Research Article

Believe It or Not

On the Possibility of Suspending Belief

Uri Hasson,[1] Joseph P. Simmons,[2] and Alexander Todorov[2]

[1]Brain Research Imaging Center, The University of Chicago, and [2]Department of Psychology, Princeton University

ABSTRACT—*We present two experiments that cast doubt on existing evidence suggesting that it is impossible to suspend belief in a comprehended proposition. In Experiment 1, we found that interrupting the encoding of a statement's veracity decreased memory for the statement's falsity when the false version of the statement was uninformative, but not when the false version was informative. This suggests that statements that are informative when false are not represented as if they were true. In Experiment 2, participants made faster lexical decisions to words implied by preceding statements when they were told that the statements were true than when the veracity of the statements was unknown or when the statements were false. The findings suggest that comprehending a statement may not require believing it, and that it may be possible to suspend belief in comprehended propositions.*

Is it possible to suspend belief in a comprehended proposition?

In attempting to answer this question, Gilbert (1991) distinguished between the philosophies of Descartes and Spinoza. According to Descartes, people first comprehend a proposition and then later assess its veracity. Thus, in Descartes' view, comprehension precedes belief, and belief suspension is therefore possible. In contrast, according to Spinoza, comprehending a proposition requires believing it to be true. Thus, in his view, comprehension and belief occur simultaneously, and belief suspension is therefore impossible.

Although Descartes' position on this issue is more intuitively appealing, existing evidence seems to favor the Spinozan view. This evidence comes in two forms. First, there is research suggesting that truth is represented more quickly and easily than falsity. People are quicker to assess the veracity of true than false statements (e.g., Just & Carpenter, 1976), and they more easily represent true than false propositions (Johnson-Laird & Savary, 1999).

Second, Gilbert and his colleagues have demonstrated that when people are under cognitive load or time pressure, they tend to misremember and misrepresent false information as being true (Gilbert, Krull, & Malone, 1990; Gilbert, Tafarodi, & Malone, 1993). In a study ostensibly about language learning, Gilbert et al. (1990, Experiment 1) presented participants with statements about the meaning of words in the Hopi language (e.g., *A tica is a fox*). Each statement was presented on a computer screen for 8 s, and then immediately followed by the word *true* or *false* to indicate whether the proposition was an accurate or inaccurate statement about the Hopi language. Critically, on some learning trials, participants were asked to respond as quickly as possible to the sound of a tone, which was meant to interrupt the encoding of the proposition's veracity by depleting participants' cognitive resources.

After this learning phase, participants were presented with some of these propositions and were required to recall whether they were originally presented as true or as false. Results were consistent with the Spinozan hypothesis; interruption decreased the recall accuracy of false but not true statements. Participants' recall accuracy was 55% for uninterrupted true propositions, 58% for interrupted true propositions, and 55% for uninterrupted false propositions, but only 35% for interrupted false propositions. Gilbert et al. (1990, Experiment 2) replicated this effect with different materials (smiling faces that were either sincere or insincere) and a more difficult interruption task that required participants to judge whether the interrupting tone was high pitched or low pitched.

Gilbert (1991) interpreted these demonstrations as evidence for a dual-process model of belief. At Stage 1, propositions are simultaneously comprehended and believed. Subsequently, at Stage 2, people effortfully "unbelieve" false propositions. Cognitive load interrupts the two-stage process before it runs to completion, causing the process to output after Stage 1. Consequently, cognitive load causes false propositions to remain believed, and therefore to be wrongly recalled as true. This dual-process model has been used to explain a variety of effects,

Address correspondence to Uri Hasson, Brain Research Imaging Center, Biological Sciences Division, The University of Chicago, 5841 S. Maryland Ave., MC-2030, Chicago, IL 60637; e-mail: uhasson@uchicago.edu.

including source-monitoring errors (Schacter, Norman, & Koutstaal, 1997), acquiescence effects (Knowles & Condon, 1999), and anchoring effects (Chapman & Johnson, 2002).

Although extant evidence is certainly consistent with the Spinozan view, the verdict is still out on the possibility of belief suspension. First, people may encode assertions of truth more easily than assertions of falsity simply because true statements are encountered more often than false statements (Skurnik, 1998), or because false statements tend to be more grammatically complex than true statements (Clark & Chase, 1972).

Second, although cognitive load may cause people to misrepresent some false propositions as true, this effect may not generalize to all propositions, because falsity may be represented differently depending on what is said to be false. In line with this argument, previous research has demonstrated that negated statements are more easily represented to the extent to which the negations promote meaningful, informative inferences (Fillenbaum, 1966; Wason, 1965). For example, Fillenbaum (1966) and Mayo, Schul, and Burnstein (2004) found that participants misrecalled negated statements as affirmations less frequently when the negation allowed for a meaningful and informative inference or could be accommodated by a preexisting schema.

These findings suggest that the representation of negation may not always constitute a simple "tag" of the affirmation, but that whether or not a tagging system is employed may depend on the informativeness of the negation. Specifically, when the negation of a term does not offer a meaningful, informative inference, then negation might be represented as a tag (Clark & Chase, 1972). However, when negated content provides an available inference, people may not represent negation as a tag of the affirmative but, instead, make the allowable inference. In a study supporting this possibility, Fillenbaum (1966) found that the most frequent memory error for informative negations involved misrecalling a negated adjective as its antonym.

Although Fillenbaum's research focused on negation rather than falsity, it seems likely that negation and falsity are represented similarly (Gilbert, 1991; Just & Carpenter, 1976). Representing a statement as false may require applying a falsity tag to the statement if its false version is uninformative, because there is no other way to represent it. However, when a proposition's false version is informative, then the false statement may be represented in terms of what its falsity implies or suggests. Thus, a false statement may be much more easily represented, and less likely to be misrecalled as true, when knowing that it is false provides a perceiver with meaningful information.

In their research, Gilbert et al. (1990) used primarily stimuli that were uninformative when false. Learning that the statement *a tica is a fox* is false does not allow for a meaningful inference, because knowing that a *tica* is not a fox implies nothing at all; *tica* could mean an infinite number of things. As we have suggested, when statements are not meaningful when false, participants may have no choice but to represent falsity in terms of the affirmative proposition, perhaps accompanied by an effortful application of a falsity tag. This in turn may lead people to incorrectly recall false statements as true when the encoding of falsity is interrupted.[1]

However, in general, many propositions that people encounter are highly informative when false. For example, consider the proposition *George owns a television*. This is an informative proposition when false, because knowing that it is false supplies you with knowledge about George—it suggests that George is unlike most people. Thus, when learning that such propositions are false, people may be able to represent them in terms of what their falsity suggests (e.g., *George is atypical; he is the bookish type*), rather than by applying a falsity tag to the affirmative proposition.

Evidence from person perception is consistent with the possibility that this process may be completed independently of cognitive load. People are remarkably good at making spontaneous person inferences from minimal information (Todorov & Uleman, 2002). Such inferences persist despite lack of explicit memory for the information that triggered them (Carlston & Skowronski, 1994; Todorov & Uleman, 2002). More important, Todorov and Uleman (2003) have shown that spontaneous person inferences occur under conditions of cognitive load, rapid presentation of information, and shallow processing, suggesting that such inferences are automatic. Perhaps, then, when statements are informative when they are false, people will not misrecall them as true, even under time pressure or cognitive load. Such a finding would cast doubt on the claim that comprehending a statement requires believing it, and that belief suspension is impossible.

EXPERIMENT 1

In our first experiment, we used a variant of the paradigm of Gilbert et al. (1990) to examine whether people under cognitive load can refrain from believing informative false statements. In this study, we presented participants with a series of true and false propositions that were pretested to be either informative or uninformative when false. Occasionally, the presentation of the statement was followed by a tone that participants had to categorize as either high pitched or low pitched. This interruption was designed to deprive people of the time and resources necessary to assess the veracity of the statement. After encoding the statements, participants were asked to recall the veracity of each proposition. We expected interruption to interfere with the encoding of falsity only among statements that were uninformative when false.

Method

Participants

Twenty undergraduates from Princeton University participated in this study for course credit.

[1]Gilbert et al. (1993) also provided information that was uninformative when false. They explicitly selected such statements "so that subjects would not assume that a false statement . . . could be negated to create a true statement" (p. 223).

Materials

We selected 48 statements from a pool of 100 statements that we pretested on the dimensions of informativeness when true and informativeness when false. In pretesting, 18 participants rated how informative each statement would be about an anonymous 30-year-old person, if that statement was the only thing they knew about the person. They made the ratings on an 11-point scale ranging from 0 (*couldn't tell anything about the person given this information*) to 10 (*could tell a lot about the person given this information*). We chose 12 statements that were informative both when true ($M = 5.1$, $SD = 0.8$) and when false ($M = 4.8$, $SD = 0.5$; e.g., *this person is a liberal*). Twelve were informative when true ($M = 5.5$, $SD = 1.1$) but not when false ($M = 0.8$, $SD = 0.4$; e.g., *this person walks barefoot to work*). Twelve were informative when false ($M = 4.8$, $SD = 1.2$) but not when true ($M = 1.0$, $SD = 1.0$; e.g., *this person owns a television*). And 12 were uninformative both when true ($M = 1.9$, $SD = 0.8$) and when false ($M = 1.0$, $SD = 0.4$; e.g., *this person drinks tea for breakfast*). In pretesting and the experiment, all the statements were of the form *this person* followed by a descriptive clause.

Design

Participants were presented with statements that were indicated to be either true or false, and the presentation was either interrupted or not interrupted by a tone. The statements themselves were of the four types just described. The design was therefore 2 (veracity: true, false) × 2 (interruption: yes, no) × 2 (informativeness when true: high, low) × 2 (informativeness when false: high, low). The materials were rotated across four lists so that each of the statements appeared once in each list as either true or false and as either interrupted or not (i.e., each statement appeared in all four lists, but in each list it was assigned to a different experimental condition). Each participant learned 48 experimental statements.

Procedure

Learning Stage. Participants were presented with information on a computer screen. On each trial, they were presented with a person's face and, underneath it, a statement that was indicated to be either true or false of that person. Participants were told to pay attention to all of the information on the screen, because they would be asked some questions about the people they were seeing. Each face and its accompanying proposition were presented for 4 s and then removed. On interruption trials, this 4-s presentation was followed by a low or high tone that was sounded for 650 ms. Participants were instructed to press "K" when hearing the high tone and "S" when hearing the low tone. There was a 1,700-ms break between trials. The first three trials served as buffers for primacy effects, and the last three trials served as buffers for recency effects. Data from these trials were not recorded.

Test Stage. After the learning stage, participants were presented with the faces and statements that they learned earlier. Their task was to determine whether the given statement was previously indicated to be true or false of the person in the picture. To dissuade participants from giving an equal number of "true" and "false" decisions, we told them that only a subset of the original faces would be presented, and that they should therefore not be alarmed if the majority of the test faces were accompanied by either true or false attributions in the learning stage (following Gilbert et al., 1990). Each face and its associated statement were presented on the screen until participants gave their answer.

Results and Discussion

Figure 1 presents the recall accuracy for statements that were uninformative versus highly informative when false. The top panel shows that for statements that were uninformative when false, we replicated the findings of Gilbert et al. (1990): Interruption had no effect on the encoding of truth, but increased the tendency to report false statements as true. The bottom panel of Figure 1 shows the recall accuracy for statements that were highly informative when false. As predicted, for these statements, interruption had no effect on remembering a statement's falsity, and no asymmetry was found between memory for truth and memory for falsity.

Statistical analyses by subjects (F_1) and by items (F_2) corroborated these observations. We subjected the error data to a 2 (veracity: true, false) × 2 (interruption: yes, no) × 2 (informativeness when true: high, low) × 2 (informativeness when false: high, low) within-subjects analysis of variance (ANOVA). Attesting to the effectiveness of the interruption task, recall was more accurate for trials that were not followed by an interruption than for interrupted trials, $F_1(1, 19) = 8.3$, $p = .01$, $\eta^2 = .31$; $F_2(1, 44) = 7.1$, $p = .01$, $\eta^2 = .14$. Although the effect of interruption was greater for the encoding of falsity than for the encoding of truth, $F_1(1, 19) = 5.0$, $p < .05$, $\eta^2 = .22$, and $F_2(1, 44) = 3.7$, $p = .06$, $\eta^2 = .08$, this effect was driven by those statements that were uninformative when false, as indicated by a reliable three-way Veracity × Interruption × Informativeness When False interaction, $F_1(1, 19) = 10.4$, $p = .005$, $\eta^2 = .36$; $F_2(1, 44) = 6.4$, $p = .015$, $\eta^2 = .13$.

The three-way interaction reflects the fact that for items that were uninformative when false (Fig. 1, top panel), there was a reliable Veracity × Interruption interaction, $F_1(1, 19) = 15.2$, $p < .001$, $\eta^2 = .45$; $F_2(1, 23) = 8.9$, $p = .007$, $\eta^2 = .28$. In contrast, as shown in the bottom panel of Figure 1, there was no interaction for statements that were informative when false (both Fs < 1). No other effects were reliable.

These results support the idea that the effect of resource depletion on the encoding of falsity ultimately depends on whether or not the proposition's false version is informative. It seems that resource depletion affects the encoding of falsity when a proposition is uninformative when false (as discovered by Gilbert et al.,

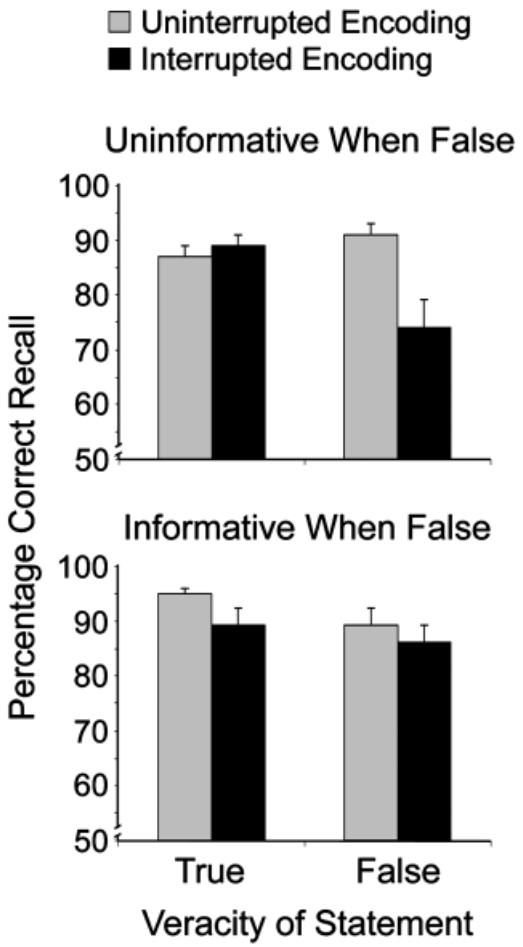

Fig. 1. Percentage of correct recall of statements' truth or falsity in Experiment 1 as a function of whether the statements were presented as true or false and whether their presentation was interrupted or uninterrupted. The top panel presents data for statements that were uninformative when false (e.g., *this person drinks tea for breakfast*), and the bottom panel presents data for statements that were informative when false (e.g., *this person is loved by family members*).

1990), but not when the statement's false version is informative. This result is consistent with the Cartesian notion that belief is not an inevitable result of comprehension. In Experiment 2, we more directly examined the question of whether people can suspend belief in a proposition.

EXPERIMENT 2

In Experiment 1, we demonstrated that interruption has no effect on the encoding of highly informative false statements. In Experiment 2, we sought to conceptually replicate this effect, while also exploring a different aspect of the Spinozan hypothesis. Gilbert (1991) noted that the Spinozan hypothesis makes at least two specific predictions. First, when people are under cognitive load or otherwise interrupted, they should represent false statements as if they were true. We investigated this hypothesis in Experiment 1. Second, "when people comprehend information whose truth value cannot be determined via assessment, they [should] act as though they believe the information to be true" (Gilbert, 1991, p. 115). We tested this prediction in Experiment 2 using a lexical decision task.

Decades of research on priming and lexical decisions indicate that people are more quickly able to identify words when those words are highly accessible in memory (e.g., Neely, 1991). Thus, when people are asked to determine whether a letter string is a word or not, they are more quickly able to do so when they have previously read a text that has made the word more accessible. In this experiment, we again presented participants with faces and statements that referred to the people depicted. After the presentation of each face-sentence pair, participants were required to rapidly make a lexical decision. There were three types of statements: True statements were indicated to be true of the person; false statements were indicated to be false of the person; and veracity-unknown statements were not indicated to be true or false until after the lexical decision task.

On critical trials, the word presented for the lexical decision task was related to either the true or the false version of the sentence preceding it. If, as the Spinozan hypothesis suggests, people represent veracity-unknown statements as if they were true, then participants would be expected to respond equally quickly to words associated with the truth of the statement (i.e., true-related words) following true statements and following veracity-unknown statements. If people do not represent veracity-unknown statements as if they were true, then participants would be expected to respond more slowly to true-related words following veracity-unknown statements than following true statements. This was our expectation. In addition, given the results of Experiment 1, we expected participants to respond more slowly to true-related words following false statements than following true statements.

Method

Participants

Sixteen undergraduates from Princeton University participated in this study for course credit.

Materials

From a pool of 80 pretested statements, we selected 24 statements that conformed to the following criteria: (a) The statements were rated as highly informative both when true and when false, and (b) the statements were strongly associated with certain words both when true and when false. For example, the statement *this person thinks that things turn out for the best* was rated as strongly associated with the terms *optimist* when true and *pessimist* when false. We also chose 24 additional statements to serve as filler materials that would be followed by legal nonwords. In this study, we used only pictures of men, because some of the statements may have different implications when referring to men and women.

Design

Statements were said to be true or false immediately, or their veracity was unknown until after the lexical decision task. In addition, each statement was followed by a lexical decision target that was related either to the meaning of the statement when true or to the meaning of the statement when false. Thus, the design was 3 (sentence veracity: true, false, unknown) × 2 (lexical decision target: true related, false related). The dependent variable was the decision latency for lexical decisions to the target words. The materials were rotated across lists so that each participant was presented with either the true- or the false-related target for each item, in one of the three priming conditions.

Procedure

On each trial, participants were presented with a photo of a person and a statement underneath the photo; they were told that their task was to learn information about the people presented to them. They were informed that in some cases they would be told immediately whether the statement was true or false of the person, and in some cases they would be told after a delay. Each photo-statement pair remained on the screen for 4 s and was immediately followed by the lexical decision letter string. This letter string remained on the screen for 250 ms, and participants were required to quickly press a button to indicate whether or not it was a word in the English language.

On half of the trials, the statements accompanying the photographs were immediately marked as either true or false of the person depicted. On the other half of the trials, participants did not learn whether the statement was true or false of the person until after they responded to the lexical decision letter string. On these veracity-unknown trials, the "true" or "false" prompt appeared after the participant's lexical decision response, and it remained on the screen for 3 s. Between trials there was a 3-s interval.

Results and Discussion

Response latencies for true-related and false-related targets for each of the three sentence types are shown in Figure 2. As predicted, sentence veracity had a reliable effect on the response times to true-related targets, $F_1(2, 30) = 6.7, p < .01, \eta^2 = .30$; $F_2(2, 46) = 4.2, p = .02, \eta^2 = .16$. The left side of Figure 2 reveals that, consistent with our prediction, lexical decisions to true-related targets were faster when the statement was said to be true ($M = 626$ ms) than when its veracity was unknown ($M = 698$ ms) or when it was false ($M = 721$ ms), $ps < .05$. For example, the word *optimist* was verified more quickly when the statement *this person thinks that things turn out for the best* was said to be true of a person than when the veracity of the statement was undetermined or when the statement was said to be false.

As shown in the right side of Figure 2, sentence veracity did not exert a reliable effect on response latencies for false-related targets ($ps > .30$). Although informative false statements are sometimes recalled in terms of the inferences they afford (Mayo et al., 2004), false statements did not speed up responding to false-related targets. This suggests that although false statements reduced the accessibility of true-related concepts, they did not immediately activate semantic content consistent with their implications (see Hasson & Glucksberg, 2005, for similar findings in an investigation of lexical negation). Perhaps people are able to immediately grasp the gist implications of informative false sentences without immediately grasping their semantic implications. As a result, people may spontaneously form a general impression (e.g., "George is a bad person") instead of automatically making an elaborative inference (e.g., "George is dishonest"; cf. McKoon & Ratcliff, 1992). We look forward to future research that investigates this intriguing possibility.

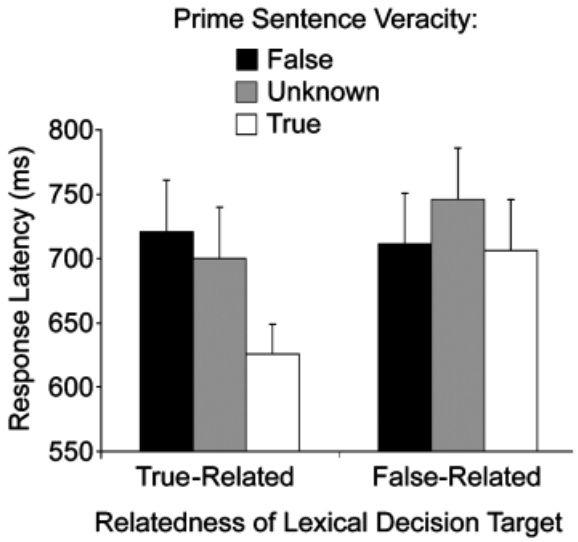

Fig. 2. Response latencies to lexical decision targets in Experiment 2 as a function of whether they were related to the truth or falsity of the preceding sentence. Results are shown separately for statements that were indicated to be true, those that were indicated to be false, and those whose veracity was unknown at the time of the lexical decision.

Most important for our central thesis is that the results of Experiment 2 contradict the Spinozan hypothesis. When the veracity of a statement is unknown or false, people do not seem to represent the statement as though they believe that it is true.

GENERAL DISCUSSION

In this article, we have presented two studies that cast doubt on the Spinozan notion that belief and comprehension occur simultaneously, and that belief suspension is therefore impossible. Previous evidence for the Spinozan hypothesis comes from studies demonstrating that cognitive load or interruption causes people to misremember explicitly false statements as being true (Gilbert et al., 1990). In Experiment 1, we replicated this effect, but only when the false versions of statements were uninforma-

tive. When the false versions of the propositions were informative such that they conveyed meaningful information about the people that they described, interruption did not impair memory. This result suggests that comprehension may not necessitate belief. When false information is informative, people can remember that the information is false—even when elaboration of that information is interrupted. It seems that extant evidence in favor of the Spinozan hypothesis does not generalize to statements that convey meaningful information when the statements are false.

In Experiment 2, we conceptually replicated this effect using a lexical decision task. When propositions were denoted as false, or when the veracity of the propositions was unknown, participants responded more slowly to true-related words than they did when the propositions were denoted as true. Contrary to the Spinozan hypothesis, this result suggests that people do not necessarily treat false propositions or veracity-unknown propositions as if they were true.

The relation between comprehension and belief is a complex one. One possibility, examined here, is that comprehension cannot exist without belief. Some researchers (e.g., Recanati, 1997) have suggested that people can believe propositions that they do not completely understand (i.e., whose representation is undetermined). Furthermore, comprehension may not be a necessary precursor for disbelief. People might not understand a proposition such as *there is not one place in which an electron exists at any given time*, yet disbelieve the proposition exactly because they cannot comprehend it. Further research on the relation between the processes of belief and comprehension may very well demonstrate that comprehension and belief are two independent processes.

Acknowledgments—We thank Dan Gilbert for his comments on a previous version of this article.

REFERENCES

Carlston, D.E., & Skowronski, J.J. (1994). Saving in the relearning of trait information as evidence for spontaneous inference generation. *Journal of Personality and Social Psychology*, *66*, 840–856.

Chapman, G.B., & Johnson, E.J. (2002). Incorporating the irrelevant: Anchors in judgments of belief and value. In T. Gilovich & D. Griffin (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 120–138). New York: Cambridge University Press.

Clark, H.H., & Chase, W.G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology*, *3*, 472–517.

Fillenbaum, S. (1966). Memory for gist: Some relevant variables. *Language and Speech*, *9*, 217–227.

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

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., Tafarodi, R.W., & Malone, P.S. (1993). You can't not believe everything you read. *Journal of Personality and Social Psychology*, *65*, 221–233.

Hasson, U., & Glucksberg, S. (2005). *Does comprehending negation entail affirmation?* Manuscript submitted for publication.

Johnson-Laird, P.N., & Savary, F. (1999). Illusory inferences: A novel class of erroneous deductions. *Cognition*, *71*, 191–229.

Just, M.A., & Carpenter, P.A. (1976). Relation between comprehending and remembering some complex sentences. *Memory & Cognition*, *4*, 318–322.

Knowles, E.S., & Condon, C.A. (1999). Why people say "yes": A dual-process theory of acquiescence. *Journal of Personality and Social Psychology*, *77*, 379–386.

Mayo, R., Schul, Y., & Burnstein, E. (2004). "I am not guilty" versus "I am innocent": Successful negation may depend on the schema used for its encoding. *Journal of Experimental Social Psychology*, *40*, 433–449.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, *99*, 440–466.

Neely, J.H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G.W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264–336). Hillsdale, NJ: Erlbaum.

Recanati, F. (1997). Can we believe what we do not understand? *Mind and Language*, *12*, 84–100.

Schacter, D., Norman, K., & Koutstaal, W. (1997). The recovered memories debate: A cognitive neuroscience perspective. In C. Martin (Ed.), *Recovered memories and false memories: Debates in psychology* (pp. 63–99). New York: Oxford University Press.

Skurnik, I.W. (1998). *Metacognition and the illusion of truth*. Unpublished doctoral dissertation, Princeton University, Princeton, NJ.

Todorov, A., & Uleman, J.S. (2002). Spontaneous trait inferences are bound to actors' faces: Evidence from a false recognition paradigm. *Journal of Personality and Social Psychology*, *83*, 1051–1065.

Todorov, A., & Uleman, J.S. (2003). The efficiency of binding spontaneous trait inferences to actors' faces. *Journal of Experimental Social Psychology*, *39*, 549–562.

Wason, P.C. (1965). The contexts of plausible denial. *Journal of Verbal Learning and Verbal Behavior*, *4*, 7–11.

(Received 6/17/04; Revision accepted 8/31/04)