

Running head: MOTIVATED ASSESSMENT OF CONTINGENCY

Motivated Reasoning and Susceptibility to the “Cell A” Bias

Erica Dawson

Yale School of Management

135 Prospect Street

New Haven, CT 06520

Thomas Gilovich and Dennis T. Regan

Department of Psychology, Uris Hall

Cornell University

Ithaca, NY 14853

Abstract

Intuitive strategies for assessing covariance, or the degree to which two variables are related, typically assign disproportionate weight to the number of instances in which both variables occur. In 2 studies, the *cell A bias* (referring to the “present-present” cell in a 2 x 2 contingency table) was attenuated by presenting some participants with a proposed bivariate relationship inconsistent with their desired beliefs about academic performance and program housing (Study 1) or the collegiate Greek system (Study 2). Participants evaluating an unpalatable proposition were more likely than others to (a) endorse a relatively stringent criterion for acceptance of the hypothesized relationship; (b) consider the entries in all 4 cells of a 2 x 2 contingency table, thus avoiding the cell A bias, and, as a result, (c) render more accurate judgments of contingency. Results are consistent with a proposed “Can I? Must I?” distinction in motivated reasoning reflecting the implicit questions that guide the evaluation of subjectively favorable and unfavorable propositions.

Motivated Reasoning and the Assessment of Contingency

I know what I believe, and I believe what I believe is right.

George W. Bush, of his position on free trade agreements with developing nations (in Sanger, 2001).

Confidence in the validity of one's beliefs is not an exclusively presidential trait. People from all walks of life are disposed to believe that what they believe is right. In one sense, this can only be true; by definition, people cannot believe things they know to be false. But a complex world rarely yields clear truths and falsehoods. There are no definitive data to clarify once and for all, for example, what the best policy in the Middle East would be, whether acupuncture is an effective remedy for arthritis, whether baseball managers should play for a tie in the late innings at home but play for the win on the road, and so on. Instead, people decide which claims should be accepted and which rejected by sorting through a considerable volume of potentially relevant information, most of it tainted by assorted limitations, anomalies, and contradictions.

People do not approach this task dispassionately. Of particular interest here, their assessments of novel information tend to reflect their expectations and desires; rather than believing what they see, they sometimes see what they believe (e.g., Gilovich, 1991; Kunda, 1990; Nisbett & Ross, 1980). When individuals occupy opposite sides of an ideological fence, it is likely that their interpretations of pertinent evidence will diverge as well. In one exemplary study, opponents and proponents of capital punishment independently extracted support for their

own view from the same (objectively inconclusive) set of evidence (Lord, Ross, & Lepper, 1979). In a parallel demonstration, partisans on both sides of the Middle East conflict were each disposed to claim media bias *against* their own cause, so that the same newscast was accused simultaneously of being pro-Israeli and pro-Arab (Vallone, Ross, & Lepper, 1985).

Of course, self-serving views lose their effectiveness if clearly unjustified, and so they must be tempered by at least a veneer of objectivity. People are thus particularly keen to establish their impartiality when evaluating themselves or issues in which they have a personal stake (e.g., Batson, Thompson, Seufferling, Whitney, & Strongman, 1999; Kunda, 1990; Pyszczynski & Greenberg, 1987), and their ability to believe what they want to believe is generally constrained by their ability to justify their position to a (real or imagined) objective observer (Kunda, 1990). The question, then, is how people come to believe in the validity of their preferred opinions, while also believing them to be derived impartially.

One strategy is to apply inferential rules most likely to yield desired outcomes. Converging accounts of motivated reasoning describe just such a process (e.g., Kruglanski & Meinholdt, 1990; Kruglanski & Webster, 1996; Kunda, 1987, 1990; Pyszczynski & Greenberg, 1987). We have suggested elsewhere, for example, that people employ different standards of evidence when evaluating favored and disfavored propositions (Dawson, Gilovich, & Regan, 2002; Gilovich, 1991). When faced with an unpalatable proposition, people appear to implicitly ask, "*Must* I believe this?," a relatively stringent standard that invokes a critical examination of pertinent evidence. Accordingly, people tend to challenge the accuracy, reliability, strength, and source of threatening information and to search for and generate exceptions and counterarguments (Dawson et al., 2002; Ditto et al., 1998; Ditto & Lopez, 1992; Edwards & Smith, 1996; Holton & Pyszczynski, 1989; Kunda, 1990; Lord et al., 1979). This

relatively thorough processing of threatening information is likely to expose any inconsistencies, ambiguities, and other flaws that may taint a body of evidence. Thus, people who are motivated to reject an unpalatable proposition are likely to find legitimate grounds for doing so.

In contrast, when faced with a palatable proposition, people appear to implicitly ask “*Can I believe this?*,” a more permissive standard that elicits reasoning biased toward confirmation. For example, when evaluating a claim compatible with a desired belief, people tend to search their memory and external sources of information for supportive evidence (Dawson et al., 2002; Kruglanski, 1980; Kunda, 1987; Pyszczynski & Greenberg, 1987), and to truncate the search once any is found (Ditto & Lopez, 1992; Ditto et al. 1998; Kruglanski & Webster, 1996). Supportive evidence thus tends to be evaluated uncritically and weighted too heavily (Ditto & Lopez, 1992; Ditto et al., 1998; Edwards & Smith, 1996; Gilovich, Dawson, & Regan, 2002; Lord et al., 1979). In short, people tend to think lightly about congenial evidence and to keep their distance from its potential deficiencies. As a result, those motivated to do so can generally adduce support for all but the most outlandish propositions.

The psychological processes invoked by the question, “*Can I believe this?*” align with the well-documented confirmation bias in reasoning. In numerous contexts, people tend to adopt a positive test strategy to evaluate hypotheses; that is, they tend to seek evidence that a hypothesis is valid rather than evidence that it is invalid (Klayman & Ha, 1987). This tendency pervades both formal hypothesis testing, when people attempt to discover or test rules of logic or categorization (Klayman & Ha, 1989; Tweeney, Doherty, Worner, Pliske, & Mynatt, 1980; Wason & Johnson-Laird, 1972), and informal hypothesis testing, when people consider propositions about social agents (Crocker, 1981; Snyder & Cantor, 1979; Snyder & Swann, 1978). In many instances, a positive test strategy is useful, allowing individuals to discriminate

between the target hypothesis and a likely alternative (Klayman & Ha, 1987; Sanbonmatsu, Posavac, Kardes, & Mantel, 1998; Trope & Bassok, 1982, 1983). Nevertheless, it sometimes leads people to misjudge the diagnosticity of the evidence consulted, exciting more confidence in its validity than a more complete examination would warrant.

People may be particularly susceptible to confirmation bias when assessing a preferred conclusion. However, skeptics--those motivated to ask "Must I believe this?" rather than "Can I believe this?"—may be more apt to avoid this pitfall as a consequence of searching for disconfirming evidence. Skeptics may thus have an advantage in reasoning about problems for which a positive test strategy is not optimal. The assessment of covariance, a reasoning task notable for both its ubiquity and apparent difficulty, is just such a case.

In the typical covariation study, participants view either a series or a summary of instances in which each of two variables is either present or absent. They then estimate the degree to which the variables are related. For example, when judging whether symptom X predicts disease A, there are four possible cases: symptom and disease both present; symptom and disease both absent; symptom present, disease absent; and symptom absent, disease present. The number of cases in a sample purported to fall into each of these categories may be displayed in a 2 x 2 table, with cells labeled A, B, C, and D, as in Figure 1. Participants might then be instructed to assess whether, or to what degree, symptom X is related to (or predicts, correlates with, or covaries with) disease A.

There are several ways to measure the relationship between two dichotomous variables, with the calculation of X^2 or the ϕ coefficient among the most common. While these calculations are familiar to statisticians, they are unlikely to be in the repertoire of naïve reasoners. Instead, people rely on intuitive strategies, which tend to reflect an underlying positive test strategy and

thus give rise to systematic error (McKenzie, 1994; Nisbett & Ross, 1980; Schustack & Sternberg, 1981; Smedslund, 1963). People reliably look first and foremost to cell A, the present-present cell consistent with the implicit hypothesis that the variables are related and, thus, occur together. A common strategy is simply to compare the number in cell A to that in cell C or, less frequently, cell B (Arkes & Harkness, 1983; Beyth-Marom, 1982; Ward & Jenkins, 1965; Wasserman, Dorner, & Kao, 1990). Even if people recognize that a judgment must be based on the entries in all four cells of the table, they often aggregate the data incorrectly, yielding conclusions that reflect a disproportionate weighting of the information in cell A, or the “cell A bias” (Beyth-Marom, 1982).

The cell A bias can be dampened by experimental variations on the traditional assessment task. People tend to perform better, for example, with data presented in summary, rather than serial, form (Smedslund, 1963) and with variables described in terms of degree of occurrence rather than absolute presence or absence (Beyth-Marom, 1982). Predictably, they perform better when explicitly instructed to compare conditional probabilities (Shaklee & Tucker, 1980) or to attempt falsification by examining all four cells (Crocker, 1982). For the most part, attempts to enhance performance have focused on the way the task is framed--instructions, presentation mode, and so on--rather than on task content, or the nature of the correlation to be assessed (although see Mynatt, Doherty, & Dragan, 1993). In a departure from this tradition, we propose that the nature of the hypothesized bivariate relationship itself, and in particular the motivational significance for the people evaluating it, may lead people to spontaneously adopt an effective strategy of falsification. Specifically, we suggest that the motivation to disprove an unpalatable proposition can diminish the cell A bias.

If people who wish to accept a proposition apply a permissive criterion, they may be satisfied that the criterion has been met upon encountering a large number in cell A relative to the numbers in one or two other cells. After this rather perfunctory consideration of the data, they may conclude, sometimes incorrectly, that the agreeable proposal has been validated by evidence. People who wish to reject the proposal may not be so easily deceived. Guided by a more stringent criterion, these individuals, we predict, should examine the numbers in all four cells and reason critically about the data in an attempt to find grounds for dismissing them. If people motivated to reject an unpalatable proposition are more likely than others to attend to the complete data, they should also be more likely to accurately assess covariation, precisely because the latter depends upon the former.

To test these predictions, we presented participants in two studies with a subjectively favorable or unfavorable proposition about the relationship between two variables. We first asked them to specify which of two criteria—corresponding to the implicit questions "Can I believe this?" and "Must I believe this?"—is the most appropriate for assessing whether the two variables are, in fact, related. We then provided data purported to have been collected to evaluate the claim under consideration. The data were presented in four cells of a 2 (variable 1: present vs. absent) x 2 (variable 2: present vs. absent) table, which participants accessed as they wished to form a judgment of association.

Our purposes were threefold. First, we wished to determine whether the "Can I? Must I?" model describes literally the differential criteria by which people evaluate palatable and unpalatable propositions. Next, and as a result, we were interested in whether those motivated to discredit an unpalatable claim—those who ask, "Must I believe this?"—would tend to examine the complete data and, thus, fare better than others on a covariation assessment task. Finally, we

explored the boundaries of skepticism to determine whether even those motivated to reject a claim would nevertheless accept it in the face of compelling evidence.

Study 1: Program Housing

This study was designed around a then-contentious issue among Cornell students: whether first-year undergraduate students should be allowed to participate in program housing, particularly housing organized by race or ethnicity. At the time of the study, the university was considering barring first-year students from living in program housing, a move hotly debated by students. Some argued that program housing isolates first-year minority students from the larger campus community and thus should be restricted to more advanced students. Others claimed that program housing helps minority students connect with others who share their interests and concerns, thus integrating them into the community in a way that would not be possible if they were denied this housing option.

We recruited students who strongly supported or opposed freshman access to program housing and asked them to evaluate a hypothetical relationship between program housing and academic performance. The effect of theme housing on students' grades was not central to the Cornell debate, which instead revolved around the social and moral aspects of voluntary segregation. We thus assumed that participants would not have substantial prior knowledge about any specific relationship between housing and academics. At the same time, we felt that most students would concede the importance of any impact on grades. We told participants that researchers had collected data in an effort to determine (in one condition) whether program housing hurts minority students' first-year grades or (in another condition) whether program housing helps minority students' grades.

We first recorded participants' articulated reasoning strategy—that is, the criterion they believed appropriate to evaluate the proposition in question. We then recorded the amount of supportive evidence they claimed to require to be convinced of the proposition's validity, the amount of information in a 2 x 2 table they examined to make an assessment of contingency, and their judgment of whether the data supported the contention that program housing is associated with improved [impaired] academic performance. We predicted that relative to people evaluating a proposition compatible with a desired belief, those evaluating a proposition incompatible with a desired belief would: (a) advocate a more stringent reasoning criterion; (b) insist on more supportive evidence for acceptance; (c) access more data from the 2 x 2 summary table; and (d) make more accurate assessments of the degree of contingency between the two variables in the 2 x 2 table.

Method

Participants were 33 Cornell undergraduates who responded to announcements in psychology classes recruiting participants with strong opinions about program housing, either pro or con. Because statistical training might influence reasoning strategy, we omitted the data of two participants who indicated that they had taken a college-level statistics course. One person failed to complete the study because of time constraints, leaving 11 male and 19 female participants.

Participants were run individually. Upon entering the lab, they provided demographic information and a list of their coursework in statistics. They recorded their general stance toward program housing by circling either "Pro; In favor of program housing" (n = 19) or "Con; Opposed to program housing" (n = 11) on a piece of paper. The experimenter could not see participants' responses and remained blind to their attitude throughout the session.

Next, participants followed a typed page as the experimenter read aloud some of the contentions surrounding students' access to program housing (e.g., some people believe program housing integrates minority students into the campus community, whereas others believe it isolates them). They were told that researchers had recently conducted a study of the relationship between program housing and minority students' first-year grade point averages (GPAs), and that their task would be to draw conclusions from the data. Specifically, participants randomly assigned to the "helps grades" condition were asked to determine whether access to program housing helps minority students' first-year grades, whereas those assigned to the "hurts grades" condition were asked to determine whether it hurts minority students' first-year grades.

Before disclosing any data, the experimenter asked participants to consider two criteria a person could use to evaluate the question at hand: "People should conclude that access to program housing helps [hurts] minority students' first-year grades only if the data compel such a conclusion--only if the data make it clear that no other conclusion is warranted" ("compel" criterion) and "People should conclude that access to program housing helps [hurts] minority students' first-year grades if the data permit such a conclusion--if there are data that support that conclusion" ("permit" criterion). The order of presentation was counterbalanced within each condition. Participants placed a check next to the criterion they believed to be most appropriate. They were then instructed to imagine they had access to all studies that have been conducted on the topic of program housing and academic performance. On a worksheet, they completed the statement, "People should accept that program housing helps [hurts] minority students' first-year grades if at least _____ percent of all studies conducted on the topic support this proposition."

The experimenter then shifted participants' attention to one specific study they were to evaluate. Although presented as authentic, the "study" was manufactured for our purposes. The experimenter said that a group of researchers had randomly selected a large number of colleges and universities from across the United States. For each, she explained, the researchers determined (a) whether the institution allowed first-year students access to program housing, and (b) whether the GPA of first-year minority students at the institution was above or below the national average for all first-year minority students. The purported results were summarized in a 2 x 2 table (Appendix A). In both conditions, the first column was labeled "University allows access to program housing" and the second "University does not allow access to program housing." In the "helps grades" condition, the first row was labeled "Minority students' GPA above average" and the second "Minority students' GPA below average." In the "hurts grades" condition, the row labels were reversed. In all conditions, the numbers in each cell were 138 (upper left cell, or "cell A"), 69 (upper right cell, or "cell B"), 46 (lower left cell, "cell C"), and 23 (lower right cell, "cell D"). We expected that the large number in cell A, when compared to the numbers in cell B or cell C or both, would prompt judgments that the variables are related. In fact, based on these "data," they are not, as $(A / A+C) = (B / B+D) = .75$.

In the initial presentation of the 2 x 2 table, the number in each cell was concealed with an index card. Participants were instructed to select as much information as they needed to test the hypothesis that "access to program housing helps [hurts] first-year minority students' grades." Participants chose a cell they wished to view, and the experimenter removed the index card to reveal the number. Participants were asked to state their thoughts after seeing the number (to encourage some degree of deliberation) and to choose another cell to view only if necessary for them to make a judgment of whether the two variables were related. They were instructed to

view as many or as few cells as they wished, but to stop as soon as they felt they had enough information, and then to state their conclusion. With the participants' permission, this portion of the session was audiotaped (all agreed).

After participants indicated they had seen as many cells as they needed, the experimenter classified and recorded each participant's judgment of contingency as (a) positive relationship/data support the proposition; (b) negative relationship/data support the opposite proposition; (c) no relationship/data inconclusive; or (d) don't know/can't make a judgment.¹ The audiotaped responses were later transcribed and coded independently by a research assistant with no knowledge of our hypotheses or participants' condition (i.e., "helps grades" or "hurts grades"). The two coders agreed on 87 percent of the responses, with discrepancies resolved by a third rater unaware of the hypothesis and participant condition. No participants reported being unable to make a judgment, effectively eliminating the final response category.

After participants had rendered their initial assessment of the relationship in question, the experimenter exposed the entries in any cells they had allowed to remain covered. She asked participants who had initially selected fewer than four cells to examine the complete data and to indicate whether (and in what way) their previous conclusion had changed. Participants' responses were recorded and later coded as above. The coders agreed on all responses. Finally, as a confirmation of reported attitude strength and valence, participants indicated their level of agreement with four statements pertaining to program housing (e.g., "I think ethnic and lifestyle program houses are a valuable component of the university learning experience" and "Program houses essentially keep people from having to interact with others who are different from them, and as such, are detrimental to people who choose to live in them"). Responses were given on a 7-point Likert scale, reverse scored where appropriate so that 1 = strongly oppose and 7 =

strongly support. Participants were then debriefed about the true purpose of the study and the use of deception.

Results

We were interested in five dependent measures: (a) which reasoning criterion participants endorsed; (b) the proportion of supportive evidence they thought necessary to accept the proposition in question; (c) the number of cells of the 2 x 2 table participants chose to view before making a judgment of contingency; (d) their judgment of contingency based on the data they chose to view; and (e) for participants who selected fewer than four cells, their judgment of contingency after being exposed to all the data.

Attitude stance. Participants were identified as supporters or opponents of program housing by whether they circled “Pro” or “Con” on the initial 1-item attitude scale. To ensure that participants had accurately reported their attitudes, and as a check of attitude strength, we analyzed their mean responses to the four-item attitude questionnaire. There was a significant difference between participants who identified themselves as supporting program housing ($M = 5.87$, $SD = .84$) and opposing it ($M = 2.25$, $SD = .71$), $t(28) = 12.07$; $p < .0001$. Furthermore, each participant’s mean response to these four items was consistent with his or her reported attitude on the initial 1-item scale. Confident that their opinions were unambiguous, we determined that of the 30 participants, 17 (11 supporters and 6 opponents of program housing) had been randomly assigned to the “helps grades” condition, and 13 (8 supporters and 5 opponents) to the “hurts grades” condition.

Reasoning criterion. Recall that before being exposed to any data, participants were asked which of two criteria they believed was most appropriate to establish whether the purported relationship existed: that one should conclude program housing helps [hurts] grades if

the evidence permits such a conclusion, or that one should conclude program housing helps [hurts] grades only if the evidence compels such a conclusion. We predicted that participants motivated to accept the proposition would tend to choose the “permit” criterion (a paraphrase of “Can I believe this?”) whereas those motivated to reject it would tend to choose the “compel” criterion (a paraphrase of “Must I believe this?”). To test this prediction, we collapsed across condition and attitude stance to form two groups comprising participants who had considered a subjectively favorable proposition (i.e., supporters in the “helps grades” condition and opponents in the “hurts grades” condition; $N = 16$) and those who had considered a subjectively unfavorable proposition (i.e., supporters in the “hurts grades” and opponents in the “helps grades” condition; $N = 14$). Sixty-three percent of participants considering a subjectively favorable proposition selected the “permit” criterion, whereas only 14 percent of those considering a subjectively unfavorable proposition did so, $X^2(1) = 7.23, p < .01$. In short, participants tended to apply a relatively permissive standard to palatable propositions, but a relatively strict standard to unpalatable propositions (Figure 2).

Percent of studies required for accepting the proposition. Participants were also asked what percentage of all studies on the topic would have to support the proposition that program housing helps [hurts] grades in order for them to be convinced of its validity. Elsewhere, we have demonstrated that opponents of an issue tend to require a greater amount of supportive evidence than supporters do (Gilovich et al., 2002). We thus anticipated that people motivated to reject the proposal about program housing and grades would state a higher percentage than those motivated to accept it. This was not the case, as all participants claimed about 75 percent to be a reasonable figure regardless of whether the proposal was subjectively favorable ($M = 72.94, SD = 16.99$) or unfavorable ($M = 77.14, SD = 8.25$), $t < 1$.

Number of cells examined. Ninety-three percent of participants chose cell A as their first or second selection. We anticipated that most would perceive support for the proposition that program housing helps [hurts] grades after comparing the relatively large number in cell A (138) with the smaller numbers in cell B (69) or cell C (46), or both. We hypothesized that participants considering a favorable proposition would truncate their search for evidence at fewer than four cells, having satisfied the implicit criterion, "Accept if evidence permits." In contrast, we anticipated that participants testing an unfavorable proposition--those guided by the more stringent criterion "Accept only if evidence compels" -- would be reluctant to concede the validity of the unpalatable proposition before examining all four cells. This was the case. As shown in Figure 3, a higher proportion of participants faced with an unpalatable proposition (64 percent) than those faced with a favorable proposition (19 percent) opted to examine the full data before attempting a judgment of contingency, $X^2(1) = 6.45, p < .02$.

Participants' initial conclusions. Participants were initially asked to judge whether there was an association between the two variables based on whatever cells they had chosen to view. Table 1 provides the percentage of participants offering each conclusion. We coded judgments as either "correct" (a judgment of "no relationship") or "incorrect" (any other judgment). As predicted, the motivation to accept or reject the proposition had a significant impact on performance. More participants (correctly) rendered a judgment of no association between the variables when the proposition in question was unfavorable (57 percent) than when it was favorable (19 percent), $X^2(1) = 4.73, p < .03$. Participants were thus three times more likely to assess contingency correctly when evaluating a proposition they were motivated to reject, rather than accept.

It is understandable that participants who chose to view fewer than all four cells would err in their judgments about the data contained within. But these participants were then given a second chance to get it right. After each had stated a conclusion, we showed them the data in any remaining cells and asked again what conclusions should be drawn from the data. We examined the final judgments of those participants who had originally selected fewer than all four cells ($N = 18$) to determine if they detected the lack of association in the table after they had been exposed to the complete data. Those considering an unfavorable proposition were more apt to render a final judgment of “no relationship” (80 percent) than those considering a favorable proposition (31 percent), $X^2(1) = 3.43, p < .06$. Thus, participants’ motivations to accept or reject the proposition impacted their conclusions even after they had seen all the information necessary for an accurate assessment of contingency.

Discussion

Some participants in this study considered a proposition compatible with their own strongly held beliefs about program housing. These individuals tended to endorse a relatively lenient evidential standard; namely, that one should accept the proposition if evidence permits. This standard, a paraphrasing of “Can I believe it?,” apparently guided their evaluation of contingency in a standard covariation assessment task. Presented with data arranged in a four-cell table, these participants tended to view only some of the data and to be misled by the relatively large number in the present-present cell. Their truncated examination of evidence and susceptibility to the cell A bias led them to see a relationship where none existed—a misperception that persisted even after exposure to the complete data.

Other participants considered a proposition at variance with their desired belief. These individuals overwhelmingly endorsed the more stringent criterion of accepting the proposition

only on the basis of compelling evidence, akin to the implicit question "Must I believe it?" Their behavior was consistent with their endorsement of a higher standard: They tended to examine all four cells of the table and to consider the overall pattern of data rather than relying on limited comparisons of two or three cells. Participants motivated to reject the proposition were likely to conclude that it was not supported by the data. The relatively few participants in this condition who failed to initially consider all four cells (and who thus drew errant conclusions about the data) tended to correct their judgments when given the opportunity.

These results suggest that the skeptical mindset elicited by an unpalatable proposition guides reasoners to a more thorough and critical evaluation of evidence than they might otherwise attempt. This stance served participants well in Study 1, leading them to the correct solution to an historically challenging task. If we are to champion skeptical inquirers, however, we must address a possible alternative interpretation of their behavior: skeptics may view unfavorable evidence not through a sharper lens, but through an intractably conservative one. Perhaps participants in our study who evaluated a proposition they did not want to accept simply set an unrealistically high standard that no body of evidence, however conclusive, could have satisfied. If this is the case, their skeptical mindset, far from helping to reduce error, would simply have introduced a different type of error. Rather than accepting evidence too readily, as did participants motivated to endorse the proposition, they may have simply refused to accept any evidence at all.

To investigate this possibility, we conducted a second study similar to the first, but with the addition of a condition in which the data relevant to the proposition in question did, in fact, support it. In this case, the conditional probabilities in the 2 x 2 table were unequal, indicating a relationship between participation in the collegiate Greek system and below-average academic

performance. Our goals in this study were twofold: 1) to replicate the results of Study 1 in the condition in which the data presented to participants indicated no relationship between the two variables, and 2) to determine if people motivated to deny the relationship would nevertheless accept it in the face of compelling evidence.

Study 2: Greek Life

We recruited college students who described themselves as strongly supportive of or opposed to the Greek system (i.e., joining a sorority or fraternity). All participants were presented with the question of whether participation in the Greek system is associated with poor grades. We anticipated that this would suit the preferences of Greek system critics, but violate the preferences of Greek system supporters. As in Study 1, participants were asked to select one of two reasoning criteria they thought was most appropriate to apply. They were then shown summarized data purported to have been collected as part of a study of grading and the Greek system. In one condition, (bogus) data were arranged so that participation in the Greek system was independent of academic performance. In a second condition, however, the data were arranged such that sorority and fraternity members in the purported sample earned disproportionately low grades.

Method

Participants were 42 male and 77 female Cornell undergraduates who responded to class announcements and posted flyers seeking individuals “with a strong opinion about the Greek system, either positive or negative.” We omitted the data of 7 participants who indicated in debriefing that they had recalled from a statistics course the appropriate way to analyze data in a 2 x 2 table like the one used in this study. In addition, we omitted the data of 12 people who

indicated on our post-experiment attitude questionnaire that they had no opinion about the Greek system. This left a sample of 31 male and 69 female participants.

The method was similar to that of Study 1. Participants followed a typed page as the experimenter read aloud some contentions surrounding participation in the Greek system (e.g., some people believe it is an integral part of college, whereas others believe it's superficial). They were told that researchers had recently conducted a study of the relationship between participation in the Greek system and academic performance, and that their task would be to evaluate the data collected as part of that study. All participants were instructed to determine whether participation in the Greek system is associated with below-average grades.

Before evaluating any data, participants indicated on a worksheet which of two reasoning criteria they believed most appropriate for this problem: "People should conclude that on the whole, participating in the Greek system hurts students' grades only if the data compel such a conclusion--only if the data make it clear that no other conclusion is warranted" ("compel" criterion) or "People should conclude that on the whole, participating in the Greek system hurts students' grades if the data permit such a conclusion--if there are data that support that conclusion" ("permit" criterion). The order of presentation was counterbalanced.

The experimenter explained that a group of researchers had randomly selected a large number of college and university students from across the United States. They recorded whether each student was a member of a sorority or a fraternity, and whether each student's grade point average was above or below the national GPA for that student's major. The purported results were summarized in a 2 x 2 table with the first column labeled "Student is member of sorority/fraternity" and the second column "Student is not a member of sorority/fraternity." The first row was labeled "Student's GPA below average" and the second row, "Student's GPA

above average.” (see Appendix B). Participants were randomly assigned to one of two contingency conditions. In the “variables independent” condition, the numbers in each cell were as before: 138 (upper left cell, or "cell A"), 69 (upper right cell, or "cell B"), 46 (lower left cell, "cell C"), and 23 (lower right cell, "cell D"). Note that $(A / A+C) = (B / B+D) = .75$. In the “variables related” condition, the numbers in cells C and D were reversed, so that $(A / A+C) > (B / B+D)$. In this condition, the probability that a member of the Greek system had below average grades was .86, whereas the probability that a nonmember had below average grades was .60. Thus, the evidence in this condition reflected an association between participation in the Greek system and poor grades. Of the 100 participants, 55 (24 supporters and 31 opponents) were randomly assigned to the “variables independent” condition, and 45 (25 supporters and 20 opponents) to the “variables related” condition.

Upon initial presentation of the 2 x 2 table, the number in each cell was concealed with an index card. Participants were instructed to access as much information as they needed to test the proposition that participation in the Greek system is associated with below-average grades. They were instructed to base their conclusion not on their overall opinion about the Greek system, but only on the data before them. As before, the experimenter removed one card at a time at the participant’s request, stopping only when the participant indicated he or she had enough information on which to base a judgment. We were concerned that participants in Study 1 who publicly committed to a judgment before seeing all four cards may have been inhibited from changing their conclusions after all cells were revealed, even if they privately believed they should do so. Therefore, at this point in Study 2, the experimenter instructed participants who had selected fewer than all four cells to withhold their judgment until they had viewed the entire data. She then exposed any remaining cells and asked participants to select one of three options:

the data support the proposition, the data are inconclusive, or the data refute the proposition. Finally, participants indicated their level of agreement with six attitude statements about the Greek system (e.g., "Sororities and fraternities are an integral part of a college experience" and "Cornell would be better off if it banned the Greek system."). The questions were reverse-scored where appropriate so that higher responses on the 7-point scale indicate more favorable attitudes toward the Greek system.

Results

Attitude stance. Of the 100 retained participants, 49 indicated on an initial 1-item scale that they were favorable and 51 unfavorable toward the Greek system. Those who identified themselves as supporters gave higher ratings on the subsequent attitude statements about the Greek system ($M = 5.74, SD = .72$) than did participants who identified themselves as opponents ($M = 2.58, SD = .64$) $t(98) = 23.2, p < .0001$.

Reasoning criterion. All participants were asked how best to evaluate the claim that participation in the Greek system is harmful to students' grades by selecting one of the two criteria. As predicted, supporters of the Greek system claimed that the data should be held to a much higher standard. Eighty-six percent of Greek system supporters endorsed the relatively strict "compel" criterion, whereas only 40 percent of those opposed to the Greek system did so, $\chi^2(1) = 22.93; p < .0001$.

Number of cells examined. Participants uncovered cells of the 2 x 2 table one at a time to view the number contained in each. Ninety-eight percent named cell A as their first or second selection. In both conditions, the number in cell A was the largest of any of the 4 cells and so a comparison of this cell with any other 1 or 2 cells might appear supportive of a relationship between the Greek system and below-average grades. We expected that participants motivated

to accept this contention (i.e., Greek system opponents) to truncate their search for evidence as soon as it appeared to be validated, after viewing only 1, 2, or 3 cells. In contrast, we expected those motivated to reject the contention (i.e., Greek system supporters) would persevere until all 4 cells had been revealed, thus maximizing their opportunity to uncover faults in the (apparently) unfriendly evidence.

An analysis of the proportion of supporters and opponents who selected all four cells supported our hypothesis. Across both conditions, more Greek system supporters (82 percent) than opponents (51 percent) opted to examine all 4 cells, $X^2(1) = 10.46, p < .01$. The same pattern was evident within each condition. Eighty-four percent of supporters and 45 percent of opponents in the “variables related” condition selected all four cells ($X^2(1) = 7.61, p < .01$), while 56 percent of supporters and 44 percent of opponents in the “variables independent” condition did so ($X^2(1) = 3.54, p < .06$).

Assessment of contingency. In this study, unlike Study 1, the experimenter asked for participants’ judgments of association only after she had revealed the numbers in all four cells of the 2 x 2 table. That is, participants selected the number of cells they felt was sufficient and necessary for such a judgment, but only rendered a judgment after viewing the entire data. Therefore, all participants within each condition based their judgments on the same data. We were interested in the proportion of participants in each condition who judged membership in the Greek system to be unrelated to poor academic performance. This is the correct assessment in the “variables independent” condition, in which $(A / A+C) = (B / B+D)$. It is incorrect in the “variables related” condition, because $(A / A+C) > (B / B+D)$, but it corresponds to the presumed desired belief of Greek system supporters. If supporters were arbitrarily dismissive of the evidence before them, they should have been equally likely to conclude that there was “no

relationship” in each condition. If, on the other hand, their skeptical stance led these participants to reason more soundly, their judgments should demonstrate a sensitivity to the evidence.

We hypothesized that Greek system proponents would be more likely than opponents to make an assessment of “no relationship” in the “variables independent” condition, but no more likely to make such a conclusion in the face of the compelling evidence provided in the “variables related” condition. This hypothesis was supported. In the “variables independent” condition, significantly more proponents (79 percent) than opponents (32 percent) correctly concluded that there was no relationship between the Greek system and grades, $X^2(1) = 11.94, p < .001$. In the “variables related” condition, however, proponents were not more likely than opponents to draw this conclusion. That is, roughly equal proportions of Greek system fans (88 percent) and foes (95 percent) in this condition conceded that the Greek system was associated with below-average grades ($X^2(1) = .67, ns$). As Figure 4 illustrates, participants motivated to reject the proposition of a relationship between the variables were more likely than others to do so only when the data provided a basis for rejection (i.e., in the “variables independent” condition). In the presence of evidence that the Greek system is indeed related to poor grades, even most participants for whom this was an unpalatable proposition acknowledged the pattern inherent in the data. A planned contrast of pro-Greek participants in the “variables independent” condition with all others confirmed the reliability of this pattern of responses, $X^2(1) = 30.44, p < .0001$.

Discussion

In this study, we replicated our previous result suggesting that people tend to ask, essentially, “Can I believe this?” when evaluating favorable propositions, but “Must I believe this?” when evaluating unfavorable ones. In one condition, the answer to “Must I believe it?”

was “no”: the evidence, although arranged to mislead, was ultimately inconclusive.

Participants motivated to reject the proposition tended to detect the flaw in the evidence that escaped the notice of participants motivated to accept it. Their skepticism, furthermore, did not make them simply rejectionists; instead, it seemed to truly facilitate sound reasoning. In a second condition, the answer to “Must I believe this?” was “yes.” Faced with apparently conclusive evidence, even participants for whom the proposed association was unpalatable conceded its validity.

General Discussion

People have a strong inclination to believe that their beliefs are well founded, even if they are not. In general, faith in the validity of one’s own cherished beliefs may be a healthy response to an uncertain world (Brown & Dutton, 1995; Taylor & Brown, 1988). However, in a number of domains, satisfactory outcomes depend upon an accurate evaluation of evidence and a willingness to abandon unfounded convictions. Scientists and philosophers alike advocate a critical approach to the evaluation of ideas (e.g., Gilovich, 1991; Sagan, 1996; Schick & Vaughn, 1999; Stanovich, 2001). For his part, Bertrand Russell (1928) championed not William James’s “will to believe,” but rather, the “will to doubt.” The participants in our studies demonstrate the soundness of this preference. People who were motivated to believe a proposition set a relatively low standard for its evaluation, asking implicitly, “Can I believe this?” They searched for information consistent with the proposition, and halted their search before examining all the evidence available to them. As a result, these participants, like others in past studies of people’s ability to assess covariation, were susceptible to the cell A bias. They were misled by the frequent co-occurrence of two variables and concluded that the variables were related when, in fact, they were not. This was true even though the task made minimal demands on attention and

memory. In short, participants with the “will to believe” satisfied their will, but at the cost of rationality.

People motivated to invalidate the proposition aligned themselves with a higher standard. They asked themselves, “Must I believe this?,” and approached the problem with a goal of disconfirmation. These participants tended to examine the available data exhaustively and to reason critically about their implications. They were relatively resistant to the cell A bias and succeeded in discovering the flaw in the evidence that allowed rejection of the objectionable proposition. At the same time, these skeptics remained sensitive to the strength of valid evidence. They were willing to accept even an unpalatable proposition after a thorough examination of the data yielded no legitimate basis for dismissal. Overall, their “will to doubt” led them toward more accurate assessments of contingency.

We are not alone in our claim that people reason differently about seemingly friendly and unfriendly data. Ditto and colleagues make a similar point with their *quantity of processing* model, in which people threatened by a proposition seek more evidence pertinent to its validity and think about it harder (Ditto & Lopez, 1992; Ditto et al., 1998). Although similar in many respects, our Can I? Must I? account and the quantity of processing model are not identical. One point of divergence is the degree to which reasoning standards are applied in anticipation of disagreeable evidence rather than in response to it. Ditto et al. depict a process whereby people impose a more stringent standard on undesirable information they encounter. That is, a person might receive a piece of evidence, deem it threatening, and only then decide that more evidence is necessary. Our Can I? Must I? account, in contrast, applies to evidence anticipated as well as evidence encountered; people may select a reasoning criterion even before being exposed to motivationally relevant information. Participants in our studies told us up-front that they

intended to be relatively accepting or skeptical of the information to follow. Their anticipatory stance guided not only how they responded to evidence, but also whether they exposed themselves to it in the first place.

A second point of departure is whether the *amount* of processing is the only distinction between skeptics and true believers. Undeniably, people tend to give more attention to evidence that supports alarming propositions than reassuring ones, and are thus more likely to detect legitimate errors—a point captured by both the quantity of processing model and our Can I? Must I? account. At the same time, part of the experimental allure of covariation judgment and similar tasks is that performance on them is remarkably unaffected by quantity of processing; thinking longer about them doesn't necessarily render them more tractable. Thinking differently about them, however, does. Specifically, trying to disconfirm an unpalatable hypothesis is a more effective strategy than trying to confirm it. With regard to covariation assessment, the *quality* of processing may be more predictive of performance than the *quantity*.

There is convergent evidence that belief-incongruent proposals elicit more effective reasoning strategies for other types of problems, as well. In two studies, people were unusually adept at solving the Wason selection task when doing so afforded an opportunity to invalidate a threatening task rule (Dawson et al., 2002). In others, participants applied complex, task-appropriate strategies of inference to problems concerning negative stereotypes about a group to which they belonged (Doosje, Spears, & Koomen, 1995; Schaller, 1992). To the degree that they adopt a strategy of disconfirmation, people evaluating unfavorable propositions may be expected to perform well on any task which requires a negative, rather than positive, test strategy. This is not to say that skeptics are necessarily less biased than others, but simply that in many cases, the

bias under which they labor leads them to adopt heuristics that coincide with normative reasoning strategies for a particular task.

Given that people must make judgments like these every day, often with important consequences, a healthy sense of incredulity has much to recommend it. Confronting a proposition people would rather not be true seems to elicit a natural skepticism. It remains to be seen whether, absent this motivation, people are able to generate a similarly critical mindset. There is considerable research indicating that reasoners can improve the quality of their decisions by generating arguments against them (e.g., Lord, Lepper, & Preston, 1984), seeking disconfirming evidence (e.g., Dawson et al., 2002), and bearing a burden of accountability for their judgments (for a review, see Tetlock, 2002). They might also do well to adopt a relatively high standard by which to evaluate their own cherished beliefs, asking not “Can I believe this?” but the more skeptical, “Must I believe it?” A French proverb translates, “Skeptics are never deceived.” While this certainly overstates the case, it is true that skeptics may, at least, be deceived less often and less easily.

References

- Arkes, H.R., & Harkness, A.R. (1982). Estimates of contingency between two dichotomous variables. *Journal of Experimental Psychology: General*, *112*, 117-135.
- Batson, C.D., Thompson, E.R., Seufferling, G., Whitney, H., & Strongman, J.A. (1999). Moral hypocrisy: Appearing moral to oneself without being so. *Journal of Personality and Social Psychology*, *77*, 525-537.
- Beyth-Marom, Ruth. (1982). Perception of correlation reexamined. *Memory & Cognition*, *10*, 511-519.
- Brown, J.D., & Dutton, K. A. (1995). Truth and consequences: The costs and benefits of accurate self-knowledge. *Personality and Social Psychology Bulletin*, *21*, 1288-1296.
- Crocker, J. (1981). Judgment of covariation by social perceivers. *Psychological Bulletin*, *90*, 272-279.
- Crocker, J. (1982). Biased questions in judgment of covariation studies. *Personality and Social Psychology Bulletin*, *8*, 214-220.
- Dawson, E., Gilovich, T., & Regan, D.T. (2002). Motivated reasoning and the Wason selection task. *Personality and Social Psychology Bulletin*, *28*, 1379-1387.
- 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,

Edwards, K., & Smith, E.E. (1996). A disconfirmation bias in the evaluation of arguments. *Journal of Personality and Social Psychology, 71*, 5-24.

Gerard, H.B., & Hoyt, M.F. (1974). Distinctiveness of social categorization and attitude toward ingroup members. *Journal of Personality and Social Psychology, 29*, 836-842.

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

Holton, B., & Pyszczynski, T. (1989). Biased information search in the interpersonal domain. *Personality and Social Psychology Bulletin, 15*, 42-51.

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

Klayman, J., & Ha, Y.-W. (1989). Hypothesis testing in rule discovery: Strategy, structure, & content. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 596-604.

Kruglanski, A. W. (1980). Lay epistemology process and contents. *Psychological Review, 87*, 70-87.

Kruglanski, A.W., & Meinholdt, C. (1990). Cognitive and motivational bases of judgmental bias: Toward a synthesis. *Polish Psychological Bulletin, 21*, 291-305.

Kruglanski, A.W., & Webster, D.M. (1996). Motivated closing of the mind: “Seizing” and “freezing.” *Psychological Review*, *103*, 263-283.

Kunda, Z. (1987). Motivated inference: Self-serving generation and evaluation of causal theories. *Journal of Personality and Social Psychology*, *25*, 272-285.

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

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.

Lord, C.G., Lepper, M.R., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology*, *47*, 1231-1243.

McKenzie, C.R.M. (1994). The accuracy of intuitive judgments strategies: Covariation assessment and Bayesian inference. *Cognitive Psychology*, *26*, 209-239.

Mynatt, C.R., Doherty, M.E., & Dragan, W. (1993). Information relevance, working memory, and the consideration of alternatives. *The Quarterly Journal of Experimental Psychology*, *46A*, 759-778.

Nisbett, R.E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.

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.

Russell, B. (1928). *Sceptical essays*. New York: W.W. Norton.

- Sagan, C. (1996). *The demon haunted world: Science as a candle in the dark*. New York: Random House.
- Sanbonmatsu, D.M., Posavac, S.S., Kardes, F.R., & Mantel, S.F. (1998). Selective hypothesis testing. *Psychonomic Bulletin & Review*, 5, 197-220.
- Sanger, D.E. (2001, July 25). Bush's European theater. *The New York Times*, p.A6.
- 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.
- Schustack, M.W., & Sternberg, R.J. (1981). Evaluation of evidence in causal inference. *Journal of Experimental Psychology: General*, 110, 101-120.
- Shaklee, H., & Tucker, D. (1980). A rule analysis of judgments of covariation between events. *Memory & Cognition*, 8, 459-467.
- Smedslund, J. (1963). The concept of correlation in adults. *Scandinavian Journal of Psychology*, 4, 165-173.
- Snyder, M., & Cantor, N. (1979). Testing hypotheses about other people: The use of historical knowledge. *Journal of Experimental Social Psychology*, 15, 330-342.
- Snyder, M., & Swann, W.B. (1978). Hypothesis testing in social interaction. *Journal of Personality and Social Psychology*, 36, 1202-1212.
- Stanovich, K.E. (2001). *How to think straight about psychology*. Boston, MA: Allyn & Bacon.

Taylor, S.E., & Brown, J.D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, *103*, 193-210.

Tetlock, P.E. (2002). Intuitive politicians, theologians, and prosecutors: Exploring the empirical implications of deviant functionalist metaphors. In Gilovich, T., Griffin, D., & Kahneman, D. (Eds.), *Heuristics and biases: The psychology of intuitive judgment*. New York: Cambridge University Press.

Trope, Y., & Bassok, M. (1982). Confirmatory and diagnostic strategies in hypothesis testing. *Journal of Personality and Social Psychology*, *43*, 22-34.

Trope, Y., & Bassok, M. (1983). Information gathering strategies in hypothesis testing. *Journal of Experimental Social Psychology*, *19*, 560-576.

Tweeney, R.D., Doherty, M.E., Worner, W.J., Pliske, D.B., & Mynatt, C.R. (1980). Strategies of rule discovery in an inference task. *Quarterly Journal of Experimental Psychology*, *32*, 109-123.

Vallone, R.P., Ross, L., & Lepper, M.R. (1985). The hostile media phenomenon: Biased perception and perceptions of media bias in coverage of the Beirut massacre. *Journal of Personality and Social Psychology*, *49*, 577-585.

Ward, W.C., & Jenkins, H.M. (1965). The display of information and the judgment of contingency. *Canadian Journal of Psychology*, *19*, 231-241.

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

Wasserman, E.A., Dorner, W.W., & Kao, S.F. (1990). Contributions of specific cell information to judgments of interevent contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 509-521.

Appendix A

Data tables presented to participants in Study 1

Condition: “Helps grades”

| | University allows access to program housing | University does not allow access to program housing |
|---|--|--|
| Minority students’ GPA above average | (A) n=138 | (B) n=69 |
| Minority students’ GPA below average | (C) n=46 | (D) n=23 |

Condition: “Hurts grades”

| | University allows access to program housing | University does not allow access to program housing |
|---|--|--|
| Minority students’ GPA below average | (A) n=138 | (B) n=69 |
| Minority students’ GPA above average | (C) n=46 | (D) n=23 |

Note: Upon initial presentation, each cell was covered with an index card labeled with the corresponding cell letter.

Appendix B

Data tables presented to participants in Study 2

Condition: “Variables independent”

| | Student is member of Greek system | Student is not member of Greek system |
|--------------------------------|--------------------------------------|--|
| Student’s GPA below average | (A) n=138 | (B) n=69 |
| Student’s GPA above average | (C) n=46 | (D) n=23 |

Condition: “Variables related”

| | Student is member of Greek system | Student is not member of Greek system |
|--------------------------------|--------------------------------------|--|
| Student’s GPA below average | (A) n=138 | (B) n=69 |
| Student’s GPA above average | (C) n=23 | (D) n=46 |

Note: Upon initial presentation, each cell was covered with an index card labeled with the corresponding cell letter.

Authors' Notes

This research was supported by National Science Foundation Research Grant SBR9809262. Portions of this research were presented at the annual meetings of the Eastern Psychological Association, April 1999, Providence, Rhode Island, and the American Psychological Society, June 2000, Miami, Florida.

Correspondence concerning this article should be addressed to Erica Dawson, Johnson Graduate School of Management, Sage Hall, Cornell University, Ithaca, NY 14853. Email: ECD6@cornell.edu, TDG1@cornell.edu, or DTR1@cornell.edu.

Footnotes

¹ Responses were coded according to each participant's specific judgment of contingency, ignoring any caveats that qualified, but did not undermine, the purported relationship. As an example, the response "Correlation doesn't prove causation" was coded as a judgment of "positive relationship" because the participant failed to detect the equal conditional probabilities. Judgments of "negative relationship" indicated that a person in the "helps grades" condition concluded that program housing actually hurts grades, or that a person in the "hurts grades" condition decided that, in fact, it helps them.

Table 1

Participants' judgments of contingency, Study 1

| Group | n | Judgment | | |
|-----------------------------|----|----------|---------|----------------|
| | | "Helps" | "Hurts" | "Inconclusive" |
| <i>Pro-program housing</i> | | | | |
| Condition: Helps grades | 11 | 73 | 0 | 27 |
| Condition: Hurts grades | 8 | 37 | 0 | 63 |
| <i>Anti-program housing</i> | | | | |
| Condition: Helps grades | 6 | 33 | 17 | 50 |
| Condition: Hurts grades | 5 | 0 | 100 | 0 |

Note: Numbers represent percent of participants making each judgment after viewing a self-selected number of cells of a 2 x 2 table in which the variables "access to program housing" and "academic performance" were statistically independent.

Figure Captions

Figure 1. Example of a 2 x 2 contingency table.

Figure 2. Percentage of participants choosing “permit” criterion, Study 1.

Figure 3. Percentage of participants choosing to view all 4 cells before attempting a judgment of contingency, Study 1.

Figure 4. Percentage of participants rendering a judgment of “no relationship,” Study 2.

Figure 4 note. All participants considered the proposition that the Greek system is detrimental to students’ grades.

| | | <u>DISEASE A</u> | |
|------------------|---------|-------------------|------------------|
| | | Present | Absent |
| <u>SYMPTOM X</u> | Present | (cell A) n=138 | (cell B) n=69 |
| | Absent | (cell C) n=46 | (cell D) n=23 |