Journal of Personality and Social Psychology
August 1993 Vol. 65, No. 2, 221-233

You Can't Not Believe Everything You Read

Daniel T. Gilbert
Department of Psychology University of Texas at Austin
Romin W. Tafarodi
Department of Psychology University of Texas at Austin
Patrick S. Malone
Department of Psychology University of Texas at Austin

ABSTRACT

Can people comprehend assertions without believing them? Descartes (1644/1984) suggested that people can and should, whereas Spinoza (1677/1982) suggested that people should but cannot. Three experiments support the hypothesis that comprehension includes an initial belief in the information comprehended. Ss were exposed to false information about a criminal defendant (Experiments 1 and 2) or a college student (Experiment 3). Some Ss were exposed to this information while under load (Experiments 1 and 2) or time pressure (Experiment 3). Ss made judgments about the target (sentencing decisions or liking judgments). Both load and time pressure caused Ss to believe the false information and to use it in making consequential decisions about the target. In Spinozan terms, both manipulations prevented Ss from "unbelieving" the false information they automatically believed during comprehension.

This article was written while Daniel T. Gilbert was a Fellow at the Center for Advanced Study in the Behavioral Sciences. The fellowship was supported by the John D. and Catherine T. MacArthur Foundation and by a Research Scientist Development Award (1—KO2—MH00939—01) from the National Institute of Mental Health. The research reported in this article was supported by grants to Daniel T. Gilbert from the National Science Foundation (BNS-8819836) and the National Institute of Mental Health (1—R01—MH49613—01) and by a Doctoral Fellowship to Romin W. Tafarodi from the Social Sciences and Humanities Research Council of Canada (SSHRC 752—91—3107). The generous support of these institutions is gratefully acknowledged.
We thank Carolyn Vu and Joe Mogab for their able assistance with the execution of these experiments and several anonymous reviewers for their critical comments on a previous version of this article.
Correspondence may be addressed to Daniel T. Gilbert, Department of Psychology, University of Texas, Mezes Hall 330, Austin, Texas, 78712.
Electronic mail may be sent to dgilbert@utxvm.bitnet
Received: April 13, 1992

Revised: January 28, 1993
Accepted: January 28, 1993

Man's most valuable trait is a judicious sense of what not to believe.

–Euripides, "Helen," *Euripides II: Four Tragedies* (412 B.C.)

One of us can still recall the day when he saw his first pair of x-ray glasses advertised on the inside cover of a comic book. The possibility of looking right through things seemed well worth the $1.99, so he mailed an envelope filled with quarters and pennies and waited to be endowed with extraordinary visual powers in just 4—6 weeks. When the x-ray glasses arrived, their red cellophane lenses were a serious disappointment. "But it said in the ad that you could see through stuff," he told his mother. "You can't believe everything you read," she said. "Oh yeah?" he replied, "Well *I* can."

The Skeptical Canon

Believing what one reads seems so easy and so natural that people must take pains to guard against it. René Descartes (1644/1984) formalized this intuition when he suggested that if one wishes to know the truth, then one should not believe an assertion until one finds evidence to justify doing so. Although this was a rather radical suggestion in a century characterized by strict obedience to civil authority and blind faith in religious doctrine, 300 years later it stands as the cardinal rule of science: One may entertain any hypothesis, but one may only believe those hypotheses that are supported by the facts. Implicit in this rule is an important assumption about human psychology. No law of science, church, or state can require people to do what they are clearly incapable of doing, and thus the injunction that one *must* control one's beliefs is predicated on the assumption that one *can* control one's beliefs. In some sense, the whole of modern science is based on the Cartesian assumption that people do not have to believe everything they read.

This capacity for skepticism is the heart not only of the scientific method, but of modern democracy as well. The First Amendment's guarantee of freedom of speech is also grounded in Cartesian psychology. John Stuart Mill (1859/1975) argued (as Milton and Locke had before him) that people can achieve true beliefs only when their society allows all ideas–true or false–to be expressed, examined, and debated. Mill disputed the common claim that if false ideas are allowed free expression, then people will be seduced into believing what they should not. He argued that there are clear benefits to entertaining false ideas (e.g., they may have a grain of truth, they may force the person who would reject them to see the truth more clearly, etc.). Moreover, Mill argued that no harm can come from allowing false doctrines to enter the "marketplace of ideas" because the educated individual is capable of choosing to accept or to reject the ideas with which he or she has contact. A marketplace of ideas draws on the same psychological assumption as does a marketplace of automobiles or fresh fruit: If people want Toyotas or Chevys or little green apples, then they can buy them. If not, they can walk on by. Just as

Descartes's (1644/1984) canon has become the essential principle of modern science, Mill's explication of that canon has become the essential principle of modern democracy.

The advent of science and of democracy are, perhaps, the two most important historical phenomena of this century. But is the psychological assumption that legitimates them correct? Do ideas really constitute a free marketplace? Can people inspect an idea and then simply walk on by if they find it wanting? Are people capable of the skepticism that good science and free speech apparently require?

One Act or Two?

To answer this question, it is important to understand why Descartes (1644/1984) was so willing to assume that people can behave skeptically. Descartes considered understanding and believing to be separate and sequential psychological operations (the second of which one is morally compelled not to perform unless there is justification for doing so). This "doctrine of separate mental acts" is the psychological foundation not only of Descartes's canon but of much modern psychological theorizing as well. Zimbardo and Lieppe (1991) summarized the conventional wisdom on this point:

Learning requires that the audience pay *attention* to the message and, in turn, gain some *comprehension* of it, understanding the new beliefs it proposes. Then, if the message has compelling arguments, *acceptance* of its conclusion and a change in attitude will follow. (p. 135)

First people comprehend a message, and then later they may accept it. Understanding and believing are today taken to be the separate and sequential operations that Descartes (1644/1984) described.

However, Descartes's (1644/1984) contemporary, Benedict Spinoza (1677/1982) , did not accept this doctrine, and he argued that understanding and believing are merely two words for the same mental operation. Spinoza suggested that people believe every assertion they understand but quickly "unbelieve" those assertions that are found to be at odds with other established facts. For Spinoza, "mere understanding" was a psychological fiction–a non sequitur that grew out of a fundamental misunderstanding of the nature of mental representation (see Gilbert, 1993). The details surrounding this misunderstanding are somewhat afield of the present concerns, but Spinoza's bottom line is not: According to Spinoza, the act of understanding is the act of believing. As such, people are incapable of withholding their acceptance of that which they understand. They may indeed change their minds after accepting the assertions they comprehend, but they cannot stop their minds from being changed by contact with those assertions.

Acceptance, then, may be a passive and inevitable act, whereas rejection may be an active operation that undoes the initial passive acceptance. The most basic prediction of this model is that when some event prevents a person from "undoing" his or her initial acceptance, then he or she should continue to believe the assertion, even when it is patently false. For example, if a person is told that lead pencils are a health hazard, he or

she must immediately believe that assertion and only then may take active measures to unbelieve it. These active measures require cognitive work (i.e., the search for or generation of contravening evidence), and if some event impairs the person's ability to perform such work, then the person should continue to believe in the danger of lead pencils until such time as the cognitive work can be done. The Cartesian hypothesis, on the other hand, makes no such prediction. That hypothesis suggests that both acceptance and rejection of an assertion are the results of cognitive work that follows comprehension of the assertion. As such, interruption should make both of these options impossible and thus leave the person in a state of nonbelief rather than one of belief or disbelief.

What the Evidence Justifies Believing

A variety of evidence suggests that people have a tendency to believe what they should not (see Gilbert, 1991 , for a review). For example, repeated exposure to assertions for which there is no evidence increases the likelihood that people will believe those assertions (Arkes, Boehm, & Xu, 1991 ; Arkes, Hacket, & Boehm, 1989 ; Begg, Armour, & Kerr, 1985 ; Hasher, Goldstein, & Toppino, 1977). Once such beliefs are formed, people have considerable difficulty undoing them (Anderson, 1982 ; Lindsay, 1990 ; Ross, Lepper, & Hubbard, 1975 ; Schul & Burnstein, 1985 ; Wilson & Brekke, 1992 ; Wyer & Budesheim, 1987 ; Wyer & Unverzagt, 1985). Moreover, several studies have suggested that under some circumstances people will believe assertions that are explicitly labeled as false (Gerrig & Prentice, 1991 ; Gilbert, Krull, & Malone, 1990 ; Wegner, Coulton, & Wenzlaff, 1985). If people are capable of withholding their acceptance of that which they comprehend, then the presentation of explicity false information would provide a propitious opportunity to do so. Yet, people do not always seem to exercise that option.

Although evidence suggests that people are prone to believe experimentally presented assertions for which there is no supporting evidence, most of these results can be explained within the Cartesian as well as the Spinozan framework. For example, studies typically use ordinary assertions that draw on the subject's real-world knowledge, thus allowing the subject to make a personal decision about the veracity of the assertion that may conflict with the experimenter's claims. In addition, when real-world assertions are used, it may be difficult to prevent people from retrieving relevant information from memory and assessing the veracity of the assertions. Gilbert et al. (1990) attempted to circumvent these problems by presenting subjects with assertions whose veracity they could not assess because one word of the assertion was in a foreign language (e.g., "A monishna is a star"). After reading each assertion, subjects were sometimes told that the assertion was true or that it was false. On some trials, subjects were interrupted by a tone-detection task just a few milliseconds after being told of the assertion's veracity. At the end of the experiment, subjects were asked to recall whether each assertion had been labeled as true or as false. The Spinozan hypothesis predicted that interruption (a) would prevent subjects from unbelieving the assertions that they automatically accepted on comprehension and would thus cause subjects to report that false assertions were true, but (b) would not cause subjects to report that true assertions were false. This asymmetry did, in fact, emerge and is not easily explained by the Cartesian hypothesis.

Unfortunately, even these studies do not directly address the key element of the Spinozan hypothesis. Gilbert et al. (1990) measured subjects' memory for the assertions they had seen, and when interrupted, subjects did mistakenly report that some assertions that were labeled as false had been labeled as true. When a person says, "The assertion 'A monishna is a star' was labeled true in the learning phase of this experiment," he or she is technically implying the presence of a belief (i.e., "I believe that a monishna is a star"). Nonetheless, remembering that a foreign phrase was paired with the label *true* does not seem to capture the essence of what most people mean by the word *belief.* Indeed, when people believe, they do more than remember that something was said to be so; they actually behave as though that something were so. Most philosophers (and virtually all psychologists) consider action to be the sine qua non of belief, but Gilbert et al. (1990) only measured memory for the veracity of an assertion and never examined subjects' tendencies to use the information that they recalled. Hastie and Park (1986) have cogently argued that what people recall and what they actually believe are often uncorrelated, and they have provided extensive documentation for this argument.

The validity of the Spinozan hypothesis, then, is still very much an open question. Our goals in performing the following experiments were as follows: First, we hoped to gather further evidence of effects that are predicted by the Spinozan hypothesis and that cannot be accounted for by the Cartesian hypothesis. Second, we hoped to gather the first evidence that interruption after comprehension leaves people in their initial state of acceptance, and that this state truly constitutes a belief insomuch as people will base consequential social behavior on it. To this end, we asked subjects in Experiment 1 to play the role of a trial judge and to make sentencing decisions about an ostensibly real criminal defendant. Subjects were given some information about the defendant that was known to be false and were occasionally interupted while reading it. We predicted that interruption would cause subjects to continue to believe the false information they accepted on comprehension and that these beliefs would exert a profound influence on their sentencing of the defendant.

Experiment 1

Method Overview

Subjects read a pair of crime reports that contained both true and false statements. The color of the text indicated whether a particular statement was true or false. One report contained false statements that exacerbated the severity of the crime, and the other report contained false statements that extenuated the severity of the crime. Some subjects performed a concurrent digit-search task as they read the false statements in the reports. Finally, subjects recommended the length of prison terms for each criminal, rated the criminal on several related dimensions, and completed a recognition memory test for some of the statements contained in the reports.

Subjects

Seventy-one female students at the University of Texas at Austin participated to fulfill a requirement in their introductory psychology course. Only native speakers of English were eligible to participate. [1]

Instructions

On arriving at the laboratory, subjects were escorted to a private cubicle, where they remained for the duration of the experiment. Subjects were given written instructions, and a female experimenter reviewed the instructions to be sure that subjects had understood them.

The crime reports.

Subjects were told that they would be reading reports of two unrelated criminal incidents that had recently occurred in Austin, TX. They were told that they would play the role of a trialcourt judge and that they should attend carefully to the details of each report because later they would be asked to make judgments about the perpetrators of the crimes and to remember the nature and circumstances of the crimes. Subjects were told that the crime reports would "crawl" across the screen of a color video monitor (much like an emergency weather bulletin crawls across the bottom of a television screen). Subjects were told that statements printed in black were true statements, but that statements printed in red were false statements. Subjects were told that the false statements were "details about the crimes that don't really belong to the reports they appear in. They were actually taken from other, unrelated police reports and then mixed in with the facts" (i.e., the true statements). This was done so that subjects would not assume that a false statement (e.g., "The robber had a gun") could be negated to create a true statement. Subjects were told to read the reports aloud.

Although no rationale for the inclusion of false information was offered, we assumed that subjects would not consider it unusual for a trial judge to be presented with both true and false testimony in the course of his or her deliberations. Indeed, despite the lack of a specific rationale, no subject questioned the presence of the false information.

The digit-search task.

Subjects were told that as the crime report crawled across the screen, a string of blue digits would occasionally crawl across the screen on a line just beneath the text (hereinafter referred to as the number line and the text line, respectively). Subjects were given a hand-held counter that was ostensibly connected by a cable to a computer in another room and were told to press the button on the counter whenever the digit *5* appeared in the number line. Subjects were told that the nearby computer would record the accuracy of their button-presses. In addition, subjects were told that when the first report ended, the text line would begin to display digits. When this happened, subjects were instructed to search both the text line and the number line for the digit *5*.

All subjects were given practice at the digit search. After this practice, half the subjects were told that they had been assigned to a control condition (hereinafter called the uninterrupted condition) and that they should be prepared to read the text of the crime reports but should ignore any numbers that appeared in the number line or the text line. The experimenter took the hand-held counter from these subjects. The remaining subjects were told to be prepared to read the text of the crime reports and concurrently to search the number line and the text line for the digit *5* (hereinafter called the interrupted condition). These subjects retained the hand-held counter.

Procedure Practice phase.

All subjects were given practice reading the crawling text. Subjects read a short narrative about the Iowa State Fair as it crawled across the text line. In both the practice and the experimental phases, subjects read the text aloud to ensure that they did read it. Digits also crawled across the number line, and all subjects practiced using the counter and searching the number line for the digit *5*. This practice was intended to make subjects feel comfortable with the experimental task. Previous research (e.g., Schneider & Shiffrin, 1977 ; Smith, 1989 ; Smith, Stewart, & Buttram, 1992) has shown that much greater amounts of practice than this are necessary before these sorts of tasks become automatized, and thus practice was not expected to diminish the effect of the digit-search task during the experiment. Subjects were allowed to perform the practice task twice.

Experimental phase.

Following the practice phase, each subject was shown two crime reports in succession. Subjects saw a white, 1.5-in. (3.8-cm) horizontal stripe across the middle of an otherwise black screen. Subjects were told that this stripe would serve as a "window" for both the text line and, beneath that, the number line. Fifteen seconds later, the prompt "GET READY" was superimposed on the stripe in black print. The prompt flashed for 15 s and then crawled to the left of the screen. The prompt was followed by the first word of the first crime report. The characters on both the text line and the number line were 0.5 in. (1.3 cm) high and .38 in. (.96 cm) wide, and crawled at approximately 16 characters per second. The number line remained blank until a false statement (printed in red) appeared in the text line. At that time, digits appeared on the number line. These digits were separated by a space of 1.25 in. (3.18 cm). The first digit appeared on the number line beneath the first letter of the first word of the first false statement on the text line, and the last digit appeared on the number line 6.25 in. (15.88 cm, or five digits) after the last letter of the last word of the first false statement appeared on the text line. The value of each digit was randomly determined with the constraints that (a) 20% of the digits in each report must be *5* s and (b) at least one *5* must occur beneath each false statement. The last word of each report was followed immediately by a sequence of 15 digits on the text line and 15 digits on the number line, thus temporarily creating two lines that the subject had to search for the digit *5*. When the last of these 15 digits had run off the screen, the text line and the number line remained blank for 30 s. The experimenter instructed the subjects to close their eyes and review the information they had just learned "in order to sift the fact from the fiction." The "GET READY" prompt appeared for a second time,

and subjects were asked to open their eyes and prepare to read the second report. The second report was presented in the same manner as the first.

The first report described how a perpetrator named Tom had robbed a stranger who had given him a ride, and the second report described how a perpetrator named Kevin had robbed a convenience store. Each report contained seven false statements that were printed in red. In one report, the false statements would have exacerbated the severity of the crime had they been true, and in the other report the false statements would have extenuated the severity of the crime had they been true. All subjects read the report about Tom before the report about Kevin. Some subjects saw a report whose false statements extenuated Tom's crime (described in the first report) and exacerbated Kevin's (described in the second report), and the remaining subjects saw a report whose false statements exacerbated Tom's crime (described in the first report) and extenuated Kevin's (described in the second report). The false statements were constructed such that their elimination did not impair the grammatical integrity of the sentences in which they were embedded or the structural integrity of the crime stories themselves. In addition, the false statements were logically independent both of each other (i.e., the content of one neither implied nor refuted the content of another) and of the true statements (i.e., the content of a true statement neither implied nor refuted the content of a false statement, and vice versa).

Dependent Measures Prison terms.

After reading the second report, subjects were asked to complete the primary dependent measure. This measure required that subjects consider the facts of each crime and recommend a prison term between 0 and 20 years for each of the perpetrators.

Other ratings.

After completing the primary measure, subjects completed three 9-point Likert scales that measured (a) their feelings toward each of the perpetrators (anchored at the extremes with the words *neutral* to *extreme dislike*), (b) how much they thought the perpetrator would be helped by counseling (anchored at the extremes with the phrases *not helped at all* to *helped a great deal*), and (c) how dangerous they thought the perpetrator was (anchored at the extremes with the words *slightly* to *extremely*). Subjects were asked to mark each scale with both a *T* and a *K* to indicate how they felt about Tom and Kevin, respectively.

Recognition memory.

After completing the ratings, subjects were shown 30 sentences and were asked to classify each as (a) a true statement from the first crime report, (b) a false statement from the first crime report, or (c) a statement that never appeared in the first crime report. A similar test was then given for the second report. Each test contained 4 sentences that had appeared as true statements in the report, 7 sentences that had appeared as false statements in the report, and 19 sentences that had never appeared in the report. None of the true or false sentences was taken verbatim from the text; rather, each captured the gist of a true or false statement that had appeared in the text. This was done to discourage

subjects from remembering unusual words or turns of phrase rather than the facts of the reports. On completion of the recognition memory tests, subjects were fully debriefed, thanked, and dismissed.

Results and Discussion

Three subjects were omitted from all analyses due to either suspicion about the procedure or excessive difficulty reading the moving text. Of the remaining 68 subjects, 34 were required to perform the concurrent digit-search task. On average, these subjects failed to detect 4.06 of the 44 digits for which they had been instructed to search. This error rate suggests that the digit-search task was demanding but not overwhelming, as intended. The incidence of false alarms was so low in pretesting that it was not measured here.

Prison Terms and Other Ratings

Subjects read aloud a pair of crime reports. Some subjects performed a digit-search task as they read the false exacerbating or false extenuating statements that were embedded in the reports. We expected these false statements to affect the prison terms recommended by subjects who performed the digit-search task (the interrupted condition), but not those recommended by subjects who performed no digit-search task (uninterrupted condition). Subjects' recommendations for prison terms were submitted to a 2 (interruption: interrupted or uninterrupted) × 2 (false statements: extenuating or exacerbating) analysis of variance (ANOVA). [2] Only the last of these was a within-subjects variable. The analysis revealed a main effect of interruption, F (1, 66) = 4.21, $p < .05$, and a main effect of false statements, F (1, 66) = 68.45, $p < .001$, both of which were qualified by the predicted Interruption × False Statements interaction, F (1, 66) = 32.00, $p < .001$. As Table 1 shows, the prison terms recommended by uninterrupted subjects were only marginally affected by the nature of the false statements they had read, F (1, 66) = 3.42, $p < .07$, but the prison terms recommended by interrupted subjects were reliably affected by the nature of the false statements they read, F (1, 66) = 97.03, $p < .001$. Interrupted subjects recommended that perpetrators serve nearly twice as much time when the false statements contained in the police reports exacerbated (rather than extenuated) the severity of the crimes.

Subjects' ratings of the perpetrators' dislikableness, dangerousness, and likelihood of deriving benefit from counseling were submitted to separate 2 × 2 ANOVAs (as earlier). As Table 1 shows, these exploratory measures revealed patterns quite similar to the pattern seen on the primary measure. Reliable Interruption × False Statements interactions emerged for dislikableness, F (1, 66) = 5.25, $p < .05$, and for dangerousness, F (1, 66) = 5.75, $p < .05$. Although the interaction for benefit from counseling was not reliable, F (1, 66) = 1.55, $p = .22$, planned comparisons revealed that even on this measure the judgments of interrupted subjects were significantly affected by the nature of the false statements they had read, F (1, 66) = 9.08, $p < .01$, whereas the judgments of uninterrupted subjects were not, F (1, 66) = 1.57, $p = .21$. In short, interruption caused subjects to act as though false statements were true.

Recognition Memory

We predicted that false statements would be particularly likely to affect interrupted subjects' judgments about perpetrators because interrupted subjects would be particularly likely to remember those statements as true. Table 2 shows the results of the recognition memory test. Three results are especially noteworthy:

- Interrupted and uninterrupted subjects were equally likely to misremember true statements. Because true statements were never interrupted, this result provides assurance that interrupted and uninterrupted subjects did not differ, either before or during the experiment, in their general memorial ability.
- Interrupted subjects were more likely than uninterrupted subjects to misremember false statements as true. This is precisely the effect that our hypothesis predicted.
- Interrupted subjects were more likely to misremember foil items as true than were uninterrupted subjects.

This last finding may be seen to suggest that interrupted subjects suffered from a general guessing bias. In other words, interrupted subjects may have been especially likely to claim that false statements were true because they were especially likely to claim that all statements were true. This alternative explanation (which would sharply undermine our interpretation of the results) is not viable for three reasons. First, interrupted subjects were not especially likely to claim that true statements were true; a general guessing bias should indeed lead to an increased number of misses (e.g., false statements misremembered as true), but it should also lead to an increased number of hits (true statements remembered as true). This did not happen in our study. Second, the tendency for interrupted subjects to remember foils as true is indeed reliable, but it is rather small when compared with the tendency for interrupted subjects to misremember false statements as true. Even if interrupted subjects' responses to foil items did represent a guessing bias, the increase created by that bias (4%) would be far too slight to account for the increase in subjects' tendency to remember false statements as true (21%). Finally, and most important, the number of false statements that subjects misremembered as true was reliably correlated with the length of the prison term they recommended; that is, subjects who misremembered as true the most false exacerbating statements about one perpetrator and the fewest false extenuating statements about the other perpetrator were also more likely to recommend longer prison terms for the former than the latter, r (68) = .29, $p < .02$. Similar correlations were found between this measure of memory and the other ratings (dislikableness, $r = .39$, $p = .001$; dangerousness, $r = .34$, $p = .005$; benefit from counseling, $r = -.12$, $p = .34$). These correlations are not what one would expect if subjects had merely claimed that the false statements were true (which is what the general guessing bias explanation suggests they do), but it is precisely what one would expect if subjects believed that the false statements were true (which is what our hypothesis suggests they do).

The data from this study, then, do a rather good job of ruling out the general guessing bias explanation. Nonetheless, they do not rule out a slightly more sophisticated version of that explanation, which we call the specific guessing bias explanation. One might

argue that subjects do not have a pervasive or general tendency to guess that an unremembered item is true, but that they do have a tendency to make such guesses when presented with unremembered items that were interrupted. In other words, if a guessing bias were to emerge only under conditions of interruption, it would predict just the pattern of memory data we obtained. There are at least two reasons why this alternative explanation should immediately be suspect. First, Gilbert et al. (1990) examined just this hypothesis and found it wanting. Guessing is associated with uncertainty, and uncertainty is associated with hesitation; when people feel uncertain about an answer, they hesitate and then they guess. Gilbert et al. measured the time it took subjects to respond to interrupted and uninterrupted items and found no reliable differences in response times. Second, the specific guessing bias explanation (like its more general cousin) cannot account for the pattern of the prison term data or the correlation between recommended prison terms and memory errors. If subjects were merely guessing that an interrupted item on a recognition memory test was true, then why did they act as though they believed the item was true when they made judgments before taking the recognition memory test?

Although the specific guessing bias explanation is weak, it cannot be dismissed entirely. However, a very simple experiment should provide a critical test of its validity. If interruption causes subjects to guess that an unremembered item is true, then it should increase the number of errors subjects make when responding to false items (which we have shown it does), but it should also decrease the number of errors subjects make when responding to true items. In Experiment 1, true items were never interrupted, and thus this part of the prediction cannot be tested. Thus, we performed a second experiment with the express purpose of determining whether interruption increases the tedency for subjects to claim that a true item is true.

Experiment 2

Method Overview

Subjects read a crime report that contained both true and false statements. The color of the text indicated whether a particular statement was true or false. The report contained two critical statements that exacerbated the severity of the crime. Some subjects performed a concurrent digit-search task as they read these two critical statements, and others did not. For some subjects, these critical statements were ostensibly true, and for some subjects they were ostensibly false. Finally, subjects completed a recognition memory test for the sentences contained in the report.

Subjects

Eighty-six female students at the University of Texas at Austin participated to fulfill a requirement in their introductory psychology course. Only native speakers of English were eligible to participate.

Instructions and Procedure

The instructions and procedures were virtually identical to those used in Experiment 1, with some notable exceptions. First, all subjects read a report in which seven key statements exacerbated the severity of a crime perpetrated by a character named Tom. The only difference between this report and the report used in Experiment 1 was a small change in two of the exacerbating statements. Half the subjects were assigned to the critical—false condition, and the remaining subjects were assigned to the critical—true condition. As Table 3 shows, subjects in the critical—false condition saw seven exacerbating statements, all of which were presented as false. Subjects in the critical—true condition saw the same seven exacerbating statements, but two of those statements (hereinafter called the critical statements) were presented as true (i.e., the color of the text was changed from red to black).

As in Experiment 1, half the subjects were assigned to the interrupted condition (i.e., they performed a digit-search task while reading the seven exacerbating statements) and half were assigned to the uninterrupted condition (i.e., they performed no such task). To balance the number of false statements across conditions, two neutral statements that were presented to critical—false subjects as true were presented to critical—true subjects as false (i.e., the color of the text was changed from black to red). This ensured that all subjects read precisely seven false statements. In short, subjects in all conditions read a crime report about a perpetrator named Tom that contained five false statements that exacerbated the severity of his crime. Subjects also saw two critical statements that exacerbated the severity of the crime. These statements were presented to some subjects as true (the critical—true condition) and to others as false (the critical—false condition). For some subjects, the two critical statements were interrupted by a digit-search task (interrupted condition), and for other subjects they were not (uninterrupted condition).

After reading the report about Tom, all subjects read a second report about Kevin. This report was identical to one of the two reports about Kevin that was used in Experiment 1. As in Experiment 1, seven statements that extenuated the severity of Kevin's crime were presented as false. For subjects in the interrupted condition, these seven statements were interrupted, and for subjects in the uninterrupted condition they were not. This second report was included for two reasons. First, it served as a buffer between the reading of the first report about Tom and the completion of the recognition memory test. Second, it provided an opportunity to replicate the general memory results of Experiment 1. After reading the second report, all subjects completed a recognition memory test (as in Experiment 1) for the content of the two reports. On completion of the recognition memory test, subjects were fully debriefed, thanked, and dismissed.

Results and Discussion

Six subjects were omitted from all analyses because of either suspicion about the procedures or excessive difficulty reading the moving text. Of the remaining 80 subjects, 40 were required to perform the concurrent digit-search task and 40 were not. Half the subjects in each group were assigned to the critical—true condition, and the remaining subjects were assigned to the critical—false condition. Subjects who performed the digit-

search task failed to detect an average of 3.63 of the 44 digits for which they had been instruted to search. As before, false alarms were not measured.

Critical Items

Subjects read aloud a crime report that contained two critical statements that exacerbated the severity of the crime, and some subjects performed a digit-search task as they read those critical statements. For some subjects these two critical statements were ostensibly false, and for others they were ostensibly true. The specific guessing bias hypothesis predicts that interruption should increase the number of true statements recognized as true as well as the number of false statements recognized as true. Our hypothesis predicts the latter effect but not the former. As the data in Table 4 show, the specific guessing bias hypothesis received no support. Interruption did indeed increase the tendency for subjects to remember false items as true, but it actually decreased their tendency to remember true items as true (primarily by increasing their tendency to claim that true items were foils). The specific guessing bias hypothesis asserts that interruption causes subjects to guess that an unremembered item is true. The data simply belie this assertion.

Noncritical Items

As Table 5 shows, the key results of Experiment 1 were replicated. (Critical items are excluded from the table and its analyses.) Across the two reports, interruption had no discernible effects on memory for true items or for foil items, but it strongly affected memory for false items. In particular, interruption increased the likelihood that subjects would misremember false items as true (and also, to a smaller extent, as foils). Interestingly, the slight and unexpected tendency for interruption to increase the likelihood that subjects in Experiment 1 would misremember foil items as true was absent in Experiment 2.

Experiment 3

Experiments 1 and 2 provide support for the Spinozan hypothesis: When people are prevented from unbelieving the assertions they comprehend, they may act as though they believe them. Subjects did not merely recall that such assertions were said to be true (as did subjects in the studies of Gilbert et al., 1990), but they actually behaved as though they believed the assertions. As the Spinozan hypothesis predicted, interruption increased the likelihood that subjects would consider a false assertion to be true but did not decrease the likelihood that they would consider a true assertion to be false, thus further refuting alternative explanations based on the notion of guessing bias.

In Experiment 3 we attempted to extend these findings (and rule out another alternative explanation) by adapting a method used by Gilbert et al. (1990, Experiment 3) . Gilbert et al. taught subjects about an imaginary animal called a glark. After subjects had learned about glarks, they were shown a series of statements about glark morphology, social habits, and so on. Subjects were required to assess the veracity of some of the statements (i.e., determine whether they were true or false of glarks) and to read other statements as

quickly as possible. Gilbert et al. found that speed reading a false statement increased the probability that subjects would later recall the statement as true, but that assessing the veracity of a false statement had no such effect. In other words, time pressure affected memory for the veracity of a false statement in much the same way that interruption did.

In Experiment 3, subjects learned about a target named Bob and were then asked to assess or read quickly some false statements about Bob. Some subjects assessed or speed read primarily positive false statements about Bob, and some assessed or speed read primarily negative false statements about Bob. We predicted that reading these false statements under time pressure would increase the probability that subjects would like or dislike Bob, but that assessing the false statements would not. One explanation of these predicted results is that reading primarily positive or negative statements may activate positive or negative constructs, respectively (see Higgins, 1989). If time pressure increased the effects of such activation, then the predicted results could be interpreted as a mere demonstration of the well-known priming, or construct activation, effect. To rule out this possibility, subjects were also asked to report their liking for another target named Jack, about whom subjects learned some biographical facts, but about whom they had neither speed read nor assessed any statements. We reasoned that if reading likable or dislikable statements merely served to activate positive or negative constructs, then liking for both Bob and Jack should be affected by the statements. We predicted that liking for Jack would not be affected by the statements subjects assessed or speed read.

Method Overview

Subjects saw a photograph and a brief biography of two targets named Bob and Jack. Subjects then learned about a series of actions performed by Bob. Next, subjects read a set of statements that described Bob performing likable, dislikable, or neutral actions. Some subjects read these statements as quickly as possible, and other subjects assessed the veracity of the statements (i.e., they attempted to determine whether the statements did or did not describe actions that they had previously been told were performed by Bob). Finally, all subjects reported their impressions of Bob and Jack.

Subjects

One hundred sixty-one female students at the University of Texas at Austin participated to fulfill a requirement in their introductory psychology course. Only native speakers of English were eligible to participate.

Pretest and Stimulus Materials

Thirty-two female judges rated the likability of 120 actions (e.g., "Roger fidgeted a lot during class") on a series of 9-point scales that were anchored at the extremes with the words *very dislikable* (1) and *very likable* (9). The 41 highest rated actions (M = 7.39) were classified as likable actions (e.g., "Tom fed the stray cat near his house"), and the 41 lowest rated actions (M = 2.67) were classified as dislikable actions (e.g., "Henry took money from his friend's wallet"). The remaining 38 actions (M = 5.17) were classified as

neutral (e.g., "Jim said the movie was bad"). These 120 statements constituted the pool from which all likable, dislikable, and neutral statements were drawn.

Procedure

Subjects were invited to take part in an experiment on impression formation. On arrival at the laboratory, subjects were greeted by a male experimenter who ushered them to an individual cubicle where they remained for the duration of the experiment. Each cubicle contained a microcomputer with keys labeled *YES* and *NO*, a video camera mounted at head level and pointed at the subject's face, and several pieces of electronic machinery that were ostensibly connected to the video camera. Subjects were told that Bob and Jack had been introductory psychology students the previous semester and that the information presented in the experiment consisted of true facts that had been collected during their respective interviews with a clinical psychologist. Photographs of two male undergraduates were attached to a pair of brief biographical sketches (which described the target's place of birth, college major, parents' occupations, etc.), and these photographs and biographies were shown to the subjects. After reading the biographies, subjects were told that the computer would present a series of facts about one of the two targets. Subjects were told that the camera and other electronic equipment was an eye-tracking device that would record their eye movements throughout the experiment (cf. Gilbert et al., 1990). In fact, the equipment was inert. The purpose of this deception is explained shortly.

The learning phase.

The experimenter pretended to calibrate the eye-tracking device and then left the room while the computer delivered written instructions to the subject. These instructions explained that during the initial learning phase the computer would randomly select one of the targets whose pictures and biographies the subject had just seen and would present the subject with a series of facts about that target. The subject's initial task was simply to learn these facts. The instructions also explained that during a subsequent testing phase, the subject would see a series of statements about this target and would be asked to make some judgments about the statements. After the subject familiarized herself with the two biographical sketches, the computer (ostensibly randomly) selected Bob as the target, and the learning phase began.

During the learning phase, 27 statements of 4—10 words (e.g., "Bob enjoyed the Mexican food") were presented on the computer screen, 1 at a time, in a random order. Four of the statements described likable actions, 4 described dislikable actions, and the remaining 19 described neutral actions. All subjects saw the same set of statements in the learning phase, although the order of presentation was randomly determined for each subject. All statements were affirmative sentences that began with Bob's name and ended with a phase describing an action. Each statement appeared on a single line at the center of the screen for 4 s and was followed by a blank screen for 2 s. During the learning phase, each statement was displayed on two separate occasions. Pilot testing indicated

that this provided ample opportunity for subjects to learn these relatively simple statements.

Trial phase.

After the learning phase, the computer presented each subject with 58 statements about Bob. The presentation of each statement was preceded by 1,500 ms either by the signal phrase "Is the following sentence TRUE?" or by the signal phrase "SPEED READ the following sentence." Subjects were instructed that when they saw the signal phrase *TRUE,* they should read the statement that followed and assess its veracity. If the statement was true (i.e., if and only if it had appeared during the learning phase), then they should press the key marked YES. If the statement was false (i.e., if it had not appeared during the learning phase), then they should press the key marked NO. Subjects were told that any novel statements encountered in the second phase of the procedure had been generated by the experimenters and should be considered false. None of the novel statements directly contradicted any statements from the learning phase. We refer to those trials that were preceded by the signal phrase TRUE as *assessment trials.* The importance of responding both quickly and accurately on assessment trials was stressed.

Subjects were told that when they saw the signal phrase SPEED READ, they should simply read the statement that followed as quickly as possible. Subjects were told that after reading the statement they should press the space bar on the computer keyboard to indicate that they had finished reading. We refer to those trials that were preceded by the signal phrase SPEED READ as *comprehension trials.* Subjects were told that on comprehension trials, their reading speed was being measured by the computer. Ostensibly, these data would be used as baseline covariates for analyses of the assessment trials. The bogus eye-tracking device was included so that subjects would feel compelled to read these statements rather than ignore them. (This deception proved quite effective for Gilbert et al., 1990 .) The importance of rapid responding was stressed for the comprehension trials. After the subject responded on either an assessment trial or a comprehension trial, the statement was erased from the screen. The next trial began 500 ms later.

Independent manipulations.

During the trial phase, each subject was presented with 58 statements. As Table 6 shows, these statements were manipulated in two ways. First, we manipulated the primary valence of the statements: Half the subjects read more likable than dislikable statements about Bob (primary valence was positive), and the remaining subjects read more dislikable than likable statements about Bob (primary valence was negative). It is important to note that all subjects actually saw the same number of true likable and true dislikable statements, which means that a perfectly rational information processor would draw the same conclusion about Bob regardless of the primary valence of the trials.

Second, we manipulated the primary task that subjects performed. Although subjects in all conditions assessed and comprehended likable, dislikable, and neutral statements,

each subject was randomly assigned to either (a) comprehend more valanced statements than she assessed (primary task was comprehension) or (b) assess more valanced statements than she comprehended (primary task was assessment). Half the subjects in the primarily positive valence condition were assigned to comprehend 20 likable statements and assess only 2 dislikable statements; the remaining subjects in that condition were assigned to assess 20 likable statements and comprehend only 2 dislikable statements. Trials with neutral statements were added so that each subject ultimately comprehended 29 statements and assessed 29 statements, thus eliminating the possibility that subjects could predict which task they would be asked to perform on the next trial. On each trial, each subject saw a statement that was randomly selected by the computer from the pool of true and false likable, dislikable, or neutral statements.

Rating phase.

After subjects comprehended and assessed the 58 statements, each subject was asked to report how much she liked Jack. These ratings were made on a 9-point scale that was anchored at the extremes with the words *very dislikable* (1) and *very likable* (9). Liking for Jack was measured first so as to maximize the possibility of detecting affective priming. Next, subjects reported their liking for Bob on a similar scale. Then, subjects completed several additional measures of liking for Bob. Each subject rated Bob on eight 9-point trait scales (friendly—bostile, warm—cold, submissive—dominant, unconfident—confident, introverted—extraverted, suspicious—trusting, unsociable—sociable, and competitive—cooperative) that were anchored at the extremes with one of the two trait adjectives in the pair. Finally, each subject read five previously unseen statements that described likable actions (e.g., Bob told good jokes at the meeting) and five previously unseen statements that described dislikable actions (e.g., Bob bossed his younger brother around). Subjects estimated the likelihood that Bob would perform each of these actions on a series of 9-point scales that were anchored at the extremes with the words *very likely* and *very unlikely.* Finally, subjects were fully debriefed, thanked, and dismissed.

Results and Discussion Omissions of Data

Thirteen subjects were omitted from all analyses (7 for failures to understand or follow the instructions, and 6 for other reasons such as reading impairments, extreme anxiety from the camera, and previous experience with eye-tracking equipment). In addition, the data were trimmed by eliminating trials on which a subject's response time (RT) was more than three standard deviations from the grand mean. This resulted in the omission of 220 of the 15,545 observations. Finally, 16 subjects were omitted from analyses of the RT data because RTs on six or more trials exceeded the criterion level. These omissions resulted in 132 subjects approximately evenly distributed among the conditions for the RT analyses and 148 subjects for all other analyses. In no case did the omission of subjects or data influence the pattern of the findings.

The Spinozan Hypothesis: Liking for Bob

Subjects' reports of their liking for Bob were analyzed by planned comparison. As Table 7 shows, the valence of the statements that subjects read quickly (i.e., the comprehension trials) had a strong impact on their liking for Bob, F (1, 144) = 6.95, $p < .01$, whereas the valence of statements whose veracity subjects assessed (i.e., the assessment trials) did not ($F < 1$). The 18 ancillary measures of subjects' liking for Bob (the eight trait scales and the 10 actions) were combined into a liking index whose internal consistency was increased by the deletion of 8 items ($\alpha = .84$ for the 10-item index and .793 for the 18-item index). This liking index was strongly correlated with the single liking item ($r = .62$, $p < .001$), and the pattern of means on this index was virtually identical to the pattern seen on the single liking item. As with the single liking item, the valence of the statements that subjects read quickly had a strong impact on their liking for Bob, F (1, 144) = 9.75, $p < .003$, whereas the valence of statements whose veracity subjects assessed did not ($F < 1$).

Construct Activation: Liking for Jack

Is it possible that reading a series of valanced statements merely activated positive or negative constructs that then influenced subjects' responses? The data suggest not. If such constructs were activated, then one would expect them to influence judgments of Jack, and as Table 7 shows, planned comparisons performed on reports of subjects' liking for Jack showed absolutely no effects in either the comprehension or the assessment condition (both F s < 1). Clearly, the valanced statements only affected judgments of Bob (and not of Jack), and did so only when they were read quickly (and not when they were assessed).

General Discussion

The folk psychology of belief is fraught with paradox. Sometimes people talk as though they can control their beliefs: "You should believe in God" or "You must not believe those awful rumors" or "Please believe that I love you." At other times people are amused by the absurdity of such a suggestion:

> "I can't believe that!" said Alice.
>
> "Can't you?" the Queen said in a pitying tone. "Try again: draw a long breath and shut your eyes."
>
> Alice laughed. "There's no use trying," she said: "One can't believe impossible things."
>
> "I daresay you haven't had much practice," said the Queen.
>
> "When I was your age I always did it for half-an-hour a day. Why sometimes I've believed as many as six impossible things before breakfast." (Carroll, 1872/1983, p. 54)

Can anyone believe six impossible things before breakfast? The Queen's paradoxical claim plays on the mistaken assumption that belief can only follow the analysis of an assertion's plausibility. It seems absurd to insist that one can believe what one has already deemed implausible, but not so absurd to suggest that one may believe the impossible before its plausibility is calculated. This, of course, was Spinoza's (1677/1982) point. People do have the power to assent, to reject, and to suspend their judgment, but only after they have believed the information to which they have been exposed. For Descartes (1644/1984) , being skeptical meant understanding an idea but not taking the second step of believing it unless evidence justified taking that step. For Spinoza, being skeptical meant taking a second step backward (unbelieving) to correct for the uncontrollable tendency to take a first step forward (believing). Both philosophers realized that achieving true beliefs required that one subvert the natural inclinations of one's own mind; for Descartes this subversion was proactive, whereas for Spinoza it was retroactive.

The Evidence for Retroactive Doubt

We have performed a half dozen experiments to examine this notion, and each has provided support for Spinoza's (1677/1982) retroactive account rather than Descartes's (1644/1984) proactive account. In addition, a wide range of other evidence is commensurate with the Spinozan position. For example, research on attribution suggests that people often draw dispositional inferences about others and then correct those inferences with information about situational constraints on the other's action (Gilbert, Pelham, & Krull, 1988 ; see also Trope, 1986 ; Newman & Uleman, 1989). Because this correction is subsequent to and more difficult than the initial dispositional inference, people display a correspondence bias (i.e., a tendency to draw dispositional inferences from the behavior of others). If one assumes that behaviors "assert" dispositions (i.e., that a friendly behavior is taken as the equivalent of the claim "I am a friendly person"), then the Spinozan hypothesis subsumes this attributional model.

Other well-known phenomena are similarly interpretable in Spinozan terms. Research on human lie detection has consistently uncovered a truthfulness bias, that is, a tendency for people to conclude that others are telling the truth when they are not (DePaulo, Stone, & Lassiter, 1985 ; Zuckerman, DePaulo, & Rosenthal, 1981). If one assumes that verbal claims "assert" their speaker's beliefs (i.e., that the claim "Abortion is evil" is taken to imply the claim "I believe that abortion is evil"), then this truthfulness bias is just the sort of mistake that a Spinozan system should make. Research on persuasive communications has shown that distraction can increase the persuasive impact of a message (Festinger & Maccoby, 1964 ; Petty, Wells, & Brock, 1976), and the Spinozan hypothesis provides a representational account that is perfectly consistent with current high-level accounts of this effect (e.g., Chaiken, 1987 ; Petty & Cacioppo, 1986). Research on hypothesis testing has shown that people often seek information that confirms the possibilities they are entertaining (Snyder & Swann, 1978 ; Wason & Johnson-Laird, 1972). The Spinozan hypothesis suggests that people may not be inept hypothesis testers; rather, they may tend to believe the possibilities that they are asked to merely entertain, in which case a confirmatory strategy may be quite rational (see Klayman & Ha, 1987). Research on the processing of linguistic denials shows that people often develop positive beliefs in

assertions that are being denied (Wegner et al., 1985 ; Wegner, Wenzlaff, Kerker, & Beattie, 1981). This phenomenon is also explicable in Spinozan terms: A denial is both an assertion and its negation, and the act of understanding the assertion includes a belief in the very thing that is being negated or denied. (See Gilbert, 1991 , for a full discussion of each of these issues and their relation to the Spinozan hypothesis.)

In short, a variety of evidence is friendly to the Spinozan account of belief, and, as far as we know, none provides a critical challenge. The Spinozan account is both a simple and a powerful theoretical tool. It suggests that belief is first, easy, and inexorable and that doubt is retroactive, difficult, and only occasionally successful. This modest contention has the potential to explain a lot. For example, it suggests that the correspondence bias, the truthfulness bias, the distraction—persuasion effect, the denial—innuendo effect, and the hypothesis-testing bias are superficially different manifestations of a single, underlying mechanism and that researchers in each of these different areas may actually share an abiding interest in human credulity. Although more tests must be performed before the Spinozan hypothesis can be unequivocally accepted, any model that can unify so many otherwise disparate phenomena deserves serious consideration.

Skepticism and Freedom of Speech

One way to characterize the Spinozan hypothesis is that information changes people even when they do not wish to be changed. Ideas are not mere candidates for belief, but potent entities whose mere communication instantly alters the behavioral propensities of the listener. This characterization of the belief process raises some difficult social issues, as Johnson (1991, p. xi) noted in his recent discussion of neural networks:

> Every time you walk away from an encounter, your brain has been altered, sometimes permanently. The obvious but disturbing truth is that people can impose these changes against your will ... Freedom of speech is based on the old dualist notion that mind and body are separate things [but] as science continues to make the case that memories cause physical changes, the distinction between mental violence, which is protected by law, and physical violence, which is illegal, is harder to understand.

The suggestion that words cause physical changes is decidedly at odds with the Millian philosophy of the free marketplace of ideas, and if one accepts this suggestion, then restrictions on speech may seem quite reasonable. If apples forced one to taste them as one passed, it might well be decided to ban rotten fruit from the marketplace (see Schauer, 1982). The Spinozan hypothesis seems to support such an argument because it suggests that ideas reprogram the individuals who encounter them so that the individuals are prepared to act as though the ideas were true. As such, those who advocate the regulation of expression in modern society might appeal to the Spinozan perspective as a scientific justification for censorship.

Such persons would have missed an important point. It is true that the Spinozan view does not portray people as especially capable skeptics, but neither does it portray them as

relentlessly gullible automata. The hypothesis suggests that people are instantly reprogrammed by the assertions they encounter, but it also suggests that they can do something to restore themselves to their previous state. Three things are required for such "self-reprogramming" to occur (see Gilbert, 1993). First, a person must have a set of rules for the logical analysis of assertions. If one does not understand that the assertions "Smith is X" and "Smith is not X" cannot both be true of the same Smith at the same time, then no amount of cognitive work will enable the person who has heard both of these assertions to unbelieve either one of them. Second, a person must have a set of true beliefs to compare to new beliefs. To some extent, all mental systems work by coherence (i.e., they evaluate the veracity of new ideas by comparing them with old ones and measuring the fit). If the system mistakenly believes that Smith is a woman, then it cannot reject the assertion that Smith is pregnant on purely logical grounds. Finally, a person must have the desire and capacity to perform work, that is, the motivation and ability to use the rules of logical analysis to compare new and old beliefs. If people are unable or unwilling to analyze an assertion (because, e.g., they are rushed or are currently attending to some other task), then the possession of logical skills and true beliefs may not matter.

People, then, do have the potential for resisting false ideas, but this potential can only be realized when the person has (a) logical ability, (b) correct information, and (c) motivation and cognitive resources. It is interesting to note that the acquisition of logical skills and true beliefs is primarily a function of education and is therefore under the control of society, whereas motivation and cognitive capacity are either fixed or under the control of the individual. Anyone who would fashion a political position from our demonstrations of the initial credulity of people must also take into account their subsequent potential for doubt and the possibility that societies can increase that potential through education rather than making it superfluous through prior restraint. Which of these methods of belief control should a society use?

Social Costs of Belief Control

Politics, it is said, make strange bedfellows. In recent years some feminists have joined right-wing conservatives to suggest that certain sexually explicit material does not deserve protection under the First Amendment because the material advances "bad ideas" about the roles of women. Religious fundamentalists in California have successfully lobbied for restrictions on the discussion of Darwinian evolution in public school textbooks because they consider evolution a bad idea to which they do not want their children exposed. A second-grade boy in Bastrop, TX, was kept in solitary confinement for almost a year because he refused to cut off his ponytail, which school authorities considered to be a symbolic speech against conformity. Although the examples are diverse, each reflects a common and perdurable social dilemma: Either bad ideas can be stopped at their source or they can be undone at their destination, and both methods have potential costs.

The Spinozan hypothesis suggests that if a bad idea is allowed to reach its destination, the person whom it reaches may not have the logical capacity, correct information, or

cognitive resources to reject it. On the other hand, as Mill (1859/1975) noted, those who are responsible for instituting prior restraints may err in their attempts to distinguish good from bad ideas, and some good ideas may never have an opportunity to reach the person. Trying to decide which of these costs is greater is what signal detection theorists know as the problem of setting beta. Should citizens be more concerned with misses (i.e., failures to encounter good ideas) or false alarms (i.e., failures to reject bad ones)? The National Organization of Women realizes that there is a cost incurred when a magazine's right to publish nude photographs is denied, but they argue that it is not so great as the cost incurred when the derogatory message they find latent in such photographs is embraced by the illogical, uninformed, or resource depleted (i.e., false alarms are more costly than misses). The American Civil Liberties Union realizes that a cost is incurred by a society that allows irredeemable obscenity, but they argue that this cost is not so great as the cost incurred when a magazine's right to publish is denied (i.e., misses are more costly than false alarms). Can the present analysis shed any light on the choice that each society must make between prior restraint and unbelieving as methods of belief control?

Neither prior restraint nor unbelieving can be relied on to weed out all bad ideas or to sow all good ones. However, although both methods are imperfect and can thus lead to error, the errors associated with prior restraint are unique in that they are resistant to social correction. It is certainly true that in a free marketplace of ideas some illogical, uninformed, or unenergetic person will encounter a bad idea and will believe it, much as subjects in our experiments did. But surely not everyone will be susceptible to the same bad idea at the same time. Surely someone will recognize the bad idea as such and, because the marketplace is free, will be able to inject his or her own "good version" of that idea into the discourse. Imagine, for example, what might have happened if subjects in various conditions of our first experiment had been allowed to chat before they rendered a verdict. One can imagine a subject from the control condition telling a subject from the interrupted condition, "No, you've got it wrong. Tom never threatened the clerk. That was a false statement that you read while you were busy counting 5s." Although the subject from the interrupted condition might refuse to believe what her partner said, there is every reason to suspect that she would be swayed (see Gilbert & Osborne, 1989). At the very least, there exists the strong possibility that the subject who suffered from a failure to unbelieve could be "cured" by social interaction (cf. Wright & Wells, 1985).

The error of believing too much may be corrected by commerce with others, but the error of believing too little cannot. When the marketplace is underregulated, the bad ideas that are present (but that one wishes were absent) may be embraced by an individual whose wrong-headed beliefs may eventually be corrected by his or her fellows. However, when the marketplace is overregulated, the good ideas that are absent (but that one wishes were present) will never be encountered. Even if the censors are entirely benevolent (itself an unlikely assumption), the intellectual anemia that their prior restraint creates is not amenable to social correction. In short, people can potentially repair their beliefs in stupid ideas, but they cannot generate all the smart ideas that they have failed to encounter. Prior restraint is probably a more effective form of belief control than is unbelieving, but its stunning effectiveness is its most troublesome cost. The social control of belief may well be a domain in which misses have irreparable consequences that false alarms do not.

The Spinozan hypothesis suggests that we are not by nature, but we can be by artifice, skeptical consumers of information. If we allow this conceptualization of belief to replace our Cartesian folk psychology, then how shall we use it to structure our own society? Shall we pander to our initial gullibility and accept the social costs of prior restraint, realizing that some good ideas will inevitably be suppressed by the arbiters of right thinking? Or shall we deregulate the marketplace of thought and accept the costs that may accrue when people are allowed to encounter bad ideas? The answer is not an easy one, but history suggests that unless we make this decision ourselves, someone will gladly make it for us.

References

Anderson, C. A. (1982). Innoculation and counter-explanation: Debiasing techniques in the perseverence of social theories. *Social Cognition, 1*, 126-139.

Arkes, H. R., Boehm, L. E. & Xu, G. (1991). Determinants of judged validity. *Journal of Experimental Social Psychology, 27*, 576-605.

Arkes, H. R., Hacket, C. & Boehm, L. (1989). The generality of the relation between familiarity and judged validity. *Journal of Behavioral Decision Making, 2*, 81-94.

Begg, I., Armour, V. & Kerr, T. (1985). On believing what we remember. *Canadian Journal of Behavioral Science, 17*, 199-214.

Carroll, L. (1983). *Through the looking glass and what Alice found there.* (Berkeley: University of California Press. (Original work published 1872)

Chaiken, S. (1987). The heuristic model of persuasion.(In M. P. Zanna, J. M. Olson, & C. P. Herman (Eds.), *Social influence: The Ontario symposium* (Vol. 5, pp. 3—39). Hillsdale, NJ: Erlbaum.)

DePaulo, B. M., Stone, J. L. & Lassiter, G. D. (1985). Deceiving and detecting deceit.(In B. R. Schlenker (Ed.), *The self in social life* (pp. 323—370). New York: McGraw-Hill.)

Descartes, R. (1984). Principles of philosophy.(In J. Cottingham, R. Stoothoff, & D. Murdoch (Eds. and Trans.), *The philosophical writings of Descartes* (Vol. 1, pp. 193—291). Cambridge, England: Cambridge University Press. (Original work published 1644))

Festinger, L. & Maccoby, N. (1964). On resistance to persuasive communications. *Journal of Abnormal and Social Psychology, 68*, 359-366.

Gerrig, R. J. & Prentice, D. A. (1991). The representation of fictional information. *Psychological Science, 2*, 336-340.

Gilbert, D. T. (1991). How mental systems believe. *American Psychologist, 46*, 107-119.

Gilbert, D. T. (1993). The assent of man: The mental representation and control of belief.(In D. M. Wegner & J. W. Pennebaker (Eds.), *Handbook of mental control* (pp. 57—87). Englewood Cliffs, NJ: Prentice Hall.)

Gilbert, D. T., Krull, D. S. & Malone, P. S. (1990). Unbelieving the unbelievable: Some problems in the rejection of false information. *Journal of Personality and Social Psychology, 59*, 601-613.

Gilbert, D. T. & Osborne, R. E. (1989). Thinking backward: Some curable and incurable consequences of cognitive busyness. *Journal of Personality and Social Psychology, 57*, 940-949.

Gilbert, D. T., Pelham, B. W. & Krull, D. S. (1988). On cognitive busyness: When person

perceivers meet persons perceived. *Journal of Personality and Social Psychology, 54,* 733-740.
Hasher, L., Goldstein, D. & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior, 16,* 107-112.
Hastie, R. & Park, B. (1986). The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. *Psychological Review, 93,* 258-268.
Higgins, E. T. (1989). Knowledge accessibility and activation: Subjectivity and suffering from unconscious sources.(In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 75—123). New York: Guilford Press.)
Johnson, G. (1991). *In the palaces of memory.* (New York: Random House)
Klayman, J. & Ha, Y. W. (1987). Confirmation, disconfirmation, and information in hypothesis-testing. *Psychological Review, 94,* 211-228.
Lindsay, D. S. (1990). Misleading suggestions can impair eyewitness' ability to remember event details. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 1077-1083.
Mill, J. S. (1975). *On liberty.* (New York: Norton. (Original work published 1859)
Newman, L. S. & Uleman, J. S. (1989). Spontaneous trait inference.(In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 155—188). New York: Guilford Press.)
Petty, R. E. & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion.(In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 19, pp. 123—205). New York: Academic Press.)
Petty, R. E., Wells, G. L. & Brock, T. C. (1976). Distraction can enhance or reduce yielding to propaganda: Thought disruption versus effort justification. *Journal of Personality and Social Psychology, 34,* 874-884.
Rosenthal, R. & Rosnow, R. L. (1985). *Content analyses: Forced comparisons in the analysis of variance.* (Cambridge, England: Cambridge University Press)
Ross, L., Lepper, M. R. & Hubbard, M. (1975). Perseverance in self-perception and social perception: Biased attribution processes in the debriefing paradigm. *Journal of Personality and Social Psychology, 32,* 880-892.
Schauer, F. (1982). *Free speech: A philosophical enquiry.* (Cambridge, England: Cambridge University Press)
Schneider, W. & Shiffrin, R. M. (1977). Controlled and automatic human information processing: Detection, search, and attention. *Psychological Review, 84,* 1-66.
Schul, Y. & Burnstein, E. (1985). When discounting fails: Conditions under which individuals use discredited information in making a judgment. *Journal of Personality and Social Psychology, 49,* 894-903.
Smith, E. R. (1989). Procedural efficiency and on-line social judgments.(In J. Bassili (Ed.), *On-line cognition in person perception* (pp. 19—38). Hillsdale, NJ: Erlbaum.)
Smith, E. R., Stewart, T. L. & Buttram, R. T. (1992). Inferring a trait from a behavior has long-term, highly specific effects. *Journal of Personality and Social Psychology, 62,* 753-759.
Snyder, M. & Swann, W. B. (1978). Hypothesis testing processes in social interaction. *Journal of Personality and Social Psychology, 36,* 1202-1212.
Spinoza, B. (1982). *The Ethics and selected letters* (S. Feldman, Ed., and S. Shirley, Trans.).(Indianapolis, IN: Hackett. (Original work published 1677)

Trope, Y. (1986). Identification and inferential processes in dispositional attribution. *Psychological Review, 93*, 239-257.
Wason, P. C. & Johnson-Laird, P. N. (1972). *The psychology of reasoning.* (Cambridge, MA: Harvard University Press)
Wegner, D. M., Coulton, G. & Wenzlaff, R. (1985). The transparency of denial: Briefing in the debriefing paradigm. *Journal of Personality and Social Psychology, 49*, 338-346.
Wegner, D. M., Wenzlaff, R., Kerker, R. M. & Beattie, A. E. (1981). Incrimination through innuendo: Can media questions become public answers? *Journal of Personality and Social Psychology, 40*, 822-832.
Wilson, T. D. & Brekke, N. (1992). *Mental contamination and mental correction: Unwanted influences on judgments and evaluations.* (Unpublished manuscript, University of Virginia)
Wright, E. F. & Wells, G. L. (1985). Does group discussion attenuate the dispositional bias? *Journal of Applied Social Psychology, 15*, 531-546.
Wyer, R. S. & Budesheim, T. L. (1987). Person memory and judgments: The impact of information that one is told to disregard. *Journal of Personality and Social Psychology, 53*, 14-29.
Wyer, R. S. & Unverzagt, W. H. (1985). The effects of instructions to disregard information on its subsequent recall and use in making judgments. *Journal of Personality and Social Psychology, 48*, 533-549.
Zimbardo, P. G. & Lieppe, M. R. (1991). *The psychology of attitude change and social influence.* (New York: McGraw-Hill)
Zuckerman, M., DePaulo, B. M. & Rosenthal, R. (1981). Verbal and nonverbal communication of deception.(In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 14, pp. 1—59). New York: Academic Press.)

1

One of the crime reports described a male perpetrator who threatened to sexually assault a female victim. We assumed that this report might provoke different reactions from male and female subjects. Thus, to avoid theoretically irrelevant effects of gender, we used only female subjects in Experiments 1 and 2. For the sake of consistency, we also used only female subjects in Experiment 3.

2

If a subject read an exacerbating story about one perpetrator (e.g., Tom), then he or she also read an extenuating story about the other perpetrator (e.g., Kevin). As such, including both false statements (exacerbating or extenuating) and perpetrator identity (Tom or Kevin) as variables in the analysis would have caused considerable statistical complexities. Specifically, each subject would have contributed two scores–one to each of two cells on the diagonal of a factorial table–which would have required the computation of both a within-subjects and a between-subjects *F* ratio for each effect. These *F* ratios are known to be overestimates and underestimates of the true *F* ratio,

which is incomputable but which lies somewhere between the two computable *F* ratios (Rosenthal & Rosnow, 1985). Because preliminary analyses revealed no interesting effects of perpetrator identity on any of the dependent measures, and because none were expected, we collapsed the data across perpetrator identity in order to simplify the analyses presented here.

Table 1
Recommended Prison Term and Ratings of Perpetrators in Experiment 1

Measure	Nature of false statements: Extenuating	Exacerbating	Difference
Uninterrupted condition			
Recommended years in prison	6.03	7.03	−1.00
Dislikableness	5.83	6.62	−0.79*
Dangerousness	5.36	6.41	−1.05*
Benefit from counseling	6.45	5.85	0.60
Interrupted condition			
Recommended years in prison	5.83	11.15	−5.32*
Dislikableness	5.18	7.00	−1.82*
Dangerousness	4.82	7.06	−2.24*
Benefit from counseling	6.45	5.03	1.42*

* $p < .05$.

Table 2
Proportion of Statements Recognized in Experiment 1

Statement type	Condition		Difference
	Uninterrupted	Interrupted	
True statements			
Recalled as true	.93	.89	.04
Recalled as false	.03	.06	−.03
Recalled as foils	.03	.04	−.01
False statements			
Recalled as true	.23	.44	−.21*
Recalled as false	.69	.34	.35*
Recalled as foils	.09	.23	−.14*
Foils			
Recalled as true	.05	.09	−.04*
Recalled as false	.01	.05	−.04
Recalled as foils	.94	.86	.08*

* $p < .05$.

Table 3

Design and Stimulus Order for Experiment 2

	Statement								
Condition	1	2	3	4	5	6	7	8	9
Interrupted									
Critical–true	FI	FI	FI	FI	FI	**FI**	**FI**	TU	TU
Critical–false	FI	FI	FI	FI	FI	**TI**	**TI**	FU	FU
Uninterrupted									
Critical–true	FU	FU	FU	FU	FU	**FU**	**FU**	TU	TU
Critical–false	FU	FU	FU	FU	FU	**TU**	**TU**	FU	FU

Note. Statements 1–7 exacerbated the severity of the crime. Statements 8 and 9 were neutral. Statements 6 and 7 (in bold) were the critical items. All other statements in Tom's report were TU. F = false statement; T = true statement; I = interrupted statement; U = uninterrupted statement.

Table 4

Proportion of Critical Statements Recognized in Experiment 2

Critical items	Condition		Difference
	Uninterrupted	Interrupted	
Presented as true			
Recalled as true	.850	.650	.200*
Recalled as false	.050	.025	.025
Recalled as foils	.100	.325	−.225*
Presented as false			
Recalled as true	.225	.475	−.250*
Recalled as false	.750	.400	.350*
Recalled as foils	.025	.125	−.100

* $p < .05$.

Table 5
Proportion of Noncritical Statements Recognized in Experiment 2

Statement type	Condition		Difference
	Uninterrupted	Interrupted	
True			
Recalled as true	.860	.879	−.019
Recalled as false	.050	.041	.009
Recalled as foils	.072	.100	−.028
False			
Recalled as true	.292	.542	−.250*
Recalled as false	.655	.194	.461*
Recalled as foils	.055	.167	−.112*
Foils			
Recalled as true	.057	.065	−.008
Recalled as false	.008	.013	−.005
Recalled as foils	.936	.923	.013

* $p < .05$.

Table 6
Distribution of Statements in Four Conditions in Experiment 3

	Primary task					
	Comprehension			Assessment		
Statement type	Trial type: Comprehension	Trial type: Assessment	Total	Trial type: Comprehension	Trial type: Assessment	Total
	Primary valence of statements: Positive					
Likable	2T/18F	2T	4T/18F	2T	2T/18F	4T/18F
Dislikable	2T	2T	4T	2T	2T	4T
Neutral	3T/4F	16T/9F	19T/13F	16T/9F	3T/4F	19T/13F
	Primary valence of statements: Negative					
Likable	2T	2T	4T	2T	2T	4T
Dislikable	2T/18F	2T	4T/18F	2T	2T/18F	4T/18F
Neutral	3T/4F	16T/9F	19T/13F	16T/9F	3T/4F	19T/13F

Note. T = true statement; F = false statement.

Table 7
Liking for Targets in Experiment 3

Measure	Primary valence of statements: Likable	Primary valence of statements: Dislikable	Difference
Primary task: Assessment			
Liking item (Jack)	6.00	5.86	0.14
Liking item (Bob)	6.50	6.35	0.15
Liking index (Bob)	106.0	106.0	0
Primary task: Comprehension			
Liking item (Jack)	5.74	5.89	−0.15
Liking item (Bob)	6.40	5.47	0.93*
Liking index (Bob)	107.4	98.4	9.00*

Note. Unmarked differences are all $F < 1$.
* $p < .01$.