Motivated Reasoning and Performance on the was on Selection Task

Erica Dawson, Thomas Gilovich and Dennis T. Regan

Pers Soc Psychol Bull 2002; 28; 1379
DOI: 10.1177/014616702236869

Motivated Reasoning and Performance on the Wason Selection Task

Erica Dawson
Thomas Gilovich
Dennis T. Regan
Cornell University

People tend to approach agreeable propositions with a bias toward confirmation and disagreeable propositions with a bias toward disconfirmation. Because the appropriate strategy for solving the four-card Wason selection task is to seek disconfirmation, the authors predicted that people motivated to reject a task rule should be more likely to solve the task than those without such motivation. In two studies, participants who considered a Wason task rule that implied their own early death (Study 1) or the validity of a threatening stereotype (Study 2) vastly outperformed participants who considered nonthreatening or agreeable rules. Discussion focuses on how a skeptical mindset may help people avoid confirmation bias both in the context of the Wason task and in everyday reasoning.

> For a man always believes more readily that which he prefers.
>
> —Francis Bacon (1620/1955, p. 111)

Bacon is not alone in the observation that people's wishes have a powerful influence on their beliefs. This sentiment, long reflected in prose and proverb, has recently received a scientific label—motivated reasoning—and enthusiastic attention from psychologists. There is now a great deal of evidence that people are inclined to draw conclusions that suggest positive outcomes for themselves; provide support for pre-existing opinions; and confirm their status, success, and well-being (Kunda, 1990). To be sure, people are also motivated to reason accurately about the world and to present a face of impartiality, both to themselves and to others (Kunda, 1987). How do people reconcile these divergent goals of reaching desirable but impartial conclusions? In other words, how do motivated reasoners maintain an "illusion of objectivity" (Kunda, 1987; Pyszczynski & Greenberg, 1987)?

Gilovich (1991) characterized motivated reasoning in terms of the different standards of evidence people typically employ to evaluate propositions they wish to be true and propositions they wish to be false. Specifically, when evaluating an agreeable proposition, they tend to ask, in essence, "*Can* I believe this?" This constitutes a rather permissive evidential standard, because some supportive evidence can be found for all but the most outlandish propositions. By employing a partial or truncated search for evidence (Dawson, Gilovich, & Regan, 2002; Ditto & Lopez, 1992; Ditto, Scepansky, Munro, Apanovich, & Lockhart, 1998; Gilovich, Dawson, & Regan, 2002), a biased assemblage of evidence (Dunning, Meyerowitz, & Holzberg, 1989; Kunda, 1987; Lord, Ross, & Lepper, 1979), and relatively superficial processing of available information (Dawson et al., 2002; Ditto et al., 1998), motivated reasoners can very often find grounds for accepting a desirable hypothesis.

When evaluating a threatening or disagreeable proposition, in contrast, people tend to invoke a more stringent evidential standard, asking implicitly, "*Must* I believe this?" Seldom is a given body of evidence beyond criticism, whether because of the method by which it was collected, the analysis to which it was subjected, the pop-

Authors' Note: This research was supported by National Science Foundation (NSF) Grants SBR9319558 and SBR9809262 and a Jacob K. Javits Fellowship Program Grant P170B980005. Portions of this research were presented at the annual meeting of the Society for Judgment and Decision Making, November 2001, Orlando, Florida. We thank Amy Barcelo and Jamie Howard for their help in collecting these data. Correspondence concerning this article should be addressed to Erica Dawson, Johnson Graduate School of Management, Sage Hall, Cornell University, Ithaca, NY 14853; e-mail: ecd6@cornell.edu.

PSPB, Vol. 28 No. 10, October 2002 1379-1387
DOI: 10.1177/014616702236869

ulation to which it is generalized, or countless other potential shortcomings. Demanding such a strict standard of evidence typically leads to a relatively thorough search through all relevant information, maximizing the chances that any flaws or limitations of the data will be spotted (Dawson et al., 2002; Ditto et al., 1998; Ditto & Lopez, 1992). Thus, people who are motivated to reject an unpalatable proposition may well find cause for doing so.

The "Can I/Must I" model is not merely metaphorical but describes literally the ways in which motivated reasoners approach evidence. When asked to select the appropriate standard of evidence that should be used to decide empirical issues, those who wanted to reject the claim under consideration endorsed a paraphrased version of the "Must I?' criterion, whereas those who wanted to accept the claim endorsed a paraphrased version of the "Can I?" criterion (Dawson et al., 2002; Gilovich et al., 2002). Thus, when evaluating cherished beliefs, people apply an evaluative standard biased toward confirmation. When evaluating unpalatable beliefs, they apply a stricter standard in an attempt to disconfirm. Among other things, this suggests that skeptical reasoners—who implicitly ask "Must I believe this?"—are less likely to fall prey to the well-documented confirmation bias in everyday reasoning.

Confirmation bias, or the tendency to seek only or primarily evidence that would confirm a hypothesis under consideration (Klayman & Ha, 1987), pervades formal and informal hypothesis testing (Lakatos, 1970; Mynatt, Doherty, & Tweney, 1977; Snyder & Swann, 1978). Even people who have no stake in a particular claim tend to approach its evaluation with a bias toward verification. The bias may be even more pronounced, however, when people consider a proposition they wish, for whatever reason, to accept as valid. In contrast, people motivated to reject a hypothesis may be oriented toward disconfirmation, searching for exceptions to the distasteful rule. This suggests that the skeptical mindset people adopt when evaluating unpalatable propositions may inoculate them from a number of reasoning errors that result from this broader confirmation bias. The Wason selection task offers a particularly informative test case (Wason, 1966, 1968).

In its original form, the Wason task consisted of four cards purported to have a letter on one side and a number on the other. Participants saw one side of each of the four cards, marked with, for example, "E," "K," "4," and "7," and considered a conditional rule such as "If a card has a vowel on one side, it has an even number on the other side." The task was to identify those cards—and only those cards—that need to be turned over to determine whether the rule is valid. The correct answer is to turn over "E" (which must reveal an even number if the rule is true) and "7" (which must not reveal a vowel). More generally, any rule that follows the form "if *p* then *q*" is tested by turning over *p* and *not-q*, the only two cards that can falsify the rule.

Perhaps the most notable feature of the Wason task is that the success rate in the typical study is only around 20% (Cosmides, 1989; Evans, 1982). The two most common errors are the unnecessary inclusion of the *q* card and failure to name the *not-q* card (Cox & Griggs, 1982). Three decades of tinkering with the task have yielded a host of modifications that boost the success rate, including increasing task familiarity and altering processing instructions (for review, see Evans, 1982). Providing familiar thematic content seems to improve performance to a modest extent (Evans, 1982; Johnson-Laird, Legrenzi, & Legrenzi, 1972; Wason & Shapiro, 1972; but see Cox & Griggs, 1982; Mankeltow & Evans, 1979), perhaps because of the availability in memory of instances that counter the task rule (Griggs & Cox, 1982). Others have shown that people perform better on the task when the content is of a particular nature, for example, when it evokes obligation or permission schemas (Cheng & Holyoak, 1985, 1989), references social contracts (Cosmides, 1989), or provides an opportunity to identify a cheater (Cosmides & Tooby, 1992; Gigerenzer & Hug, 1992). The literature on the relative impact of abstract reasoning schemas versus social contracts is mixed, however, and the contending theoretical positions remain controversial. What is clear is that performance is often improved when the Wason task is draped in meaningful content rather than presented as a rule about letters and numbers and that in its original, abstract form, the task remains stubbornly challenging even to well-educated, experienced reasoners.

Wason attributed people's systematic errors on the selection task to a general tendency toward verification; in other words, a confirmation bias (Wason & Johnson-Laird, 1972). In the context of the selection task, Wason hypothesized, confirmation bias leads people to turn over cards that could confirm the rule, when the correct response is to seek information that potentially disconfirms the rule. This interpretation finds support from studies in which participants were given explicit instructions to falsify the rule or to take the perspective of one who should aim to falsify it, such as when a police officer looks for "law violators" (Fiedler & Hertel, 1994; Griggs, 1984; Valentine, 1985; Yachanin, 1986; Yachanin & Tweney, 1982). Such instructions generally yield modest gains in performance. Similar improvements obtain when a falsification strategy is elicited indirectly by asking respondents to test rules they know or strongly suspect to be false based on prior learning tasks (Pollard & Evans, 1981, 1983). Fiedler and Hertel (1994) observed more dramatic results when they combined falsification

instructions with contextual information designed to elicit some suspicion about the task rule's validity. For example, the rule "If a person takes (the drug) Sanotren, he/she becomes immune to AIDS" was presented with background information suggesting the company that produces Sanotren may be disreputable, thus casting doubt on the drug's effectiveness. With combined falsification instructions and suspicious circumstances, participants correctly identified cards *p* and *not-q* about 35% of the time.

Task instructions and thematic content, then, are external cues that encourage respondents to consider the implications of the typically neglected *not-q* card (Fiedler & Hertel, 1994). We propose that internal cues can achieve the same end. Specifically, motivations elicited by a task rule may influence subsequent reasoning about that rule. Note that in this context "motivation" refers to a desire to reach a specific conclusion rather than a desire to perform well. In fact, an individual's general motivation to provide the correct answer seems to have little bearing on his or her ability to do so. Incentives such as offering money (or threatening to take some away) tend to create highly motivated but largely unsuccessful participants (Wason & Johnson-Laird, 1972). But where general motivation fails, specific motivation to reject the aversive implications of a task rule may guide people toward a successful reasoning strategy.

We propose that the fundamentally different orientations described by the "Can I/Must I" model should give rise to predictable patterns of response on a version of the Wason task framed to make the task rule palatable or unpalatable to respondents. To the degree that a person wishes to reject a rule, he or she is likely to look for exceptions to it. If the person can identify even one exception, the rule loses its claim to validity. A person who wishes to accept the rule or who is neutral toward it is less likely to search for invalidating information than a person who wishes to reject it. Therefore, we predicted that individuals faced with a subjectively unpalatable rule would be more likely than participants faced with a subjectively palatable or neutral rule to select *p* and (critically) *not-q*, the potentially invalidating cards. Because this is precisely the strategy that leads to a correct solution to the Wason task, we anticipated that people testing a subjectively unfavorable hypothesis would perform better than those who were neutral or favorable toward the hypothesis.

There is some evidence that processing goals may influence performance on the Wason task. Participants in one series of studies (George, 1991) were instructed to adopt a hypothetical goal such as purchasing a car and then to consider a relevant conditional rule that implied either a satisfactory outcome (in this example, "If one buys a K-make car, one is always satisfied") or an unsatisfactory outcome (" . . . one is always dissatisfied"). The four task cards were labeled "K-make car," "other make car," "satisfied buyer," and "unsatisfied buyer." Participants responded differently to task rules implying positive and negative outcomes. In one instance, 64% of participants facing an unsatisfactory outcome selected *not-q*, compared with 27% of those considering a satisfactory outcome.

One interpretation of these results is that conditional rules reflecting unsatisfactory outcomes were more successful at eliciting *not-q* selections than those reflecting satisfactory outcomes. Note, however, that the card labeled "satisfied buyer" corresponds to the card *q* for satisfactory outcome rules and *not-q* for unsatisfactory rules and is more frequently chosen than the "unsatisfied buyer" card regardless of the rule being tested. Each of the task rules and stimuli in these studies suffers from the same confound. Therefore, it remains unclear whether participants were seeking to disconfirm an unpleasant (but not a pleasant) rule or whether they were simply inclined to view a pleasant card (i.e., "satisfied buyer" rather than "unsatisfied buyer") regardless of its pertinence to the rule. A second concern is that the hypothetical and thus largely inconsequential goals imposed on participants in these studies may not be typical of the more compelling implicit goals that guide real-life motivated reasoners. It is unclear whether reasoning about a hypothetical car purchase, for example, can provide insight into how people evaluate information regarding vital interests such as their health, abilities, future happiness and success, and so on.

To obtain more unequivocal support for the impact of motivated reasoning on performance on the Wason task, we presented identical materials to all participants but varied whether the outcomes were likely to be viewed as satisfactory or unsatisfactory. The unsatisfactory outcomes, furthermore, were compellingly negative, corresponding to the likelihood of an early death (Study 1) or the accuracy of an odious, self-relevant stereotype (Study 2). We examined responses to these thematic Wason tasks for evidence that motivation to disprove the task rule elicits better performance.

STUDY 1: THE DENIAL OF DEATH

In this study, we induced motivational perspectives via the Emotional Lability Inventory (ELI) developed by Greenberg et al. (1993). Their measure consists of 24 questions about the strength and volatility of respondents' emotions. The questions are such that nearly everyone makes either predominantly high or predominantly low responses—on an 11-point scale—across the 24 items (e.g., "How often do you have strong emotional reactions?" and "To what extent do others see you as emotional?"). After completing the inventory, partici-

pants are told that people who circle mostly low numbers are low in emotional lability, whereas those who circle mostly high numbers are high in emotional lability. Thus, the structure of the inventory ensures that most participants are readily classified as high or low in emotional lability and, critically, that their classification "rings true" and is both unambiguous and highly salient to them.

In our study, after participants completed the ELI and received feedback about their own classification, some tested a rule implying that high emotional lability is associated with early death and others that low emotional lability may be deadly. Presumably, this information is threatening to those who are told that their own emotional lability level may be a cause of premature death (Greenberg et al., 1993). We predicted that participants testing the threatening hypothesis that early death is associated with their own emotional lability classification would select the Wason task cards with the potential to disconfirm this unhappy proposition more often than those testing the nonthreatening hypothesis that early death is associated with a different emotional lability classification than their own. Put simply, subjects motivated to reject the task rule should be more likely to select the correct cards.

Method

Participants were 44 undergraduate students at Cornell University. They received course credit for their participation and were randomly assigned to conditions.

Participants were run individually or in groups of two to six. The experimenter presented the study as an examination of how people evaluate scientific evidence. She asked participants to complete an ELI so that "we can statistically control for individual differences." In reality, the purpose was to provide participants with salient feedback about their own emotional lability classification. After they finished the ELI, the experimenter explained that "Most people are interested in where they fall. It's fairly straightforward: If you circled mostly low numbers, you're low in emotional lability, and if you circled more high numbers, you're high in emotional lability." Having thus drawn participants' attention to their own lability classification, the experimenter proceeded to read aloud as participants followed a page of written text. In one condition, participants read that scientists have reason to believe that there is a positive relationship between low emotional lability and early death; in a second condition, they read that the purported relationship was between high emotional lability and early death. The experimenter told participants that a group of scientists had recently conducted a study in which they obtained emotional lability ratings and age of death for each person in a sample of Benedictine monks and Orthodox rabbis. She explained that the scientists had classified each participant as having experienced either an early or a late death in comparison to the median age of death for all men in the population. Participants learned that the hypothesis the scientists wished to investigate was whether everyone in the study who was low (high) in emotional lability, without exception, experienced an early death.

Participants saw four cards representing four individuals randomly chosen from the study sample of monks and rabbis. One side of each card indicated whether the person was low or high in emotional lability; the other indicated whether the person experienced an early or late death. The experimenter displayed one side of each of the four cards, arranging them so that participants read the following: Subject A: Low Emotional Lability; Subject B: High Emotional Lability; Subject C: Early Death; Subject D: Late Death. The experimenter instructed participants to choose two cards they would need to turn over to assess the hypothesis that subjects in the "monks and rabbis" study who were low (high) in emotional lability experienced, without exception, an early death. Participants were given as much time as they needed to write their selections on a piece of paper. Finally, the experimenter conducted a thorough debriefing in which she stressed that no known or suspected relationship exists between longevity and emotionality as captured by the ELI.

Results

The midpoint of the scale for each question on the ELI was 6. Participants with average scores less than 6 were classified as scoring low on the ELI (M ELI response = 3.1, $n = 22$) and those with average scores greater than 6 were classified as scoring high (M ELI response = 8.5, $n = 22$). Of the 22 high lability participants, 9 had been randomly assigned to test the hypothesis that low emotional lability leads to early death and 13 tested the hypothesis that high emotional lability leads to early death. The 22 low lability participants were evenly split between conditions.

Our dependent measure was whether participants named the two cards that would adequately test the hypothesis under consideration (cards corresponding to "Subject A" and "Subject D" when the hypothesis was that low emotional lability results in early death or "Subject B" and "Subject D" when the hypothesis was that high emotional lability results in early death). We anticipated that participants who perceived a potential threat in the hypothesis (i.e., participants whose own emotional lability categorization was implicated in early death) would arrive at the correct solution more often than those who did not. As illustrated in Figure 1, this was indeed the case. Only 10% of participants who consid-

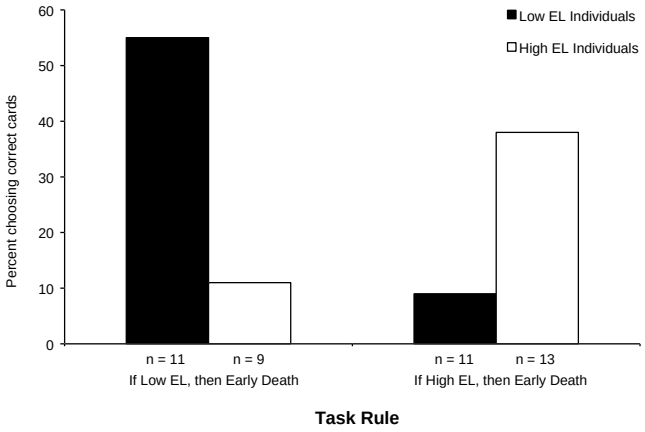

Figure 1 Percentage of participants in each condition selecting *p*, *not-q*, Study 1.

NOTE: Only participants whose own emotional lability (EL) corresponded to the EL referenced in the task rule were considered to have tested a subjectively threatening hypothesis.

ered a nonthreatening hypothesis correctly solved the problem, a success rate comparable to that generally observed in studies of the Wason selection task. In contrast, 46% of participants who tested a subjectively threatening hypothesis correctly named the cards corresponding to *p* and *not-q*, $\chi^2(1, N = 44) = 6.72$, $p < .01$.

Furthermore, of the 13 participants who solved the problem, all but 2 had been presented with task rules that, if true, would imply early death for people in their own emotional lability category. Note that neither test rule nor emotional lability level alone affected the success rate, $\chi^2(1, N = 44) = .52$ and .74, respectively, both *ns*. Finally, note that all but 7 participants selected as one of the two cards the equivalent of the *p* card ("Subject A" or "Subject B," depending on condition)—three who were high in emotional lability and were testing whether low emotional lability leads to early death, two who were low in emotional lability and were testing whether high emotional lability leads to early death, and one each in the two remaining conditions. The results, then, are clearly due to motivational influences on whether participants selected the typically overlooked *not-q* card ("Subject D").

Discussion

Some participants in this study considered a rule that, because of their own emotional lability status, did not threaten them personally. These participants experienced the usual difficulty in reasoning through the Wason task, demonstrating a failure to seek disconfirming evidence to evaluate the rule's validity. Other participants, in contrast, considered a rule that, if valid, implied a highly aversive outcome for themselves, namely, a high probability of experiencing an early death. Faced with a hypothesis they would rather reject, these participants were 4 times as likely to select both *p* and (critically) *not-q*, the cards that could potentially invalidate the task rule and hence are the correct responses to the Wason task.

It should be noted that in Study 1, participants were instructed to choose two cards to test the given rule. The original Wason selection task did not impose such a constraint: instructions were to choose "those cards and only those cards" necessary to determine the validity of the rule. We applied the two-card restriction in Study 1 to give respondents every opportunity to solve the notoriously difficult problem. Having obtained positive results on this presumably easier version, we conducted a similar experiment with the original, more challenging version in which respondents may choose any number of cards they feel are necessary. Study 2 is thus a conceptual replication of our first experiment, one that employs the instructions from the original Wason task.

STUDY 2: PROTECTING ONE'S OWN

In Study 2, we used ethnic and gender stereotypes to create propositions that participants would (or would not) be loath to accept. We asked participants to generate these stereotypes themselves and then adapted the content to fit the "If *p*, then *q*" format. For example, the stereotype that the Irish tend to like their drink was transformed into a Wason rule, "All Irish are alcoholics." Some participants tested a rule that, if true, implied something negative about themselves. We reasoned an Irish American testing the rule about alcoholics, for example, or a woman testing the rule that "All women are bad drivers" would be threatened by the stereotype in a way that, say, a Chinese American man would not. Still other participants tested rules that, far from being threatening, implied something positive about themselves: "All Asian Americans are smart" or "All women are empathetic."

As in Study 1, participants considered appropriate content-related cards corresponding to *p*, *not-p*, *q*, and *not-q*. Unlike participants in Study 1, however, they were encouraged to turn over as many cards as necessary (but no more than were necessary) to determine the validity of the task rule. We anticipated that individuals who were faced with a task rule embodying a negative stereotype about their own ethnic or gender group would be more motivated to falsify the rule—and therefore to select the correct cards *p* and *not-q*—than those who either considered a positive stereotype or considered a negative stereotype about an outgroup.

Method

Eighty Cornell undergraduates received course credit for their participation. Participants were run in pairs or groups of three. Seating was arranged so that no partici-

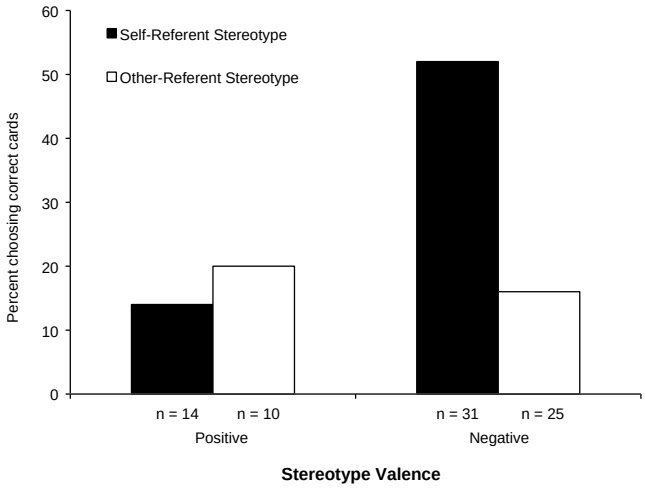

Figure 2 Percentage of participants in each condition selecting *p*, *not-q*, Study 2.

pant could read another's responses to the dependent measure.

Participants were first asked to think of a group to which they belonged, "for instance, on the basis of race or ethnicity," that had a strong stereotype associated with it. Participants were told that the stereotype need not (indeed, probably would not) be valid, only that it be widely known. In one condition, participants were asked to think of a negative stereotype pertaining to their group; in a smaller control condition, they were asked to think of a positive stereotype. After each person indicated that he or she had thought of a pertinent stereotype, the experimenter randomly selected one participant to state to the other participant(s) the group and associated stereotype he or she had in mind. The experimenter then wrote a rule on a blackboard in this form: "*All* (members of group) are (stereotypic feature)." For example, if a participant volunteered the stereotype that African Americans are musical, the experimenter wrote "All African Americans are musical."

Participants were then instructed to test the validity of the rule with respect to a set of four individuals. The experimenter drew four circles to represent these people and provided one piece of information for each: Person 1 was said to be a member of the group in question (corresponding to *p* in the classic Wason task), Person 2 was said *not* to be a member of the group in question (*not-p*), Person 3 was said to have the trait in question (*q*), and Person 4 was said *not* to have the trait in question (*not-q*). Thus, participants knew the group membership of two members of the set (but not whether the trait applied to them) and whether the trait applied to the two other members of the set (but not their group membership). This information was recorded on four index cards displayed next to the circle representations of each "Person"; the missing information, then, was supposedly contained on the back of each card. Participants were instructed to select "those people and only those people for whom you would need to discover the missing piece of information to determine whether it is true or false that all (p) are (q)." These instructions are the same as those typically provided in the Wason selection task (Wason & Johnson-Laird, 1972). Participants indicated on a piece of paper their selection or selections and whether they considered themselves a member of the group identified in the example. The experimenter collected the papers and debriefed participants taking great care to explain that no element of the study was to be taken as an implicit statement of the validity of the stereotype they had considered.

Results and Discussion

There are two independent variables of interest: the valence of the stereotype being tested and whether each respondent considered himself or herself to be a member of the stereotyped group. A third variable, whether the group stereotype was self- or other-generated, had no effect and will not be discussed further.

As in Study 1, we predicted that participants presented with a subjectively threatening hypothesis would be disproportionately successful on the Wason task. Our hypothesis was strongly supported by the data. As illustrated in Figure 2, individuals confronted with a negative stereotype about a group to which they belonged were more likely to select the correct cards (52%) than were participants who tested a rule about a negative stereotype that did not apply to themselves (16%) and those who tested rules about a positive stereotype that either did (14%) or did not (20%) apply to themselves. This pattern is statistically significant, as indicated by a chi-square test of independence between condition (negative self-relevant, negative other-relevant, positive self-relevant, and positive other-relevant) and response (correct or incorrect), $\chi^2(3, N = 80) = 11.35$, $p < 01$. As in Study 1, these results were due entirely to whether participants chose the *not-q* card, because all but six participants (as evenly spread across the four conditions as it is possible to be) chose the *p* card.

Although we included four conditions, the comparison of primary interest is between people who tested a negative, self-referent stereotype and all others. We coded people in the former group as having tested a subjectively threatening hypothesis ($n = 31$) and all others as having tested a subjectively nonthreatening hypothesis ($n = 49$). The difference in success rates was highly significant. In the absence of a subjectively threatening task rule, only 16% of participants arrived at the correct solu-

tion. In contrast, 52% of participants in the critical "threat" group did so, $\chi^2(1, N = 80) = 11.26$, $p < .001$.

We conducted a final analysis to rule out the possibility that it was their greater familiarity with the stereotype being tested (rather than its unpalatable quality) that was responsible for the superior performance by participants testing a personally threatening hypothesis. Members of a stereotyped group, certainly, are typically more familiar with all of the group's associated stereotypes. However, if familiarity alone was responsible for the superior performance observed in this study, participants testing a self-referent task rule should have outperformed others regardless of the rule's valence. The proposition that "all women, without exception, are nurturing" is no less familiar than the proposition that "all women, without exception, are bad drivers" and should have elicited equally successful performance according to this alternative interpretation. But this was not the case. Participants to whom the stereotypes referred were more likely than others to select the cards that could disconfirm a negative stereotype, $\chi^2(1, N = 56) = 7.65$, $p < .006$, but were *not* more likely than others to select the cards that could disconfirm a positive stereotype, $\chi^2(1, N = 24) < 1$. Thus, only the consideration of a test rule with negative connotations for the self elicited a higher rate of correct responses than other test rules, and indeed, a much higher success rate than is generally observed in Wason task studies.

GENERAL DISCUSSION

Consider two standards by which one could choose to evaluate a body of evidence. One is to accept the validity of a proposition if the data collected to assess it seem on the whole to be consistent with it. Another, stricter standard is to require the data to be not merely consistent with a proposition but overwhelmingly supportive of it. Considered in isolation, each seems reasonable. Indeed, each is formally endorsed by the American legal system, which sees no contradiction in declaring O. J. Simpson innocent in one trial but guilty in a second. Presumably, the famously undersized glove was as much a misfit during the second proceeding as the first. The disparate verdicts were the result of a relatively high standard of evidence in a criminal trial, where guilt must be established beyond reasonable doubt, and a lower standard in a civil trial, where guilt must be only probable (often interpreted as supported by at least 51% of the evidence; Wrightsman, Nietzel, & Fortune, 1998). In everyday reasoning, people appear to employ similarly divergent criteria to assess propositions they are motivated either to reject or accept. In so doing, people are often able to believe what they wish to believe about the world around them and, at the same time, remain convinced of their own objectivity.

Faced with an unpalatable hypothesis, people tend to ask, in essence, "Must I believe this?" rather than "Can I believe this?" (Gilovich, 1991). The implicit question guides the ways in which they access and evaluate potentially informative data. For example, reasoners motivated to reject a proposition tend to examine all potentially informative data rather than just a portion of it (Dawson et al., 2002) and to contemplate it more thoroughly (Ditto et al., 1998; Ditto & Lopez, 1992). As we demonstrated with the Wason selection task, they also tend to approach friendly and hostile hypotheses with qualitatively different goals. Favorable propositions appear to elicit a search for confirmation; unfavorable propositions elicit a search for disconfirmation. In the context of the Wason selection task, disconfirmation is precisely the approach that leads to the correct response. Thus, individuals in our studies who sought to reject the task rule were more likely than others to solve the problem.

The behavior of participants in our studies was not a function of the task rule alone but of the subjective interpretation of the rule. Specifically, some participants confronted a rule that, although negative in tone, was not directed at them, whereas others considered a rule that, if true, would imply their own heightened chance of experiencing an early death (Study 1) or the confirmation of a personally offensive stereotype (Study 2). Thus, we were able to distinguish between the effects of propositions that were negative in a general sense and those that specifically targeted reasoners' personal well-being or worldview. We found no evidence that negative information alone elicited a disconfirmatory approach to the Wason task. In Study 1, people exposed to a negative but personally irrelevant hypothesis (essentially "the other group will, without exception, experience early death") showed a rate of success similar to that observed in other studies using a content-free Wason task. In Study 2, a negative test rule alone inspired no more correct responses than a positive test rule. Only when participants considered negative test rules directed at themselves were they drawn to the potentially disconfirming case and, by extension, to the correct answer to the Wason selection task.

Although the skepticism that motivated participants brought to the task led them to the correct answer in our studies, we do not suggest that it would necessarily transfer to superior performance on subsequent, more abstract versions of the Wason problem. The phenomenology with which they approach the task is one dominated by thoughts such as "this can't be right" rather than "the right way to approach this is to look for disconfirmations." Thus, selecting the cards that could invalidate a distasteful hypothesis does not necessarily indicate more skillful logical reasoning or greater general mastery of the problem. What it does demonstrate is

that skeptics were led by their motivations to look where they needed to look to evaluate a claim, even if they were not aware of an explicit strategy for doing so.

In general, a healthy skepticism entails a proclivity to thoroughly evaluate data and to think critically about their implications. It is this feature that has led the many advocates of "critical thinking" to argue that a skeptical mindset promotes sound reasoning (Gilovich, 1991; Sagan, 1996; Schick & Vaughn, 1999; Shermer, 1997; Stanovich, 2001). A reasoner who adopts a skeptical mindset may be better able to spot illusory correlations (Schaller, 1991, 1992), to rely on proportions rather than on raw numbers to assess contingency between variables (Dawson et al., 2002), and to resist generalizing from small or highly variable samples (Doosje, Spears, & Koomen, 1995). In contrast, the motivation to find support for preferred beliefs may render people particularly susceptible to errors. Such a motivation often leads a person to overlook even glaring faults in the data, because it is difficult to find what is not sought. What is needed, then, are ways to encourage a general or decontextualized skepticism that can facilitate sound reasoning in situations in which a strong motivation to disconfirm a particular proposition is absent.

But does a skeptical mindset make people unwilling to accept unpalatable propositions however strong their evidentiary bases? Typically not. Although instances of such radical rejectionism can be found (the Flat Earth Society retains an active membership to this day and many cults contain absurd foundational beliefs), most often even those with a strong motivation to reject a hypothesis tend to accept it in the face of strongly supportive evidence. In other studies, for example, we have shown that those who are disinclined to accept a given proposition will, after careful scrutiny of the pertinent evidence, come to accept it if the evidence is in fact solidly supportive (Dawson et al., 2002; Gilovich et al., 2002). Sometimes the answer to "Must I believe?" is a resounding "yes," and motivated reasoners seem willing to accept that answer when the evidence so indicates. More generally, individuals motivated to reject a proposition have a number of strategies at their disposal, from categorical rejection ("nothing could convince me otherwise") to careful scrutiny of available evidence with an eye toward spotting flaws ("but their survey only asked people who . . . "). It is those strategies at the latter end of the continuum, however, that best serve the twin goals of maintaining one's cherished beliefs and preserving one's conviction that one is unbiased (Kunda, 1987; Pronin, Ross, & Gilovich, 2002). It is thus no great surprise that the latter strategies are more typically pursued, or at least more typically pursued first (Kunda, 1987). In the end, individuals who pursue such strategies can end up believing what they want to believe, and remain convinced that they are fair-minded evaluators able to let the evidentiary chips fall where they may. Along the way, as we have shown here, they can also successfully solve challenging inferential problems that stump less skeptical inquirers.

REFERENCES

Bacon, F. (1955). Novum organum. In R. M. Hutchins (Ed.), *Great books of the Western world* (Vol. 30, pp. 105-195). Chicago: Encyclopedia Britannica. (Original work published 1620)

Cheng, P. W., & Holyoak, K. (1985). Pragmatic reasoning schemas. *Cognitive Psychology, 17*, 391-416.

Cheng, P. W., & Holyoak, K. (1989). On the natural selection of reasoning theories. *Cognition, 33*, 285-313.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition, 31*, 187-276.

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

Cox, J. R., & Griggs, R. A. (1982). The effects of experience on performance in Wason's selection task. *Memory & Cognition, 105*, 496-502.

Dawson, E., Gilovich, T., & Regan, D.T. (2002). *Motivated reasoning and susceptibility to the "Cell A" bias*. Manuscript submitted for publication.

Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology, 63*, 568-584.

Ditto, P. H., Scepansky, J. A., Munro, G. D., Apanovich, A. M., & Lockhart, L. K. (1998). Motivated sensitivity to preference-inconsistent information. *Journal of Personality and Social Psychology, 75*, 53-69.

Doosje, B., Spears, R., & Koomen, W. (1995). When bad isn't all bad: Strategic use of sample information in generalization and stereotyping. *Journal of Personality and Social Psychology, 69*, 642-655.

Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology, 57*, 1082-1090.

Evans, J. St. B. T. (1982). *The psychology of deductive reasoning*. London: Routledge Kegan Paul.

Fiedler, K., & Hertel, G. (1994). Content-related schemata versus verbal-framing effects in deductive reasoning. *Social Cognition, 12*, 129-147.

George, C. (1991). Facilitation in the Wason selection task with a consequent referring to an unsatisfactory outcome. *British Journal of Psychology, 82*, 463-472.

Gigerenzer, G., & Hug, K. (1992). Reasoning about social contracts: Cheating and perspective change. *Cognition, 43*, 127-171.

Gilovich, T. (1991). *How we know what isn't so: The fallibility of human reason in everyday life*. New York: Free Press.

Gilovich, T., Dawson, E., & Regan, D. T. (2002). *Motivated skepticism, motivated credulity: The "Can I? Must I?" distinction in preference-driven reasoning*. Manuscript in preparation.

Greenberg, J., Pyszczynski, T., Solomon, S., Pinel, E., Jordan, K., & Simon, L. (1993). The effects of self-esteem on vulnerability-denying defensive distortions: Further evidence that self-esteem serves an anxiety-buffering function. *Journal of Experimental Social Psychology, 29*, 229-251.

Griggs, R. A. (1984). Memory cueing and instructional effects on Wason's selection task. *Current Psychological Research and Reviews, 3*, 3-10.

Griggs, R. A., & Cox, J. R. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology, 73*, 407-420.

Johnson-Laird, P. N., Legrenzi, P., & Legrenzi, M. (1972). Reasoning and a sense of reality. *British Journal of Psychology, 63*, 395-400.

Klayman, J., & Ha, Y. W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94*, 211-228.

Kunda, Z. (1987). Motivated inference: Self-serving generation and evaluation of causal theories. *Journal of Personality and Social Psychology, 53*, 636-647.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*, 636-647.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91-105). London: Cambridge University Press.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology, 37*, 2098-2109.

Mankeltow, K. I., & Evans, J. St. B. T. (1979). Facilitation of reasoning by realism: Effect or non-effect? *British Journal of Psychology, 70*, 477-488.

Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology, 29*, 85-95.

Pollard, P., & Evans, J. St. B. T. (1981). The effects of prior beliefs in reasoning: An associational interpretation. *British Journal of Psychology, 72*, 73-81.

Pollard, P., & Evans, J. St. B. T. (1983). The effect of experimentally contrived experience on reasoning performance. *Psychological Research, 45*, 287-301.

Pronin, E., Ross, L., & Gilovich, T. (2002). *The bias blindspot: Detecting bias in others, but not the self.* Manuscript in progress.

Pyszczynski, T., & Greenberg, J. (1987). Toward an integration of cognitive and motivational perspectives on social inference: A biased hypothesis-testing model. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 20, pp. 297-340). New York: Academic Press.

Sagan, C. (1996). *The demon haunted world: Science as a candle in the dark.* New York: Random House.

Schaller, M. (1991). Social categorization and the formation of group stereotypes: Further evidence for biased information processing in the perception of group-behavior correlations. *European Journal of Social Psychology, 21*, 25-35.

Schaller, M. (1992). In-group favoritism and statistical reasoning in social inference: Implications for formation and maintenance of group stereotypes. *Journal of Personality and Social Psychology, 63*, 61-74.

Schick, T., & Vaughn, L. (1999). *How to think about weird things: Critical thinking for a new age.* Mountain View, CA: Mayfield.

Shermer, M. (1997). *Why people believe weird things: Pseudoscience, superstition, and other confusions of our time.* New York: Freeman.

Snyder, M., & Swann, W. B. (1978). Hypothesis-testing processes in social interaction. *Journal of Personality and Social Psychology, 36*, 1202-1212.

Stanovich, K. E. (2001). *How to think straight about psychology.* Boston: Allyn & Bacon.

Valentine, E. R. (1985). The effect of instructions on performance in the Wason selection task. *Current Psychological Research and Reviews, 4*, 214-223.

Wason, P. C. (1966). Reasoning. In B. Foss (Ed.), *New horizons in psychology.* London: Penguin.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology, 20*, 271-281.

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content.* Cambridge, MA: Harvard University Press.

Wason, P. C., & Shapiro, D. (1972). Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology, 23*, 63-71.

Wrightsman, L. S., Nietzel, T., & Fortune, W. H. (1998). *Psychology and the legal system.* Pacific Grove, CA: Brooks/Cole.

Yachanin, S. A. (1986). Facilitation in Wason's selection task: Content and instructions. *Current Psychological Research and Reviews, 5*, 20-29.

Yachanin, S. A., & Tweney, R. D. (1982). The effect of semantic content in cognition strategies in the four card selection task. *Bulletin of the Psychonomic Society, 19*, 87-90.

Received August 14, 2001
Revision accepted March 14, 2002