (2005). "Are Two Heads Better Than One? Team versus Individual Play in Signaling Games," *American Economic Review*, 95, 477-509.

Ganguly, Ananda, John Kagel, and Donald Moser. (2000). "Do Asset Market Prices Reflect Traders' Judgment Biases," *Journal of Risk and Uncertainty*, 20, 219-245.

Grether, David. (1980). "Bayes' Rule and the Representativeness Heuristic: Some Experimental Evidence," *Journal of Economic Behavior and Organization,* 17, 31-57.

Grether, David. (1992). "Testing Bayes' Rule as a Descriptive Model: The Representativeness Heuristic," *Quarterly Journal of Economics*, 95, 537-557.

Kagel, John, and Dan Levin. (1986). "The Winner's Curse and Public Information in Common Value Auctions," *American Economic Review*, 76, 894-920.

Kahneman, Daniel, and Amos Tversky. (1972). "Subjective Probability: A Judgment of Representativeness," *Cognitive Psychology*, 3, 430-454.

Kerr, Norbert, Robert MacCoun, and Geoffrey Kramer. (1996). "Bias in Judgment: Comparing Individuals and Groups," *Psychological Review* ,103, 687-719.

Ouwersloot, Hans, Peter Nijkamp, and Piet Rietveld (1998). "Errors in Probability Updating Behaviour: Measurement and Impact Analysis," *Journal of Economic Psychology*, 19, 535-563.

Tversky, Amos, and, Daniel Kahneman. (1971). "Belief in the Law of Small Numbers," *Psychological Bulletin*, 76, 105-110.

Tversky, Amos, and, Daniel Kahneman. (1973). "Availability: A Heuristic for Judging Frequency and Probability," *Cognitive Psychology*, 5, 207-232.

Zizzo, Daniel, Stephanie Stolarz-Fantino, Julie Wen, and Edmund Fantino. (2000). "A Violation of the Monotonicity Axiom: Experimental Evidence on the Conjunction Fallacy," *Journal of Economic Behavior and Organization*, 41, 263-276.