# Is probability matching smart?
# Associations between probabilistic choices
# and cognitive ability

RICHARD F. WEST
*James Madison University, Harrisonburg, Virginia*

and

KEITH E. STANOVICH
*University of Toronto, Toronto, Ontario, Canada*

In three experiments involving over 1,500 university students ($n = 1,557$) and two different probabilistic choice tasks, we found that the utility-maximizing strategy of choosing the most probable alternative was not the majority response. In a story problem version of a probabilistic choice task in which participants chose from among five different strategies, the maximizing response and the probability-matching response were each selected by a similar number of students (roughly 35% of the sample selected each). In a more continuous, or trial-by-trial, task, the utility-maximizing response was chosen by only one half as many students as the probability-matching response. More important, in both versions of the task, the participants preferring the utility-maximizing response were significantly higher in cognitive ability than were the participants showing a probability-matching tendency. Critiques of the traditional interpretation of probability matching as nonoptimal may well help explain why some humans are drawn to the nonmaximizing behavior of probability matching, but the traditional heuristics and biases interpretation can most easily accommodate the finding that participants high in computational ability are more likely to carry out the rule-based cognitive procedures that lead to maximizing behavior.

An important research tradition in the cognitive psychology of reasoning—the heuristics and biases approach—has firmly established that people's responses often deviate from the performance considered normative on many reasoning tasks (for summaries of the large literature, see Baron, 2000; Dawes, 1998; Evans & Over, 1996; Kahneman, Slovic, & Tversky, 1982; Kahneman & Tversky, 2000; Manktelow, 1999; Shafir & LeBoeuf, 2002; Stanovich, 1999). It has been common for these empirical demonstrations of a gap between descriptive and normative models of reasoning and decision making to be taken as indications that systematic irrationalities characterize human cognition. However, over the last decade, an alternative interpretation of these findings has been championed by various theorists (Anderson, 1990, 1991; Brase, Cosmides, & Tooby, 1998; Chater & Oaksford, 2000, 2001; Cosmides

& Tooby, 1992, 1994, 1996; Gigerenzer, 1996a; Oaksford & Chater, 1998, 2001; Rode, Cosmides, Hell, & Tooby, 1999; Todd & Gigerenzer, 2000). They have reinterpreted the modal response in most of the classic heuristics and biases experiments as indicating an optimal information-processing adaptation on the part of the participants. For example, Rode et al. (1999) have argued that error in human probabilistic judgments is more apparent than real.

As an example, consider the probabilistic contingency experiment, which has many versions in psychology (Fantino & Esfandiari, 2002; Gal & Baron, 1996; Peterson & Ulehla, 1965; Shanks, Tunney, & McCarthy, 2002; Tversky & Edwards, 1966). In one version, the participant sits in front of two lights (one red and one blue) and is told that he or she is to predict which of the lights will be flashed on each trial and that there will be several dozen such trials (participants are often paid money for correct predictions). The experimenter has actually programmed the lights to flash randomly, with the provision that the red light will flash 70% of the time, and the blue light 30% of the time. Participants do quickly pick up the fact that the red light is flashing more, and they predict that it will flash on more trials than they predict that the blue light will flash. Most often, they switch back and forth, predicting the red light roughly 70% of the time and the blue light roughly 30% of the time. This strategy of probability matching is suboptimal because it ensures that, in this example, the participant will predict correctly only 58% of the time ($.7 \times .7 + .3 \times$

.3), as compared with the 70% hit rate achieved by predicting the more likely color on each trial.

An alternative procedure, in the tradition of research exploring human probability judgments (Kahneman & Tversky, 1972), involves presenting verbal problems in which the frequencies of hypothetical outcomes are either directly given or easily inferable from the outset. Gal and Baron (1996), for example, asked participants what global strategy they would use when betting on the most likely outcome for each roll of a die in a hypothetical game. In this game, a die with four red faces and two green faces was to be rolled several times. Approximately two thirds of college students failed to use the maximizing strategy of predicting the most probable color for each roll of the die (Gal & Baron, 1996).

Although the response of probability matching has been viewed as irrational in the heuristics and biases literature (see Fiorina, 1971; Gal & Baron, 1996; Shanks et al., 2002), this response pattern has been defended for a variety of reasons ranging from evolutionary concerns to concerns that the normative models being applied to the task are wrong (e.g., Cooper, 1989; Cooper & Kaplan, 1982; Fiorina, 1971; Gigerenzer, 1996b; Keren & Wagenaar, 1985; Rode et al., 1999; Skyrms, 1996).

One aspect of performance that has not received enough attention in these debates about response patterns in the heuristics and biases literature has been individual differences. We have previously demonstrated (e.g., Stanovich, 1999; Stanovich & West, 1998a, 1999, 2000) that the nature of these individual differences and their patterns of covariance might have implications for explanations of why human behavior often departs from normative models. With respect to several other effects in the heuristics and biases literature (e.g., the conjunction fallacy, selection task performance, or belief bias effects), we have previously observed that the modal response on the task was not the response given by those of higher cognitive ability (i.e., participants closer to the modal response were lower, rather than higher, in cognitive ability). We have argued that debates about which responses are optimal or prescriptive on probabilistic reasoning tasks might be leavened by a more detailed knowledge of just *who* was making which response and of how these people responded on other indicators of cognitive functioning. For example, it might be assumed that the responses of organisms with fewer algorithmic limitations would be closer to the prescriptive response because the optimal strategy is computationally more complex and only those with the requisite computational power are able to compute it. Alternatively, the optimal strategy might not be more computationally complex, but its greater efficiency might be more readily recognized as such by those with more intelligence. Either way, we might expect individual differences in cognitive capacity to predict differential responding—a trend that has been observed on other heuristics and biases tasks (Byrnes & Overton, 1986; Jepson, Krantz, & Nisbett, 1983; Stanovich & West, 1998a, 2000).

Theorists in the heuristics and biases camp are largely committed to this prediction because they often explain divergences between normative models and actual performance by claiming that limitations in computational capacity prevent the normative response. When the modal response is not normative under their interpretation, we would expect the more intelligent participants to disproportionately give the nonmodal response (a trend that has sometimes, but not always, been obtained; Stanovich & West, 2000).

In three experiments, we examined whether the probability-matching response in probabilistic choice tasks is an example of a modal response not chosen by the most intelligent individuals. The first two experiments used a probabilistic choice task adapted from the die problem of Gal and Baron (1996), described above. In Experiment 1, the participants were asked to identify the global strategy they would use in trying to maximize their winnings when predicting the outcome of each roll of a hypothetical die with four red faces and two green faces. Experiment 2 replicated the method of Experiment 1 and added a measure that probed the beliefs about probabilities of participants choosing different strategies. In contrast to Experiments 1 and 2, which examined holistic or global strategies, Experiment 3 used a more continuous, or trial-by-trial, betting task that asked the participants to indicate their specific predictions for each of 10 cards after a random shuffle of a hypothetical 10-card deck that contained 7 "a" and 3 "b" cards. The card task also differed from the die task in that, although predicting the most probable letter for each trial was the utility-maximizing strategy and would, on the average, result in winning the most money, a matching strategy might appear especially tempting because it offered at least the possibility of predicting the outcome on more than seven trials. Previous research using story problems has rarely been used to examine the cognitive characteristics of individuals making differing probabilistic choices. The present research explores how cognitive abilities and beliefs about probabilities differ for participants making optimal or suboptimal choices in probabilistic choice tasks.

## EXPERIMENT 1

### Method

**Participants**. The participants were 445 undergraduate students (108 males and 337 females) recruited through an introductory psychology participant pool at a medium-sized state university. Their mean age was 18.8 years (*SD* = 2.2). The majority of these students were freshmen (267 students) or sophomores (120 students), and almost 90% of them identified themselves as White (398 White, 16 African American, 21 Asian American, and 10 other). The demographics form filled out by the students included questions on their educational history in mathematics and statistics courses. We constructed a 0- to 4-point scale that assessed the students' mathematics/ statistics course background. Students received 1 point if they had taken a statistics course in college (193 students), 1 point if they had taken a statistics course in high school (80 students), 1 point if they had taken a mathematics course at the university level (288 students), and 1 point if they had had at least 4 years of high school mathematics (399 students). The mean score on this mathematics experience composite was 2.16 (*SD* = 0.92).

**Cognitive ability measure**. Students were asked to indicate their verbal and mathematical Scholastic Aptitude Test (SAT) scores on the demographics sheet. The mean reported verbal SAT score of the

students was 589 ($SD = 65$), the mean reported mathematical SAT score was 589 ($SD = 67$), and mean total SAT score was 1,178 ($SD = 104$). These self-reported scores closely matched the averages for this institution (582, 587, and 1,169, respectively; James Madison University Office of Institutional Research, 1999–2000). A further indication of the validity of such self-reported scores was obtained by Stanovich and West (1998b), who found that the correlation between a vocabulary test and self-reported SAT total scores (.49) was quite similar to the .51 correlation between the vocabulary test and *verified* total SAT scores in a previous investigation using the same vocabulary measure (West & Stanovich, 1991). A final indication of the validity of the SAT self-reports is that the vocabulary test in the Stanovich and West (1998b) study displayed a higher correlation with the verbal SAT scores (.61) than with the mathematical SAT scores (.13). In the experiment that follows, the relationships with verbal SAT and mathematical SAT scores were very similar (there were no significant interactions with SAT subscale). Thus, the more reliable total SAT score will be used in most of the analyses to be reported.

Our reliance on the SAT as a measure of cognitive ability is justified by the fact that it loads highly on psychometric *g* (general intelligence) and that it is strongly associated with working memory—the quintessential indicator of computational capacity in cognitive science (Engle, Tuholski, Laughlin, & Conway, 1999; Kyllonen, 1996).

**Probabilistic choice task: Dice problem**. The probabilistic choice task used in this investigation was a dice problem adapted from the work of Gal and Baron (1996). The problem was stated as follows:

A die with 4 red faces and 2 green faces will be rolled 60 times. Before each roll you will be asked to predict which color (red or green) will show up once the die is rolled. You will be given one dollar for each correct prediction. Assume that you want to make as much money as possible. What strategy would you use in order to make *as much money as possible* by making the most correct predictions?
Strategy A: Go by intuition, switching when there has been too many of one color or the other.
Strategy B: Predict the more likely color (red) on most of the rolls but occasionally, after a long run of reds, predict a green.
Strategy C: Make predictions according to the frequency of occurrence (4 of 6 for red and 2 of 6 for green). That is, predict twice as many reds as greens.
Strategy D: Predict the more likely color (red) on all of the 60 rolls.
Strategy E: Predict more red than green, but switching back and forth depending upon "runs" of one color or the other.
Which Strategy is best? ___

Strategy C is the probability-matching (Match) strategy and Strategy D is the maximizing (Max) strategy. Strategy A is in essence an intuitive strategy that mixes a touch of the gambler's fallacy with intuition (GF-Intuition). Strategy B recognizes *runs*, a gambler's fallacy, with almost matching (GF-Almost), and Strategy E more purely reflects the gambler's fallacy (GF-Pure).

## Results and Discussion

Table 1 displays the strategy choices of the participants in Experiment 1, along with the mean SAT total scores of the individuals choosing each strategy. The most popular strategy choice ($n = 168$) was Match. However, a close second in popularity ($n = 150$) was the Max response of choosing the most probable option on each trial. The other three strategies, which tend to reflect the gambler's fallacy, received fewer choices (combined $n = 127$) than did the two dominant responses.

As the third column of the table indicates, the participants choosing the Max strategy tended to have higher SAT scores than did the participants choosing one of the other strategies. An analysis of variance on the SAT scores confirmed this impression by indicating that there was a significant overall effect of strategy [$F(4,440) = 8.08, MS_e = 10,087.33, p < .001$]. Students choosing the Max strategy had higher SAT scores (1,215, $SD = 104$) than did those choosing the Match (1,161, $SD = 95$), GF-Almost (1,163, $SD = 92$), GF-Pure (1,147, $SD = 106$), and GF-Intuition (1,151, $SD = 128$) strategies. Scheffé post hoc tests indicated that students choosing the Max strategy had significantly higher SAT scores than did those choosing the Match ($p < .001$), GF-Almost ($p < .05$), and GF-Pure ($p < .01$) strategies (not surprisingly, the difference involving the GF-Intuition strategy did not achieve significance, because only 15 students selected this strategy). As is apparent in the fourth column of the table, the mathematics experience composite scores of the five strategy groups were very similar [$F(4,440) < 1, MS_e = 0.85$, n.s.]. The fifth column of the table indicates that there were no appreciable differences between the five strategy groups in the number of university level statistics course they had taken [$F(4,440) \le 1, MS_e = 0.25$, n.s.].

Gal and Baron (1996) reported that the Max response was given more often by male students in their study than by female students. We replicated this finding [$\chi^2(4, N = 445) = 27.9, p < .001$], as is indicated in Table 1. The Max

**Table 1**
**Mean Scholastic Aptitude Test (SAT) Scores, Mathematics Experience Composite Scores, and Number of University Statistics Courses Taken, as a Function of Strategy Choice, and Number and Percentages of Males (M) and Females (F) Using the Strategies, in Experiment 1**

| Prediction | | Score | | No. of | Sex | | | |
| Strategy | *n* | SAT Total | Math Composite | Statistics Courses | M *n* | M % | F *n* | F % |
|---|---|---|---|---|---|---|---|---|
| GF-Intuition | 15 | 1,151 | 2.4 | 0.53 | 3 | 3 | 12 | 4 |
| GF-Almost | 64 | 1,163$_a$ | 2.2 | 0.38 | 13 | 12 | 51 | 15 |
| GF-Pure | 48 | 1,147$_a$ | 2.0 | 0.38 | 11 | 10 | 37 | 11 |
| Match | 168 | 1,161$_a$ | 2.1 | 0.47 | 23 | 21 | 145 | 43 |
| Max | 150 | 1,215$_b$ | 2.1 | 0.43 | 58 | 54 | 92 | 27 |

Note—GF-Intuition, gambler's fallacy with intuition strategy; GF-Almost, gambler's fallacy with an almost-matching strategy; GF-Pure, pure gambler's fallacy strategy; Match, probability matching strategy; Max, maximizing strategy. Means in the same row that show different subscripts (a, b) are significantly different at $p < .05$ in the Scheffé post hoc test.

strategy was chosen by a majority of males in the study (54%), but only by a minority of the females (27%). Although female students' total SAT scores ($M = 1,169$) were somewhat lower than those of the male students ($M = 1,203$), this difference did not appear to account for all of the sex differences observed in strategy. For example, when strategy was converted to a 0/1 variable (Max = 1, and all other strategies = 0) and was simultaneously regressed on total SAT score and sex, both variables remained significant predictors of strategy [$t(442) = 5.05$, $p < .001$, and $t(442) = 4.57$, $p < .001$, respectively]. The results were essentially similar when the regression was run using the mathematical SAT score rather than the total score [$t(442) = 4.50$, $p < .001$, and $t(442) = 4.75$, $p < .001$, respectively].

## EXPERIMENT 2

Experiment 1 provided another example of the trend discussed in the introduction: The modal response in the probabilistic choice task was not the response of the most cognitively able participants.[1] Across the entire sample, the participants choosing the instrumentally rational Max response were higher in cognitive ability than were those choosing the evolutionarily rational Match response. In Experiment 2, we sought to solidify this finding by examining performance with a larger sample size and by probing the beliefs about probability of people choosing different strategies. Subsequent to choosing a strategy, the participants responded to a questionnaire that probed various beliefs about the trials and contingencies in the hypothetical die-throwing situation. In addition, the large sample size of Experiment 2 gave us a chance to examine the replica-

bility of the rather large and surprising sex difference displayed in the dice problem.

### Method

**Participants**. The participants were 715 undergraduate students (168 males and 547 females) recruited through an introductory psychology participant pool at a medium-sized state university. None of these students had participated in Experiment 1. Their mean age was 18.9 years ($SD = 2.1$). The majority of these students were freshmen (394 students) or sophomores (242 students), and almost 90% of them identified themselves as White (633 White, 20 African American, 40 Asian American, and 22 other). The demographics form contained the same questions on educational history in mathematics and statistics courses as in Experiment 1.

**Cognitive ability measure**. The mean reported verbal SAT score of the students was 581 ($SD = 65$), the mean reported mathematical SAT score was 585 ($SD = 68$), and the mean total SAT score was 1,167 ($SD = 104$).

**Probabilistic choice task: Dice problem**. The dice problem from Experiment 1 was also used in the present experiment as the probabilistic choice task.

**Probabilistic belief questionnaire**. Subsequent to choosing a strategy in the probabilistic choice task, the participants completed a seven-item questionnaire that probed various beliefs about the trials and contingencies in the hypothetical die-throwing situation. The response format for each item was *agree strongly* (scored as 6), *agree moderately* (5), *agree slightly* (4), *disagree slightly* (3), *disagree moderately* (2), and *disagree strongly* (scored as 1). Students received the following instructions for this task: "After considering your choice of the best strategy, how much do you agree or disagree with the following statements?" The specific questionnaire items are listed in Table 2.

### Results and Discussion

Table 3 displays the strategy choices of the participants in Experiment 2, along with the mean SAT total scores of the individuals choosing each strategy. There it can be

**Table 2**
**Mean Agreement Score as a Function of Probabilistic Belief Item and Die Prediction Strategy in Experiment 2**
**(With Correlation Between Item and SAT Total Score Shown in Parentheses)**

| | Die Prediction Strategy | | | | |
| | GF-Comb ($n = 175$) | Match ($n = 260$) | Max ($n = 280$) | $MS_e$ | $F(2,712)$ |
| Questionnaire Item | | | | | |
|---|---|---|---|---|---|
| 1. My method is best because it leaves more room for intuition to help me. ($-.26$) | $4.20_a$ | $2.91_b$ | $2.10_c$ | 1.60 | 149.05* |
| 2. After a long run of red, green is more likely, so after a long run of red I should switch my prediction to green. ($-.22$) | $4.03_a$ | $3.21_b$ | $2.30_c$ | 1.64 | 102.24* |
| 3. Because in the long run two thirds of the rolls will come out red, I should predict red on two thirds of the 60 rolls. ($-.21$) | $4.03_a$ | $4.62_b$ | $3.06_c$ | 1.76 | 94.65* |
| 4. Red is more likely to come up on *any* one roll, given that there are more red sides on the die than green. (.22) | $5.09_a$ | $5.14_a$ | $5.60_b$ | 0.79 | 25.49* |
| 5. I should predict red on most rolls, but *randomly* predict green on a few others. ($-.26$) | $4.59_a$ | $3.87_b$ | $3.26_c$ | 1.47 | 64.98* |
| 6. I thought that the outcome of any one roll of the die is unrelated to the previous rolls. (.13) | $4.44_a$ | $4.50_a$ | $5.15_b$ | 2.01 | 19.51* |
| 7. The method I thought was best is preferred because in the long run it *could* provide a higher gain, although at a higher risk. ($-.21$) | $4.25_a$ | $3.93_a$ | $3.40_b$ | 1.91 | 22.39* |

Note—GF-Comb, the three gambler's fallacy strategies are combined; Match, probability matching strategy; Max, maximizing strategy. Means in the same row that show different subscripts (a, b, c) are significantly different at $p < .05$ in the Scheffé post hoc test. The number in parentheses is the correlation between that item and SAT total score. All of these correlations are significant at the .001 level.    *$p < .001$.

**Table 3**
**Mean Scholastic Aptitude Test (SAT) Scores, Mathematics Experience Composite Scores, and Number of University Statistics Courses Taken, as a Function of Strategy Choice, and Number and Percentages of Males (M) and Females (F) Using the Strategies in Experiment 2**

| Prediction Strategy | n | Score | | No. of Statistics Courses | Sex | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | M | | F | |
| | | SAT Total | Math Composite | | n | % | n | % |
| GF-Intuition | 22 | 1,169 | 2.4 | 0.50 | 5 | 3 | 17 | 3 |
| GF-Almost | 83 | 1,140$_a$ | 2.3 | 0.48 | 18 | 11 | 65 | 12 |
| GF-Pure | 70 | 1,137$_a$ | 2.2 | 0.53 | 16 | 10 | 54 | 10 |
| Match | 260 | 1,152$_a$ | 2.4 | 0.55 | 29 | 17 | 231 | 42 |
| Max | 280 | 1,196$_b$ | 2.3 | 0.46 | 100 | 60 | 180 | 33 |

Note—GF-Intuition, gambler's fallacy with intuition strategy; GF-Almost, gambler's fallacy with an almost-matching strategy; GF-Pure, pure gambler's fallacy strategy; Match, probability matching strategy; Max, maximizing strategy. Means in the same row that show different subscripts (a, b) are significantly different at $p < .05$ in the Scheffé post hoc test.

seen that the most popular strategy choice ($n = 280$) was the Max response of choosing the most probable option on each trial. However, a close second in popularity ($n = 260$) was the Match strategy. As in Experiment 1, the other three strategies, which tend to reflect the gambler's fallacy, received substantially fewer choices (combined $n = 175$) than did the two dominant responses.

As the third column of Table 3 indicates, the participants choosing the Max strategy tended to have higher SAT scores than did the participants choosing one of the other strategies. An analysis of variance on the SAT scores confirmed this impression by indicating that there was a significant overall effect of strategy [$F(4,709) = 9.91$, $MS_e = 10,372.15, p < .001$]. The students choosing the Max strategy had higher SAT scores (1,196, $SD = 103$) than those choosing the Match (1,152, $SD = 100$), GF-Almost (1,140, $SD = 114$), GF-Pure (1,137, $SD = 92$), and GF-Intuition (1,169, $SD = 84$) strategies. Scheffé post hoc tests indicated that the students choosing the Max strategy had significantly higher SAT scores than did those choosing the Match ($p < .001$), GF-Almost ($p < .01$), and GF-Pure ($p < .01$) strategies (not surprisingly, the difference involving the GF-Intuition strategy did not achieve significance, because only 22 students selected this strategy). As is apparent in the fourth column of Table 3, the mathematics experience composite scores of the five strategy groups were very similar [$F(4,709) = 1.24, MS_e = 0.76$, n.s.], as were the number of university level statistics course they had taken [Table 3, fifth column; $F(4,709) = 1.21, MS_e = 0.25$, n.s.].

In Experiment 2, we confirmed the surprising finding (first demonstrated by Gal & Baron, 1996) that the Max response was given more often by male participants [$\chi^2(4, N = 715) = 45.5, p < .001$]. The Max strategy was chosen by a majority of males in the study (60%), but only by a minority of the females (33%). In contrast, the Match response was chosen only by a minority of the males in the study (17%), but the Match response was chosen by a plurality of the females (42%). As in Experiment 1, the female students' total SAT scores ($M = 1,160$) were somewhat lower than those of the male students ($M = 1,187$),

but this difference did not account for all of the sex differences observed in strategy choice. When strategy was converted to a 0/1 variable (Max = 1, and all other strategies = 0) and was simultaneously regressed on total SAT score and sex, both variables remained significant predictors of strategy [$t(712) = 5.56, p < .001$, and $t(712) = 5.84, p < .001$, respectively]. The results were essentially similar when the regression was run using the mathematical SAT score, rather than the total score [$t(712) = 5.31$, $p < .001$, and $t(712) = 5.35, p < .001$, respectively].

The participants in Experiment 2 filled out the probabilistic belief questionnaire subsequent to performing the dice task. Table 2 presents the mean scores on the 6-point (*disagree strongly* to *agree strongly*) scale for each of the items on this questionnaire as a function of strategy. In that table, the Match and the Max strategies are compared with a combined gambler's fallacy (GF-Comb) strategy that was formed by combining the set of three strategies that place an emphasis on the gambler's fallacy (GF-Intuition, GF-Almost, and GF-Pure)—that is, switching predictions on the basis of *runs*.

The three strategy groups differed significantly on each of the seven probabilistic belief items [from $F(4,709) = 19.51$ ($MS_e = 2.01$) to 149.05 ($MS_e = 1.60$); all $p$s < .001]. The students in the GF-Comb group were significantly more likely to agree that intuition was important (Item 1), that a color would become less likely after a long run (Item 2), and that it was important to randomly predict the less likely color on some trials (Item 5). The students in the Max group were least likely to agree with each of these three beliefs.

The Match group was significantly more likely to agree that it was important to match the most probable long-run outcome pattern (Item 3), whereas the Max group was least likely to agree with this belief. Although the GF-Comb and the Match groups did not differ significantly from each other in their belief that higher gain could be achieved by sometimes selecting the riskier color (Item 7), the students in both of these groups were significantly more likely to agree with this item than were the students in the Max group (Scheffé post hoc test, both $p$s < .001).

Over 95% of the students in each of the three strategy groups agreed to various degrees (*agree slightly* to *agree strongly*) that red was the more likely outcome of any one roll (Item 4), but the Max group was significantly firmer in this belief. Over 70% of the students in each strategy group agreed to various degrees that the outcome of any roll of the die was independent of the previous rolls (Item 6), but again, the students in the Max group were significantly stronger in their affirmation. In fact, in contrast to both the GF-Comb and the Match groups, the *only* beliefs strongly endorsed by the Max group were those that emphasized the higher probability of red on each independent roll of the die (Items 4 and 6).[2]

## EXPERIMENT 3

Although in Experiment 2 slightly more participants selected the normative Max response over the Match response ($n = 280$ vs. 260), the Max responders constituted only 39% of the total sample. The gambler's fallacy responses—particularly Strategy B and Strategy E, it should be noted—also suggest a Match rather than a Max strategy. Although only a minority of the participants chose the Max strategy, those that did displayed higher SAT scores.

Before elaborating our interpretation of these results, we will report a third experiment, in which we examined the generality of the patterns observed in probabilistic choice tasks. The paradigm for probabilistic choice that we employed in the first two experiments involved a story problem with a holistic, or global, strategy choice. Following the task used by Gal and Baron (1996), the participants in Experiments 1 and 2 were asked to indicate which single global strategy they would use in predicting the color of each of 60 rolls of a die that had four red faces and two green faces. In Experiment 3, we employed a more continuous, or trial-by-trial, task that asked the participants to indicate what their specific prediction would be for each of 10 cards after a random shuffle of a 10-card deck that contained 7 cards with the letter "a" and 3 cards with the letter "b" on the down side. Although the Max strategy of selecting "a" for each card is the optimal strategy with a fixed hit rate of 70% and the Match strategy would result in an average hit rate of only 58%, the latter strategy may appear additionally tempting because it offers the possibility of predicting the outcome on more than seven trials. We were interested in whether the patterns of individual differences in cognitive ability (and the sex difference) would replicate in a paradigm with quite different cognitive processing requirements.

## Method

### Participants

The participants were 422 undergraduate students (173 males and 249 females) recruited through an introductory psychology participant pool at a medium-sized state university. None of these students had participated in Experiment 1 or 2. Their mean age was 19.1 years ($SD = 1.8$). The majority of these students were freshmen (272 students) or sophomores (109 students), and almost 90% of them iden-

tified themselves as White (368 White, 10 African American, 32 Asian American, and 12 other). The demographics form contained the same questions on educational history in mathematics and statistics courses as in Experiment 1.

**Cognitive ability measure**. The students were asked to indicate their verbal and mathematical SAT scores on the demographics sheet. The mean self-reported verbal SAT score of the students was 581 ($SD = 78$), the mean reported mathematical SAT score was 594 ($SD = 69$), and the mean total SAT score was 1,174 ($SD = 105$).

**Probabilistic choice task: Trial-by-trial task**. The probabilistic choice task asked the participants to indicate what their specific prediction would be for each of 10 cards following a random shuffle of a 10-card deck that contained 7 cards with the letter "a" and 3 cards with the letter "b" on the down side. The problem was stated as follows:

Card Guessing Game
A card deck has only 10 cards:
7 of the cards have the letter "**a**" on the down side.
3 of the cards have the letter "**b**" on the down side.
The 10 cards are randomly shuffled.
Your task is to guess the letter on the down side of each card before it is turned over. Pretend that you will win $100 for each card's down side letter you correctly predict. Indicate your predictions for each of the 10 cards:

1.    Card #1 will be **a** or **b**?
2.    Card #2 will be **a** or **b**?
3.    Card #3 will be **a** or **b**?
4.    Card #4 will be **a** or **b**?
5.    Card #5 will be **a** or **b**?
6.    Card #6 will be **a** or **b**?
7.    Card #7 will be **a** or **b**?
8.    Card #8 will be **a** or **b**?
9.    Card #9 will be **a** or **b**?
10.   Card #10 will be **a** or **b**?

The students who predicted any combination of 7 "a" cards and 3 "b" cards used the Match strategy, and the students who predicted "a" for each of the 10 cards used the Max strategy. Strategies other than Match and Max were used very rarely with this paradigm and were dropped from the analyses that follow (only 5.9% of the sample gave a pattern other than Max or Match: 7 students selected "b" for 1 or 2 of the cards, and 18 students selected "b" on more than 3 of the cards). The remaining sample contained 397 students.

## Results and Discussion

Table 4 displays the strategy choices of the participants in Experiment 3, along with the mean SAT total scores of the individuals choosing between the Match and the Max strategies. The Match strategy for the trial-by-trial card task was preferred by about two thirds of the students ($n = 264$). In contrast, the Max strategy of choosing the most probable option on each trial was preferred by only about a third of the students ($n = 133$). As the third column of Table 4 indicates, the participants choosing the Max strategy tended to have significantly higher SAT scores (1,202, $SD = 104$) than did the participants choosing the Match strategy [1,163, $SD = 102$; $t(395) = 3.57, p < .001$]. The mathematics experience composite scores of the two strategy groups were very similar [$t(395) = -0.83$, n.s.], and there were no appreciable differences between the two strategy groups in the number of university level statistics course they had taken [two-tailed $t(395) = -0.28$, n.s.].

Once again, the Max response was given more often by the male students than by the female students [$\chi^2(1, N = 397) = 6.40, p < .05$], as is indicated in the second row of Table 4. The Max strategy was chosen by 41% of the males, but only by 28% of the females. As in Experiments 1 and

**Table 4**
**Mean Scholastic Aptitude Test (SAT) Scores, Mathematics Experience Composite Scores,**
**and Number of University Statistics Courses Taken, as a Function**
**of Trial-by-Trial Card Prediction Strategy Choice, and Numbers and Percentages**
**of Males (M) and Females (F) Using the Strategies, in Experiment 3**

| | | Score | | No. of | Sex | | | |
| | | | | | M | | F | |
| Prediction Strategy | $n$ | SAT Total | Math Composite | Statistics Courses | $n$ | % | $n$ | % |
|---|---|---|---|---|---|---|---|---|
| Match | 264 | 1,163 | 2.3 | 0.50 | 98 | 59 | 166 | 72 |
| Max | 133 | 1,202 | 2.3 | 0.48 | 67 | 41 | 66 | 28 |

Note—Match, probability matching strategy; Max, maximizing strategy.

2, female students' total SAT scores ($M = 1,164$) were somewhat lower than those of the male students ($M = 1,192$), but this difference did not appear to account for all of the sex differences observed in strategy. When strategy was converted to a 0/1 variable (Match = 0, Max = 1) and was simultaneously regressed on total SAT score and sex, both variables remained significant predictors of strategy [$t(394) = 3.29, p < .005$, and $t(394) = 2.13, p = .034$, respectively]. The results were essentially similar when the regression was run using the mathematical SAT score rather than the total score [$t(394) = 3.68, p < .001$, and $t(394) = 2.04, p = .042$, respectively].[3]

## GENERAL DISCUSSION

In three different experiments utilizing two different probabilistic choice paradigms, we observed that the Max response in both tasks was the choice of the most cognitively able participants. This was despite the fact that the nonmaximizing Match response tended to be the modal response. On the trial-by-trial version of the task (Experiment 3), it was by far the dominant response. It was chosen by a plurality of the participants in Experiment 1 and was at least equal in magnitude to the Max strategy in Experiment 2 (recall that several of the GF strategy choices incorporate matching components).

This pattern—the modal response on the task being the response given by those of lower cognitive ability—is not without precedent (Jepson et al., 1983; Stanovich & West, 1998a, 2000). We interpret the pattern in this experiment and others (e.g., Stanovich, 1999) as consistent with assumptions behind the heuristics and biases framework (e.g., Gilovich, Griffin, & Kahneman, 2002; Kahneman & Frederick, 2002; Kahneman et al., 1982) that the default processing strategy of most participants is a nonnormative cognitive shortcut that does not stress cognitive capacity. Participants who, on an individual difference analysis, are higher in cognitive capacity are less likely to make use of this cognitive shortcut. In situations like the probabilistic choice tasks used here—where the heuristic primes a nonnormative response—high analytic intelligence will be more strongly associated with the instrumentally rational response.[4]

The results from the probabilistic belief questionnaire employed in Experiment 2 provided strong evidence that the Max strategy group held beliefs about the implications of the probabilistic processes that were remarkably different from those held by the two other groups. The GF-Comb group strongly endorsed the importance of using intuition, making some random or risky choices, and avoiding long runs or streaks of one color. The Match group tended to be somewhat more restrained in their endorsement of these beliefs, while at the same time very strongly endorsing the view that the most optimal pattern of choices needed to match the most likely long-run pattern. The strategies of both the GF-Comb and the Match groups appear to have been strongly influenced by their knowledge that, although red may have been the more likely outcome of any specific roll of the die, a generally mixed pattern of red and green was still the most likely, or representative (see Kahneman & Frederick, 2002; Kahneman & Tversky, 1972), long-run outcome. The GF-Comb group appeared to be focused primarily on avoiding runs or streaks, whereas the Match group appeared to be focused primarily on what the most likely global pattern of results would generally look like. In contrast to both the GF-Comb and the Match groups, the only beliefs strongly endorsed by the Max group were those that emphasized the higher probability of red on each independent roll of the die (see Items 4 and 6 in Table 2).

Replicating the earlier finding of Gal and Baron (1996), we found that the male participants in each of our three experiments were significantly more likely to select the normative Max strategy than were the female participants. The male participants in Experiment 2 were also less likely than the female participants to endorse beliefs consistent with the gambler's fallacy. A portion of this male/female difference was accounted for by differences in cognitive ability, but a significant proportion of the variance in strategy selection was independently associated with sex in each of the three experiments. Evolutionary ideas in foraging and sexual selection theory that may be relevant to explaining the difference (see Geary, 1996) are beyond the scope of what can be adjudicated with the present data, but the literature on sex differences in mathematical ability may be relevant. Males, for example, are often found to perform somewhat better on mathematical reasoning tasks, whereas females are often found to perform somewhat better on mathematical calculation tasks (e.g., Geary, 1996; Kimura, 1999; but see Chipman, 1996). Although

our data contribute empirically to the description of this sex difference, we remain theoretically agnostic with respect to the origin of these differences.

## REFERENCES

ANDERSON, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.

ANDERSON, J. R. (1991). Is human cognition adaptive? *Behavioral & Brain Sciences*, **14**, 471-517.

BARON, J. (2000). *Thinking and deciding* (3rd ed.). Cambridge, MA: Cambridge University Press.

BRASE, G. L., COSMIDES, L., & TOOBY, J. (1998). Individuation, counting, and statistical inference: The role of frequency and whole-object representations in judgment under uncertainty. *Journal of Experimental Psychology: General*, **127**, 3-21.

BYRNES, J. P., & OVERTON, W. F. (1986). Reasoning about certainty and uncertainty in concrete, causal, and propositional contexts. *Developmental Psychology*, **22**, 793-799.

CHATER, N., & OAKSFORD, M. (2000). The rational analysis of mind and behaviour. *Synthèse*, **122**, 93-131.

CHATER, N., & OAKSFORD, M. (2001). Human rationality and the psychology of reasoning: Where do we go from here? *British Journal of Psychology*, **92**, 193-216.

CHIPMAN, S. F. (1996). Still far too sexy a topic. *Behavioral & Brain Sciences*, **19**, 248-249.

COOPER, W. S. (1989). How evolutionary biology challenges the classical theory of rational choice. *Biology & Philosophy*, **4**, 457-481.

COOPER, W. S., & KAPLAN, R. H. (1982). Adaptive coin-flipping: A decision-theoretic examination of natural selection for random individual variation. *Journal of Theoretical Biology*, **94**, 135-151.

COSMIDES, L., & TOOBY, J. (1992). Cognitive adaptations for social exchange. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind* (pp. 163-205). New York: Oxford University Press.

COSMIDES, L., & TOOBY, J. (1994). Beyond intuition and instinct blindness: Toward an evolutionarily rigorous cognitive science. *Cognition*, **50**, 41-77.

COSMIDES, L., & TOOBY, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, **58**, 1-73.

DAWES, R. M. (1998). Behavioral decision making and judgment. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 1, pp. 497-548). Boston: McGraw-Hill.

ENGLE, R. W., TUHOLSKI, S. W., LAUGHLIN, J. E., & CONWAY, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, **128**, 309-331.

EVANS, J. S. B. T., & OVER, D. E. (1996). *Rationality and reasoning*. Hove, U.K.: Psychology Press.

FANTINO, E., & ESFANDIARI, A. (2002). Probability matching: Encouraging optimal responding in humans. *Canadian Journal of Experimental Psychology*, **56**, 58-63.

FIORINA, M. P. (1971). A note on probability matching and rational choice. *Behavioral Science*, **16**, 158-166.

GAL, I., & BARON, J. (1996). Understanding repeated simple choices. *Thinking & Reasoning*, **2**, 81-98.

GEARY, D. C. (1996). Sexual selection and sex differences in mathematical abilities. *Behavioral & Brain Sciences*, **19**, 229-247.

GIGERENZER, G. (1996a). On narrow norms and vague heuristics: A reply to Kahneman and Tversky (1996). *Psychological Review*, **103**, 592-596.

GIGERENZER, G. (1996b). Rationality: Why social context matters. In P. B. Baltes & U. Staudinger (Eds.), *Interactive minds: Life-span perspectives on the social foundation of cognition* (pp. 319-346). Cambridge: Cambridge University Press.

GILOVICH, T., GRIFFIN, D., & KAHNEMAN, D. (Eds.) (2002). *Heuristics and biases: The psychology of intuitive judgment*. New York: Cambridge University Press.

JAMES MADISON UNIVERSITY OFFICE OF INSTITUTIONAL RESEARCH (1999–2000). *Table 2-20: Distribution of SAT scores of entering class by sex, fall 1998–99*. Retrieved November 11, 2001, from http://www.jmu.edu/instresearch/statsum/199900/2-20.htm.

JEPSON, C., KRANTZ, D., & NISBETT, R. (1983). Inductive reasoning: Competence or skill? *Behavioral & Brain Sciences*, **6**, 494-501.

KAHNEMAN, D., & FREDERICK, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49-81). New York: Cambridge University Press.

KAHNEMAN, D., SLOVIC, P., & TVERSKY, A. (Eds.) (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

KAHNEMAN, D., & TVERSKY, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, **3**, 430-454.

KAHNEMAN, D., & TVERSKY, A. (Eds.) (2000). *Choices, values, and frames*. Cambridge: Cambridge University Press.

KEREN, G., & WAGENAAR, W. A. (1985). On the psychology of playing blackjack: Normative and descriptive considerations with implications for decision theory. *Journal of Experimental Psychology: General*, **114**, 133-158.

KIMURA, D. (1999). *Sex and cognition*. Cambridge, MA: MIT Press.

KOKIS, J., MACPHERSON, R., TOPLAK, M., WEST, R. F., & STANOVICH, K. E. (2002). Heuristic and analytic processing: Age trends and associations with cognitive ability and cognitive styles. *Journal of Experimental Child Psychology*, **83**, 26-52.

KYLLONEN, P. C. (1996). Is working memory capacity Spearman's *g*? In I. Dennis & P. Tapsfield (Eds.), *Human abilities: Their nature and measurement* (pp. 49-76). Mahwah, NJ: Erlbaum.

MANKTELOW, K. I. (1999). *Reasoning and thinking*. Hove, U.K.: Psychology Press.

OAKSFORD, M., & CHATER, N. (1998). *Rationality in an uncertain world*. Hove, U.K.: Psychology Press.

OAKSFORD, M., & CHATER, N. (2001). The probabilistic approach to human reasoning. *Trends in Cognitive Sciences*, **5**, 349-357.

PETERSON, C. R., & ULEHLA, Z. (1965). Sequential patterns and maximizing. *Journal of Experimental Psychology*, **69**, 1-4.

RODE, C., COSMIDES, L., HELL, W., & TOOBY, J. (1999). When and why do people avoid unknown probabilities in decisions under uncertainty? Testing some predictions from optimal foraging theory. *Cognition*, **72**, 269-304.

SHAFIR, E., & LEBOEUF, R. A. (2002). Rationality. *Annual Review of Psychology*, **53**, 491-517.

SHANKS, D. R., TUNNEY, R. J., & MCCARTHY, J. D. (2002). A re-examination of probability matching and rational choice. *Journal of Behavioral Decision Making*, **15**, 233-250.

SKYRMS, B. (1996). *The evolution of the social contract*. Cambridge: Cambridge University Press.

STANOVICH, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah, NJ: Erlbaum.

STANOVICH, K. E., & WEST, R. F. (1998a). Individual differences in rational thought. *Journal of Experimental Psychology: General*, **127**, 161-188.

STANOVICH, K. E., & WEST, R. F. (1998b). Who uses base rates and *P*(D/~H)? An analysis of individual differences. *Memory & Cognition*, **26**, 161-179.

STANOVICH, K. E., & WEST, R. F. (1999). Discrepancies between normative and descriptive models of decision making and the understanding/acceptance principle. *Cognitive Psychology*, **38**, 349-385.

STANOVICH, K. E., & WEST, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral & Brain Sciences*, **23**, 645-726.

TODD, P. M., & GIGERENZER, G. (2000). Precis of Simple heuristics that make us smart. *Behavioral & Brain Sciences*, **23**, 727-780.

TVERSKY, A., & EDWARDS, W. (1966). Information versus reward in binary choice. *Journal of Experimental Psychology*, **71**, 680-683.

WEST, R. F., & STANOVICH, K. E. (1991). The incidental acquisition of information from reading. *Psychological Science*, **2**, 325-330.

## NOTES

1. Strictly speaking, this was true for females only, since only for this subgroup was the Match strategy the modal choice.

2. These were the only two questionnaire items for which the female students' mean agreement scores were not significantly higher than those of the male students [Items 4 and 6, respectively; two-tailed

$t(713) = 1.84$, $p = .066$, and $t(713) = 1.48$, $p = .138$; Items 1, 2, 3, 5, and 7, all two-tailed $t(713) = -3.20$ to $-6.12$; all $p$s $= .0014$ to $<.0001$]. This pattern is consistent with the weaker tendency toward the Max strategy in females.

3. We assumed that the participants would construe the task in Experiment 3 as one in which they were to make predictions with no knowledge of which letters had actually been on the down sides of previous cards. An insightful reviewer pointed out that our directions were somewhat ambiguous and that the participants might have construed the task as one in which the outcomes of previous predictions were revealed (i.e., imagined) prior to subsequent predictions. Under this alternative construal, if several "a" cards had already been revealed, the probability of a "b" might exceed that for an "a" and become optimal. The viability of this alternative construal was explored by examination of data from the first 4 cards, since "a" was always optimal under both construals for the first 4 cards. An examination of these data indicated that very few participants (3.3%) predicted the low probability letter "b" for Card 1. More important, contrary to the implications of the suggested alternative construal, the low-probability "b" was predicted more often for Card 3 (34.0%) than for any of the other 9 cards. Furthermore, the total number of "b" predictions for Cards 2–4 did not differ from the total number of "b" predictions for Cards 8–10 [$t(396) = 1.05$, n.s.]. Finally, when data from only the first 4 cards were analyzed, the pattern of SAT findings was essentially identical to the pattern reported in Table 4 with all 10 cards. Thus, the evidence unambiguously supports the view that the participants construed this task as we had intended and that "a" was always the optimal choice.

4. It should be noted that the range restriction in analytic intelligence of our university sample may have attenuated the magnitude of this association. Many of our participants had had at least minimal exposure to probability theory and notions of randomness in previous math and/or statistic courses, and the extent to which our findings can be generalized to populations without these educational experiences remains largely unexplored. Although there have been relatively few studies with noncollege students in which story-type probabilistic choice tasks have been used, some evidence suggests that increasing age and cognitive ability are associated with increased normative behavior in children on such tasks. For example, Gal and Baron (1996) studied a small sample of sixth- and ninth-grade children with a version of the die task and observed a nearly universal Match strategy. Kokis, Macpherson, Toplak, West, and Stanovich (2002) used a very different type of probabilistic reasoning task with 10- to 13-year-old children and found that the children who made optimal probabilistic decisions had significantly higher scores on a measure of cognitive ability than did children who made suboptimal decisions.