Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Memory and Language 52 (2005) 535–550
Journal of Memory and Language
www.elsevier.com/locate/jml

Social metacognitive judgments: The role of retrieval-induced forgetting in person memory and impressions☆

Benjamin C. Storm*, Elizabeth Ligon Bjork, Robert A. Bjork

Department of Psychology, University of California, 1285 Franz Hall, Box 951563, Los Angeles, CA 90095-1593, USA

Received 17 September 2004, revision received 14 January 2005
Available online 23 February 2005

Abstract

Retrieval is a "memory modifier" (R.A. Bjork, 1975) in two senses: information retrieved from memory becomes more recallable; and, other information associated with the same cues becomes less recallable. Over time, therefore, retrieval processes act to update, shape, and sometimes distort our memories, including, presumably, memories involved in our feelings toward others. This paper reports several experiments exploring the role of retrieval-induced forgetting in maintaining and modifying impressions; more specifically, whether we monitor or use the retrieval accessibility of memories about other people to make and adjust our impressions of them. Participants who practiced retrieving neutral characteristics associated with target individuals exhibited retrieval-induced forgetting of positive or negative characteristics also associated with such individuals, but such impaired retrieval did not affect subsequent likeability ratings. Findings suggest that impairing retrieval access to specific information does not, in of itself, lead to changes in judgments originally based on such information.
© 2005 Elsevier Inc. All rights reserved.

Keywords: Social metacognition; Retrieval inhibition; Retrieval-induced forgetting; Impression formation; Person memory; Negativity bias

Introduction

Memory dynamics, such as the effects of primacy and recency, have long been of interest to social psychologists and other researchers interested in topics such as judgment and decision making, impression formation, and person memory. In general, what is recallable about a person, a group, or an interaction would seem, intuitively, to constitute the raw material for social metacognitions of various types, such as our judgments and impressions of others. And, further, it seems reasonable—as with other metacognitive assessments, such as

☆ The authors thank the following undergraduate research assistants who worked on this project: Nancy Sin, Shirine Sateei, Megan Vu, Jennifer Tolkan, Lucero Barajas, Estela Barajas, Michelle Yoo, Farid Holakouee, Andrea Vander Hyde, Gabriela Pocasangre, Blair Tasker, John F. Nestojko, Sarah Smith, Sejal Patel, Susy Hassan, Roya Rahmanian, Jennifer Storm, and Virgie Mamaradlo, and, especially Leila Kushan who worked on a pilot version of Experiment 2. We also thank the members of Cogfog for their valuable input at various stages of the present research, the attendees at the Social Meta-Cognitive Conference (University of Heidelberg, July, 2004) for comments regarding aspects of the present research presented at that conference and, also, Janet Metcalfe, Collen Kelley, and two anonymous reviewers for their thoughtful comments and suggestions.

* Corresponding author. Fax: +1 310 206 5895.
E-mail address: storm@ucla.edu (B.C. Storm).

0749-596X/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2005.01.008

feelings of knowing—that when we alter what is recallable about others, we should also alter our judgments and impressions about them. The goal of the present research is to examine whether *retrieval-induced forgetting* (Anderson, Bjork, & Bjork, 1994; Anderson & Spellman, 1995), which refers to the loss of access to information in memory when other information associated with the same cues is repeatedly recalled, can both alter our memory representations of others and, as a consequence, also alter the impressions we previously formed on the basis of those representations. Such a finding would be consistent with interpretations of feeling-of-knowing and other judgments in terms of accessibility (e.g., Koriat, 1995, 1998), where it has been found that the fluency of access to a target item influences the judgments about that item.

One adaptive function of retrieval-induced forgetting, among other types of "goal-directed" forgetting (Bjork, Bjork, & Anderson, 1998), involves finding and selecting correct information from memory, such as a given person's name. As we search our memories for such desired information, we need to forget or suppress the recall of closely related, but incorrect, information—such as a similar, but wrong, name. The retrieval-induced forgetting that results when one name is retrieved in preference to others plays a critical role in the resolution of such competition. In the context of social judgments and person memory, however, it seems likely that retrieval-induced forgetting sometimes biases and distorts our memories of other individuals and groups of individuals. That is, because it is possible to retrieve only a subset of information about another person at any given time, we repeatedly engage in what might be thought of as naturally occurring selective retrieval practice when we think about other people in our lives. Over time, the repeated recollection of a subset of information about a given person, or group of individuals, may lead to the diminished ability to recall other relevant information about that person or group. By continuously retrieving positive (or negative) information about another person or group, for example, retrieval access to the associated negative (or positive) information will suffer, and, potentially, alter our impression of that person or group. From the standpoint of social metacognition, the extent to which such selective retrieval does, indeed, alter our impressions has important implications for understanding how changes in memory are monitored, and whether social judgments are generally adjusted in response to changes in retrieval access.

The retrieval-induced forgetting paradigm

As the name implies, *retrieval-induced forgetting* describes the detrimental effect that retrieving some items from memory has on the later recall of competing items. The paradigm commonly used to study this phenomenon involves three distinct phases: a study phase, a retrieval-practice phase, and a final testing phase. In the paradigm as initially implemented (Anderson et al., 1994), participants learn a series of word pairs in the form of a category and an exemplar of that category (e.g., *Fruit Orange*, *Metal Silver*, *Fruit Lemon*, and *Metal Bronze*). Following such study, participants retrieve half of the exemplars from half of the studied categories via guided retrieval practice in which category names together with a letter stem (e.g., *Fruit Or____*) are presented and participants are instructed to retrieve the previously studied exemplar that fits the combined category-stem cue. Exemplars receiving retrieval practice are referred to as Rp+ items (e.g., orange); exemplars within the practiced category, but not practiced themselves, are referred to as Rp− items (e.g., lemon); and unpracticed exemplars from unpracticed categories are referred to as Nrp items (e.g., silver, bronze) and serve as baseline controls. Following an intermediary distracter phase, participants are given a final category-cued recall test for all members of each category.

Not surprisingly, on such a final test, recall of the Rp+ items is significantly facilitated relative to that of the Nrp baseline items. Of more interest, however, recall of the Rp− items—that is, recall of the unpracticed exemplars from the practiced categories—is significantly impaired relative to that of the Nrp baseline items. As has been argued by Anderson et al. (1994), this impaired recall of the Rp− items results from inhibitory processes—or retrieval inhibition—initiated during the preceding retrieval practice. More specifically, it is thought that during the attempt to retrieve a particular Rp+ item (e.g., *Orange* in response to *Fruit Or____*) other associated exemplars (e.g., *Lemon*) compete for recall as well. Thus, to retrieve the target exemplar (*Orange*), in the presence of this competition, the competing exemplar (*Lemon*) must be selected against, resulting in its becoming temporarily inhibited or forgotten.

In the retrieval-induced forgetting paradigm, therefore, the inhibition of the Rp− items does not occur as a result of an intentional attempt to forget such items—as might, for example, be the case in the impaired retrieval of to-be-forgotten items in the intentional or directed-forgetting paradigm (e.g., Bjork et al., 1998; Bjork, 1989) in which participants attempt to forget certain items in response to an explicit cue. In contrast, in the retrieval-induced forgetting paradigm, the cue to forget is implicit to the task and the forgetting or inhibition of the Rp− items occurs as an unintentional byproduct of the participant's attempts to retrieve the Rp+ items.

A growing body of evidence suggests that retrieval-induced forgetting occurs in a wide variety of contexts. Not only are semantic categories subject to the effects of retrieval inhibition, so are visual scenes and event narratives in eye-witness memory (Saunders & MacLeod, 2002; Shaw, Bjork, & Handal, 1995), visuospatial

materials (Ciranni & Shimamura, 1999), examination materials (Macrae & MacLeod, 1999), and autobiographical memories (Barnier, Hung, & Conway, 2004).

Relevant prior findings

Given this apparent generality, it seems highly likely that retrieval-induced forgetting influences our social memories, and, if so, might alter our relevant social meta-cognitions as well. There is no actual evidence, however, that retrieval-induced forgetting plays such a role, or that it does not play such a role. There is evidence, that traits and characteristics of individuals are subject to retrieval-induced forgetting, but whether such forgetting alters existing impressions or attitudes is unknown. Macrae and MacLeod (1999; Experiment 1) for example, found that participants, after being presented with 10 positive traits for each of two target individuals (Bill or John), exhibited retrieval-induced forgetting of half the traits for a given individual when induced to practice retrieving that person's other traits. In Macrae and MacLeod's research, however, the point was not to assess the participant's evaluative reactions to Bill or John, nor how such reactions might have been changed by retrieval-induced forgetting.

In another relevant study, Dunn and Spellman (2003) presented participants with stereotypical and individuating traits of hypothetical individuals, described as an Asian woman or a mother. Examples of such traits are *studious* vs. *wealthy* for an Asian woman and *patient* vs. *musical* for a mother. Practice retrieving traits of one type resulted in a subsequent reduction in the ability to recall traits of the other type, but Dunn and Spellman also did not examine whether such retrieval-induced forgetting changed participants' impressions of the individuals in question.

What also remains unclear is whether negative information is affected differently by retrieval-induced forgetting than positive information. Attrill and MacLeod (2004) examined the extent to which retrieval-induced forgetting affects memory for valenced information, whether relevant to one's self or to another person. Pairs of participants, when meeting each other for the first time, were asked to choose positive and negative traits that best characterize themselves and the other person. After selective retrieval practice, evidence of retrieval-induced forgetting for positive aspects of other-relevant information was obtained, but not for negative aspects, and no evidence of retrieval-induced forgetting for self-referential information of either type was found.

Attrill and MacLeod's results, however, need to be interpreted in light of the particular interpersonal dynamics that were present in their research. Being forced to describe a stranger using negative traits may well have been an emotionally distinctive experience, leading to the type of processing normally associated with self-referential material (in typical real-world conditions, the negative traits that we use to characterize another person are not disclosed). To the extent that was the case, their findings can be viewed as largely consistent with those of MacLeod and Roseveare (2002), who found that information highly relevant to oneself, or encoded in terms of the self, is often protected from retrieval inhibition. Describing someone else in terms of positive traits, on the other hand, might not have involved distinctive processing to the same extent.

Paradoxically, although research on impression formation has demonstrated a negativity bias in the perception of other people (e.g., Fiske, 1980; Rozin & Rozyman, 2001), existing research on retrieval-induced forgetting suggests that such a negativity bias in person memory might facilitate forgetting of negative information due to the particular dynamics of competitive retrieval. Using category–exemplar pairs that varied in associate strength, for example, Anderson et al. (1994) demonstrated that strong exemplars are subject to more retrieval-induced forgetting than weak exemplars. This unintuitive finding appears attributable to the dynamics of response competition. During retrieval practice (e.g., trying to retrieve *Orange* to the cue *Fruit Or____*), competing responses need to be selected against or inhibited. The exemplars that are inhibited, therefore, are those exemplars most strongly associated with the category cue—the ones, so to speak, that immediately come to mind. That is, when asked to recall a fruit, exemplars like apples and bananas are what come to mind, not kiwis and pomegranates, and they, then, are the most subject to suppression. Given such results, it may be the stronger and more salient negative information about an individual that is most susceptible to retrieval-induced forgetting.

In summary, then, the question as to whether negative traits associated with another person are or are not susceptible to retrieval-induced forgetting is in need of further examination. Analogous to there being necessary conditions for producing retrieval inhibition in laboratory experiments using category–exemplar materials (demonstrated in the research of Anderson & McCulloch, 1999; Smith & Hunt, 2000; Macrae & Roseveare, 2002), there are likely to be necessary conditions for inhibiting what we know about other people in the social world. Furthermore, positive and negative information may introduce different competitive retrieval dynamics and may, therefore, be influenced in different ways by retrieval-induced forgetting, which might, in turn, have important implications for our understanding of person memory and impression formation.

Finally, and importantly, it may also be the case that changes in the overt recallability of traits or attributes simply do not alter existing impressions. In research on directed forgetting in social and legal contexts, for

example, instructions to disregard, suppress, ignore, or discount information has typically been shown *not* to affect impression or judgments based on that information (for summaries, see Johnson, 1994, and multiple chapters in Golding & MacLeod, 1998). Bjork (1998) has argued that the conditions in such experiments often differ from those in memory research—in such contexts, for example, it is not necessarily in one's interest to forget when told to do so (e.g., when a juror is instructed to disregard the testimony of a witness). It is also plausible, however, that once formed, impressions are abstractions that remain largely unchanged across changes in the explicit recall of the original information (e.g., Klein, Loftus, & Plog, 1992). Priming effects, for example, on implicit-memory tests such as word-fragment completion, have been shown to be undiminished by the instruction to forget the items that are the source of such effects, even though the subsequent recall of those same items is impaired by such an instruction (see, e.g., Bjork & Bjork, 1996).

Evidence from case studies designed to assess the self-knowledge of amnesic and autistic patients also suggest that impressions may not be malleable. Klein, Loftus, and Kihlstrom (1996) reported the case of a 21-year-old college student who, after a severe head injury, could not remember her entire first year of college. Despite this loss of episodic access, she was nonetheless still accurate in providing relevant trait-descriptiveness ratings of herself. Similarly, Klein, Chan, and Loftus (1999) reported the case of a high functioning autistic individual who demonstrated accurate trait-knowledge despite having a severely diminished ability to retrieve trait-relevant personal experiences from memory. These findings suggest a functional distinction between the metacognitive impressions we have of ourselves and the accessibility of relevant information about ourselves. Thus, they would appear to contradict the accessibility account of metacognitive judgments (Koriat, 1995, 1998) and the hypothesis addressed in the present research: namely, that as what we can retrieve about others is altered, our impressions of them should be altered in a corresponding fashion. In contrast, such a distinction would imply that reductions in retrieval access to positive or negative information about another person might not alter our previously formed impressions of them.

Experiment 1a

In Experiment 1a, we focused on the following three questions: (a) is valenced information about other people vulnerable to inhibition via retrieval-induced forgetting in the same way that exemplars of semantic categories are; (b) are negative and positive traits encoded about another person affected differently by retrieval-induced forgetting; and (c) can the selective retrieval-induced forgetting of positive or negative attributes of a given person change an earlier impression we have formed of that person. Our basic procedure involved having participants study traits associated with different individuals, who were represented by photos as well as names; then engage in retrieval practice for some of the traits (Rp+ items), but not all of the traits (Rp− items), for some of the people but not for others (Nrp items); and, finally, judge the likeability of each of the studied individuals.

We had two general hypotheses. First, we expected that both negative and positive information would be inhibited, but not to the same extent. In consideration of prior findings (Attrill & MacLeod, 2004), and the pervasiveness of the negativity bias (Rozin & Rozyman, 2001; Ybarra, 2001), it seemed possible that negative information could be less vulnerable to retrieval inhibition. Based on the competitive retrieval dynamics presumed to underlie retrieval-induced forgetting, however, the opposite prediction can also be made. That is, when trying to recall other information, the negativity bias should exacerbate the interference between negative information and one's target of retrieval, thereby making the need to suppress the negative information all the more necessary.

Second, we expected that retrieval-induced forgetting, to the extent that it impaired access to either positive or negative information about a given studied individual, would shift the affective impression of that person accordingly. This second prediction was made cautiously, however. As discussed earlier, research in the directed-forgetting paradigm (e.g., Bjork & Bjork, 1996, 2003; Johnson, 1994) suggests that impressions might persist despite the inhibition of the memories on which they were based. Also, and more broadly, it seemed possible that a dissociation between the implicit feelings associated with explicit memories, and the accessibility of those explicit memories per se, might be the way in which our impressions in person perception overcome explicit forgetting. Said differently, people do not necessarily have to remember everything about someone as long as they can remember how they should act and feel around that person.

Method

Participants

A total of 96 undergraduate students from the University of California, Los Angeles, averaging 19.9 years of age, participated in the experiment for class credit in an introductory psychology class. The sample consisted of 32 men and 64 women, all of whom were fluent English speakers. Equal proportions of male and female participants were randomly assigned to each counterbalanced experimental condition.

Materials

In constructing target stimuli, 28 pictures were taken of 14 male and 14 female college students around the UCLA campus. The pictures, which were taken from approximately five feet away and showed the shoulders and head, were pasted onto index cards and labeled with a commonly used name (e.g., Dave). As part of another experiment, 64 participants judged 14 of these individuals (either all the males or all the females) on a scale of one to seven on the following five dimensions: likeability, honesty, intelligence, morality, and attractiveness. Four males and four females were then selected as being rated most similarly. In particular, the selected targets were rated highly similar in terms of likeability and attractiveness. Any individual who was actually known by two or more of the 64 pilot participants was removed from consideration.

Anderson's (1968) likeability norms were used to generate three types of traits (20 neutral, 10 negative, and 10 positive) to be associated with the targets (see Appendix A for a total listing). The 20 neutral traits selected, such as *quiet*, *shy*, and *blunt*, were divided into four lists with average ratings between 297 and 303 on the likeability scale; the 10 positive traits selected, such as *loyal*, *friendly*, and *helpful*, were divided into two lists of five traits with average likeability ratings of 514 and 517, respectively; and the 10 negative traits selected, such as *phony*, *selfish*, and *nosey*, were divided into two lists of five traits with average likeability ratings of 77 and 79, respectively. To ensure that participants generated the correct traits during selective retrieval practice, no two selected traits began with the same two letters.

Traits were then assigned to targets to create four types of targets (either all male or all female) to be presented to each participant. Two of those targets were characterized by five neutral and five positive traits and two were characterized by five neutral traits and five negative traits. One of the positive targets and one of the negative targets was selected to receive retrieval practice on their neutral traits (making them +Rp and −Rp targets, respectively), and the traits of the other two targets received no retrieval practice (making them +Nrp and −Nrp targets, respectively).

Each list of five neutral traits was attached to one target person. Thus, every participant was exposed to and, when applicable, received retrieval practice for the same five traits for each target. The two lists of five positive or five negative traits, respectively, that were added to make up the 10-trait list for each target, however, were counterbalanced across subjects and targets. Thus, across subjects, when assessing whether negative and positive traits had suffered retrieval-induced forgetting as a consequence of the retrieval practice given to neutral traits, the neutral traits practiced and the sets of negative and positive traits, respectively, were the same. During the study phase, the neutral and valenced traits to be associated with a given target were presented in an interleaved and semi-random order.

Procedure

To create a somewhat realistic impression-formation context, a cover story was employed. Participants were told that they were going to learn traits describing four individuals, and in the process, they should form impressions about them. Participants were asked to assume that other participants had interacted with these individuals in a prior study and had used the to-be-learned traits to describe these individuals. The participants were told that they, too, might be asked to play a game with one of the individuals at the end of the experiment, and that the more they remembered about that individual, the better they would do in the game. Participants were assured that the target individuals would not know anything about them, what they remembered, or how they gave their ratings. The purpose of this cover story was to encourage the participants to treat the targets like real individuals with whom they would potentially have to interact in the future.

After the instructions, the experiment began and consisted of three main phases for each target individual in turn: a study phase, a retrieval-practice phase, and a testing phase, all within the 10-min block allocated to a given target. This blocking of targets and their associated traits was adopted because a previous pilot study, using the more traditional study procedure in which all targets and their associated traits are presented in an interleaved format, proved problematic: participants had difficulty remembering what traits were associated with which person.

The study phase involved the participants reading 10 word pairs that appeared underneath a target's picture on the computer screen (e.g., John: *average*; John: *loyal*) and remained in view for 5 s each. For each participant, the 10 word pairs consisted of five neutral and five positive traits for two of the targets, and five neutral and five negative traits for two other targets. Next, in the retrieval-practice phase, half of the subjects (for any particular target) received retrieval practice for the neutral traits (e.g., John: *av*____) three times, while the other half of the subjects received retrieval practice for an unrelated category (e.g., fruit: *or*____), thus creating Rp− and Nrp conditions, respectively, for the unpracticed valenced traits. Positive and negative traits are considered to be Rp− items when the target's neutral traits are practiced, and are considered Nrp items when some other category is practiced. In both conditions, however, the target's image remained on the screen and the participant had 5 s to write down the missing word. After an unrelated five-min distracter task (for which the target's face was removed), a category-cued recall test was administered. The target's face and name were placed on the screen, and the participant had 60 s to recall as many traits describing the target as possible.

To assess the participant's impression of each target, two judgment tasks were inserted into the above procedure. First, immediately following the study phase, the participants were asked to evaluate the target on six dimensions: likeability, honesty, intelligence, morality, competence, and attractiveness. Rather than have participants report a number score, however, they crossed a 15.4 cm line to make a judgment about that individual. If judging likeability, for example, the far left end of the line was labeled, "very unlikable," while the far right-end of the line was labeled, "very likeable." This method was used to minimize the subject's ability to remember his or her previous rating. As likeability was the measure of interest, we placed it first to remove any possible ordering effects. The additional rating judgments were included to limit further the participant's ability to remember his or her previous likeability rating. The second rating task was administered in the same way after the 5 m distraction, and immediately before the recall task.

After the above process was repeated for each of the four targets, the participants were informed that they would not have to meet any of the individuals, but that they did have a couple tasks left to complete. Following a five-min delay, during which they completed a word-search puzzle, the participants were asked to rate each of the individuals once again and then to engage in a surprise, final cued-recall task. These final rating and recall tasks were added to examine whether any retrieval-induced forgetting effects would persist after a delay and whether impression changes might become more salient when the four studied individuals had to be rated on the same line, which might encourage comparisons among the individuals in order, in effect, to rank order them on the likeability scale. As it turned out, in Experiment 1a—and in Experiments 1b and 2 as well—the pattern of findings on the final rating and recall tests did not differ from those on the earlier tests in any significant way. The results of the final rating and recall tests are not, therefore, reported below, but are available from the authors.

Results and discussion

Neutral traits

Retrieval-practice performance. The mean success rate for retrieving neutral traits during the retrieval-practice phase was .87 ($SD = .14$).

Recall performance. Correct recall of neutral traits when they had been Rp+ items vs. Nrp items, when they had been presented with positive vs. negative traits, and when they had been associated with male vs. female targets was analyzed in a 2 (Rp+ vs. Nrp) × 2 (positive vs. negative) × 2 (male target vs. female target) mixed-design, Analysis of Variance (ANOVA), with sex of target being the only between subjects variable. As expected, a significant main effect of retrieval practice was obtained, with .73 ($SE = .02$) of the neutral traits receiving retrieval practice (Rp+ traits) being correctly recalled, whereas only .46 ($SE = .02$) of the neutral traits not receiving retrieval practice (Nrp traits) being correctly recalled, $F(1,94) = 158.00$, $p < .001$. This advantage was similar across valence conditions: when the neutral traits were presented with positive traits, retrieval practice increased the proportion correctly recalled to .75 ($SE = .02$) from .49 ($SE = .02$), and, when the same neutral traits were presented with negative traits, retrieval practice increased the proportion correctly recalled to .71 ($SE = .02$) from .43 ($SE = .03$).

There was, however, a significant main effect on the recall of neutral traits depending on the nature of the other traits with which they had been presented, $F(1,94) = 5.89$, $p < .05$. Specifically, when paired with positive traits, the proportion of neutral traits correctly recalled was .62 ($SE = .02$), and when paired with negative traits, the proportion correctly recalled was .57 ($SE = .02$). Perhaps neutral traits, when combined with positive traits, seem relatively more negative and were therefore better recalled owing to the negativity bias. There was no difference in the recall performance of neutral traits describing male and female targets, and no interactions were significant.

Recall of positive and negative traits

The mean correct cued-recall proportions for positive and negative traits as a function of whether they had been associated with a male or female target and whether the target's neutral traits had or had not been given retrieval practice (i.e., their status as Rp− or Nrp items) on the initial recall test are shown in Table 1. The data summarized in this table were analyzed in a 2 (positive vs. negative) × 2 (male target vs. female target) × 2

Table 1
Mean correct trait-recall proportions (and standard errors) in relation to trait valence, item type (Rp− or Nrp), and sex of the target in Experiment 1a

Sex of the target	Trait valence	
Item type	Positive	Negative
Male		
Rp−	.38 (.03)	.40 (.03)
Nrp	.43 (.03)	.52 (.03)
Difference	*−.05*	*−.12*
Female		
Rp−	.30 (.03)	.41 (.03)
Nrp	.43 (.03)	.51 (.03)
Difference	*−.13*	*−.10*
Combined		
Rp−	.34 (.02)	.40 (.02)
Nrp	.43 (.02)	.51 (.02)
Difference	*−.09*	*−.11*

(Rp− vs. Nrp), mixed-design ANOVA, with sex of target the only between-subjects variable. The participant's sex was included as an additional variable in a separate ANOVA, but no significant differences between male and female participants were obtained; thus, all results reported in the present section combine across male and female participants.

A significant retrieval-induced forgetting effect was observed, with Rp− items (whether positive or negative traits and whether associated with a male or female target) being recalled significantly less well ($M = .37$; $SE = .02$) than their Nrp counterparts ($M = .47$; $SE = .02$), $F(1,94) = 19.96$, $p < .001$. In other words, the selective retrieval of neutral traits significantly impaired the participants' ability to recall competing positive or negative traits on a later cued recall test.

A negativity bias was found in the recall of traits, as participants recalled significantly more negative traits ($M = .46$; $SE = .02$) than positive traits ($M = .39$; $SE = .02$), $F(1,94) = 12.02$, $p < .001$. Negative traits, however, were not found to be affected differently (M for Rp− = .40, $SE = .02$ vs. M for Nrp = .51, $SE = .02$) by retrieval-induced forgetting than positive traits (M for Rp− = .34, $SE = .02$ vs. M for Nrp = .43, $SE = .02$), $F(1,94) < 1$. And, although the means in Table 1 suggest the possibility that positive traits are less inhibited than negative traits when associated to a male target (top two rows) whereas the opposite is true for female targets (next two rows), this potentially intriguing three-way interaction between valence, item type, and sex of the target did not reach significance, $F(1,94) = 1.56$, $p = .21$.

Impression ratings

The mean likeability ratings obtained as a function of (a) whether the target was male or female, (b) had been presented as a positive or negative target, (c) whether the associated neutral traits had or had not been given retrieval practice, and (d) the point in the experiment at which they were made (1st: immediately after the study phase, or 2nd: immediately before the cued recall test) are shown in Table 2. The resulting ANOVA was a 2 (male vs. female target) × 2 (positive vs. negative target) × 2 (Rp− vs. Nrp) × 2 (points at which ratings were made), mixed-design, with sex of target being the only between-subjects variable.

Importantly, a significant main effect for valence confirmed that participants formed different impressions of individuals depending on whether they were associated with five positive or five negative traits, with positive targets receiving overall likeability ratings of 10.7 cm ($SE = .3$ cm) and negative targets receiving overall likeability ratings of 4.9 cm ($SE = .4$ cm), $F(1,94) = 420.88$, $p < .001$. No interaction was found between target valence and sex of the target, $F(1,94) < 1$.

When the neutral traits associated with positive targets were given retrieval practice, likeability ratings dropped from a mean of 10.7 cm to 10.5 cm (SEs = .2 cm) *from the first to the second rating.* When no retrieval practice was given to their associated neutral traits, however, the likeability ratings for the same positive targets maintained a constant mean of 10.8 cm (SEs = .2 cm) across those same ratings. A 2 (Rp− vs. Nrp) × 2 (before-retrieval-practice vs. after-retrieval-practice) within-subjects ANOVA confirmed that the reductions in likeability ratings did not significantly interact with the retrieval-practice manipulation between the first and second ratings.

When neutral traits associated with negative targets were given retrieval practice, likeability ratings went from a mean of 4.8 to 5.1 cm (SEs = .2 cm) across the times ratings were gathered. When no retrieval practice was given to their neutral traits, however, the likeability ratings for the same negative targets made the same change, from a mean of 4.8 to 5.1 cm (SEs = .2 cm). Once again, an ANOVA confirmed that these changes in impression did not significantly interact with the retrieval-practice manipulation between the first and second ratings.

In keeping with the expectation that when the memory on which an impression was formed becomes altered, the impression based on that memory should be consistently altered, we would expect likeability ratings for negative targets to increase when their associated negative traits were made less accessible by the retrieval practice of their associated neutral traits; and, conversely, that the likeability ratings for positive targets would decrease when their positive traits were made less accessible by the retrieval practice of their associated

Table 2
Mean (and standard errors) of likeability ratings (in cm on the rating scale) in relation to sex of target, trait valence, and item type (Rp− or Nrp) in Experiment 1a

Trait valence	Time of rating	
Item type	1st (Pre-Rp)	2nd (Post-Rp)
Male targets		
Positive		
Rp−	10.9 (.3)	10.4 (.3)
Nrp	10.5 (.3)	10.6 (.3)
Negative		
Rp−	4.4 (.3)	4.7 (.3)
Nrp	4.8 (.3)	5.1 (.3)
Female targets		
Positive		
Rp−	10.5 (.3)	10.4 (.3)
Nrp	11.5 (.3)	11.1 (.3)
Negative		
Rp−	5.2 (.4)	5.4 (.4)
Nrp	5.0 (.3)	5.1 (.3)

Note. Maximum rating = 15.4.

neutral traits. Inconsistent with this expectation, however, there was no change in the likeability of either negative or positive targets as a function of whether the neutral traits associated with those targets did or did not receive retrieval practice—even though such practice did reduce participants' ability to recall a given target's positive or negative traits.

The pattern of results calls into question whether there is *any* relationship between the ability to retrieve traits about an individual and one's impression of that individual. It might be argued that the level of retrieval-induced forgetting in Experiment 1a was too weak to induce a significant change in impression, but such an argument predicts that there should be a significant correlation between number of valenced traits recalled and the corresponding likeability rating—and no such correlation was observed. For targets associated with negative traits, the Pearson's correlation between number of negative traits recalled and likeability rating was not significant, ($r = .02$, $p > .5$), nor was the corresponding correlation for targets associated with positive traits ($r = -.03$, $p > .5$).

Experiment 1b

In Experiment 1b, we created more realistic individuals by associating both positive and negative traits with each target—with the idea that such a design might also be more sensitive to detecting differences in the inhibitory susceptibility of positive and negative information. If negative traits are more likely to come to mind than are positive traits during the retrieval of neutral traits, they might then, according to the suppression account of retrieval-induced forgetting (Anderson et al., 1994), suffer more retrieval-induced forgetting. By describing a target individual with both positive and negative traits, in addition to neutral traits, we hoped to increase our ability to detect any such difference.

Method

Participants

Sixty-four undergraduate students at the University of California, Los Angeles, averaging 19.6 years of age, participated in the experiment for credit in an introductory psychology course. The sample consisted of 32 males and 32 females, and equal numbers of males and females were randomly assigned to all experimental conditions.

Materials

The materials used in Experiment 1a were used again in Experiment 1b, but with one adjustment. To create four lists consisting of four neutral, three positive, and three negative traits, four neutral traits were removed and replaced by two new positive traits and two new negative traits. The 10 traits in each list were placed in a semi-random order. All materials were once again counterbalanced across subjects so that each list, list order, and target individual was equally likely to be used in a given experimental condition. This counterbalancing also ensured that each counterbalanced condition was equally likely to be placed in a given order.

Procedure

As in Experiment 1a, participants were given a cover story informing them that they would learn about and form impressions of four different individuals. Rather than learning about four males or four females, however, each participant learned about two males and two females, one in each of four separate 10-min blocks, with each block—as in Experiment 1a—consisting of a learning phase, a retrieval-practice phase, a distracter, and a cued recall phase.

Results and discussion

Neutral traits

Retrieval-practice performance. The mean success rate for retrieving neutral traits during the retrieval-practice phase was .85 ($SD = .20$).

Recall performance. Correct recall of neutral traits when they had been Rp+ items vs. Nrp items and when presented with a male vs. a female target was analyzed using a 2 (Rp+ vs. Nrp) × 2 (male target vs. female target) within-subjects ANOVA. As expected, a significant effect of retrieval practice was obtained, $F(1, 63) = 111.16$, $p < .001$. The proportion of neutral traits recalled correctly was greater when given retrieval practice ($M = .71$, $SE = .02$) than when not given retrieval practice ($M = .42$, $SE = .02$). There were no significant differences between recall performance for male and female targets.

Recall of positive and negative traits

The mean correct cued-recall proportions for positive and negative traits as a function of whether they had been associated with a male or female target and whether the target's neutral traits had or had not been given retrieval practice (i.e., their status as Rp− or Nrp items) on the cued-recall test in Experiment 1b are shown in Table 3. The data summarized in this table were analyzed in a 2 (positive vs. negative) × 2 (male target vs. female target) × 2 (Rp− vs. Nrp) repeated-measures ANOVA. The participant's sex was included as an additional variable in a separate ANOVA, but no significant differences between male and female participants were obtained; thus, all results reported in the present section combine across male and female participants.

Table 3
Mean correct trait-recall proportions (and standard errors) in relation to trait valence, item type (Rp− or Nrp), and sex of the target in Experiment 1b

Sex of the target	Trait valence	
Item type	Positive	Negative
Male		
Rp−	.30 (.03)	.40 (.03)
Nrp	.39 (.04)	.46 (.03)
Difference	*−.09*	*−.06*
Female		
Rp−	.34 (.03)	.42 (.04)
Nrp	.40 (.03)	.48 (.03)
Difference	*−.06*	*−.06*
Combined		
Rp−	.32 (.02)	.41 (.03)
Nrp	.40 (.03)	.47 (.03)
Difference	*−.08*	*−.06*

Importantly, the selective retrieval practice of neutral traits significantly inhibited the subsequent recall of competing positive and negative traits, with Rp− items being recalled significantly less well ($M = .36$, $SE = .02$) than the comparable Nrp items ($M = .44$, $SE = .02$), $F(1, 63) = 7.585$, $p < .01$. There was a negativity bias, with significantly more negative traits ($M = .44$; $SE = .02$) being recalled than positive traits ($M = .36$, $SE = .02$), $F(1, 63) = 12.92$, $p < .001$. And as in Experiment 1a, we found no difference in the retrieval-induced forgetting effect for positive and negative traits, $F(1, 63) < 1$. Positive traits suffered essentially the same inhibition (M for Rp− = .32, $SE = .02$ vs. M for Nrp = .40, $SE = .03$) as did negative traits (M for Rp− = .41, $SE = .03$ vs. M for Nrp = .47, $SE = .03$). As indicated in Table 3, a significant three-way interaction between trait valence, item type, and sex of the target was not found $F(1, 63) < 1$.

Impression ratings

Likeability ratings were examined in relation to (a) whether the target was male or female, (b) whether the target's neutral traits had or had not been given retrieval practice, and (c) with respect to the point in the experiment in which the likeability rating was given (immediately after the study phase or immediately before the initial recall test) using a 2 (male vs. female target) × 2 (Rp− vs. Nrp) × 2 (Time 1 vs. Time 2) repeated-measures ANOVA. As expected from the results of Experiment 1a, no significant differences or interactions were found. As also shown by the ANOVA and post hoc t test comparisons, the ratings were similar ($M = 7.7$, $SD = 2.3$) across all experimental conditions.

The results from both Experiments 1a and 1b demonstrate that retrieval-induced forgetting can play a role in what characteristics we are later able to recall about other individuals. In a context where participants were presumably motivated to remember and form impressions about others, their ability to recall valanced descriptive traits about others was inhibited by prior selective retrieval practice of competing neutral traits. Moreover, despite theoretical reasons to expect a different amount of inhibition for positive and negative traits, they were impaired equally. Furthermore, as shown by the results of Experiment 1b, both types of traits were subject to retrieval-induced forgetting when simultaneously associated with the same individual. The retrieval-induced forgetting of either positive or negative traits, however, failed to alter a target's perceived likeability, and the participants' likeability judgments bore little if any relationship to the number of valenced traits that they could explicitly remember.

Experiment 2

The finding in Experiments 1a and 1b that the recall of valenced traits was impaired by the prior retrieval practice of neutral traits associated with the same target individual was anticipated. That there was no corresponding change in participants' impressions of such individuals was, at least to some extent, a surprise. In Experiment 2, we tested the replicability and generality of both findings by using what seemed to be a richer and more natural way of describing the individuals about whom impressions were to be formed. In addition, one other aspect of the procedure was changed in order to examine an alternative explanation of why impression ratings did not change with the selective inhibition of either positive or negative traits: namely, that participants attempted to remember and then to repeat their earlier ratings.

In Experiments 1a and 1b, targets were described via lists of traits, which may not be an adequate representation of the type of information individuals use to form their impressions of others. Very rarely, for instance, do people learn a list of traits that are said to describe another person. More commonly, people observe various behaviors on the part of another person, which they encode and evaluate in the process of forming an impression of that individual. Forming an impression of someone on the basis of traits attributed to that person vs. on the basis of behaviors attributed to that person might, perhaps, involve different processes.

According to Semin and Fiedler's (1988, 1991) linguistic category model, for example, trait and behavioral descriptions lie at opposite ends of an abstractness-concreteness continuum and reflect fundamentally different ways in which social phenomena can be represented. Concrete behavioral descriptions are believed to be easier to remember and to provoke more attentional resources than their abstract trait counterparts. And, indeed, in a recent study by Ter Doest, Semin, and Sherman (2002), it was found that, in comparison to traits,

more attention was paid to behavioral stimuli, which were recalled more effectively and were susceptible to fewer intrusions. The way in which judgments and impressions are made also appears to differ in relation to whether trait or behavioral descriptions are learned. When evaluative impressions are formed on the basis of traits, for example, they tend to have a stronger relationship with prior expectations than when formed on the basis of behaviors (Ter Doest et al., 2002).

Experiment 2 was designed to explore how the dynamics of retrieval-induced forgetting and impression formation might differ when impressions were to be formed on the basis of concrete behavioral information. Is it possible, for example, that participants, when exposed to concrete behavioral descriptions about others, spontaneously inter-relate the behaviors in order to create an aggregate impression of that person to a greater extent than they do with trait information. If so, such integrative processing could largely eliminate the occurrence of retrieval-induced forgetting (cf. Anderson & McCulloch, 1999). It is also possible, however, that the more deeply processed behavioral descriptions would provide significantly more interference during selective retrieval practice, thereby intensifying the need to suppress them.

Additionally, we wondered if the expected influence of retrieval-induced forgetting on the valence of one's impressions might be more likely to emerge with these new materials. That is, we thought the new materials—in addition to adding more real-world validity to the impression-formation task—would introduce new social-cognitive dynamics that might alter the way in which participants formed and maintained their impressions. Perhaps, for example, by learning behaviors as opposed to traits, participants will be less locked into their initially formed impressions. Because they now must infer the meaning of the behaviors themselves, the effect of retrieval-induced forgetting may have a greater capacity to induce changes in the valence of their impressions.

Finally, in Experiment 2, we made a small, but potentially important procedural change. In Experiments 1a and 1b, it is possible that participants—in spite of the steps we took to make it difficult for them to do so—tried to remember and to repeat their initial impression ratings when asked to rate an individual target again. To rule out this relatively uninteresting explanation as to why impressions of others were not altered by the observed retrieval-induced forgetting, participants in Experiment 2 were only required to rate a given target *after* completing the retrieval-practice phase.

Method

Participants

Forty undergraduate students (30 females and 10 males) from the University of California, Los Angeles, averaging 20.0 years of age, received either credit in an introductory psychology course, or monetary payment, for their participation.

Materials

The same targets used in Experiments 1a and 1b were employed in Experiment 2, and they were presented in the same manner, except that a behavioral description together with the target's name, rather than the target's name and a trait, appeared under the target's picture. The new behavioral descriptions, however, were written so that their meaning, or the inference about the target that could be drawn from them, would be roughly equivalent in terms of valence to the traits used in Experiments 1a and 1b. The traits, *cautious*, *rude*, and *kind*, for example, led to the behavioral descriptions: *looks both ways before crossing the street, interrupts during conversations, and runs errands for his elderly aunt*, respectively. As in Experiment 1a, each target was characterized either by five neutral and five positive behaviors or by five neutral and five negative behaviors. As illustrated in the three examples, the behavioral descriptions were fairly brief, ranging from three to eight words in length. A list of all the behavior descriptions used, categorized by valence, is presented in Appendix B.

Procedure

The procedure for Experiment 2 was very similar to Experiment 1a except for the following changes. First, in the retrieval-practice phase, rather than retrieving a trait to a name and a letter-stem cue (e.g., John: av____), participants retrieved two missing words from a behavioral description (e.g., John: rides his bi____ to sc____). Second, participants made only one likeability rating, which they did after completing the retrieval-practice phase. Third, in order to accommodate the more complex stimulus materials, the timing used for various parts of the experiment was lengthened appropriately. In the retrieval-practice phase, participants were given 7 s to generate and write down the two missing words from the behavioral description. And, in the cued-recall task, participants were given 90 s (instead of 60) to write down as many of the behavioral descriptions associated with that target as they could remember.

Results and discussion

Neutral behaviors

Retrieval-practice performance. The mean success rate for retrieving the missing words from the neutral behavioral descriptions during the retrieval-practice phase was .98 ($SD = .07$).

Recall performance. Correct recall of neutral behavioral descriptions when they had been Rp+ items vs. Nrp items, when they had been presented with positive vs.

negative behaviors, and when they had been associated with male vs. female targets was analyzed in a 2 (Rp+ vs. Nrp) × 2 (positive vs. negative) × 2 (male target vs. female target) mixed design ANOVA, with sex of the target as the only between-subjects variable. As expected, a significant effect of retrieval practice was obtained, with the proportion of neutral behaviors correctly recalled being .74 ($SE = .02$) when given retrieval practice vs. only .51 ($SE = .03$) when not given retrieval practice, $F(1,38) = 45.81$, $p < .001$, and this advantage was similar whether the neutral traits had been paired with positive or negative behaviors. Specifically, when neutral behaviors were presented with positive behaviors, retrieval practice increased the proportion recalled from .53 ($SE = .03$) to .77 ($SE = .03$), and when the same neutral behaviors were presented with negative behaviors, retrieval practice increased the proportion recalled from .48 ($SE = .04$) to .72 ($SE = .03$). The main effect of more neutral items being recalled when paired with positive behaviors approached significance, $F(1,38) = 3.31$, $p = .08$, but no differences were found in the recall of neutral traits describing male vs. female targets.

Recall of positive and negative behaviors

The mean correct recall proportions for behaviors on the cued-recall test as a function of whether they were positive or negative behaviors, had been associated with a male or female target, and whether the target's neutral behavioral descriptions had or had not been given retrieval practice are shown in Table 4. To evaluate these results a 2 (positive vs. negative) × 2 (Rp− vs. Nrp) × 2 (male target vs. female target) mixed design ANOVA, with sex of target being the only between-subjects variable, was performed.

As in Experiments 1a and 1b, a significant retrieval-induced forgetting effect was observed, with Rp− items being recalled significantly less well ($M = .44$, $SE = .03$) than their Nrp counterparts ($M = .53$, $SE = .02$), $F(1,38) = 10.43$, $p < .005$. Thus, as with traits, the selective retrieval of neutral behaviors, as opposed to some other arbitrary category, significantly reduced the participants' ability to recall competing positive or negative behaviors in a later cued recall test. As for differences in the extent of retrieval-induced forgetting for positive vs. negative behaviors, no interaction was found. Although negative behavioral descriptions appeared to suffer more inhibition (M for Rp− $= .46$, $SE = .04$ vs. M for Nrp $= .59$, $SE = .03$) than positive behavioral descriptions (M for Rp− $= .43$, $SE = .04$ vs. M for Nrp $= .47$, $SE = .03$), this difference was not significant, $F(1,38) = 1.65$, $p = .21$.

Table 4
Mean correct behavior-recall proportions (and standard errors) in relation to behavior valence, item type (Rp− or Nrp), and sex of the target in Experiment 2

Sex of the target	Behavior valence	
Item type	Positive	Negative
Male		
Rp−	.44 (.05)	.49 (.05)
Nrp	.54 (.04)	.53 (.04)
Difference	*−.10*	*−.04*
Female		
Rp−	.41 (.05)	.43 (.05)
Nrp	.39 (.04)	.64 (.04)
Difference	*−.02*	*−.21*
Combined		
Rp−	.43 (.04)	.46 (.04)
Nrp	.47 (.03)	.59 (.03)
Difference	*−.04*	*−.13*

To assess the possibility of a negativity bias in the recall of behaviors, independent of the effects of retrieval practice, recall performance for behaviors when appearing as Nrp items was separately analyzed in a 2 (positive vs. negative) × 2 (male target vs. female target) mixed ANOVA. As in Experiments la and lb, a significant negativity bias in recall was observed, with participants recalling .47 ($SE = .03$) of the positive behaviors compared to .59 ($SE = .03$) of the negative behaviors, $F(1,38) = 9.32$, $p < .005$. Furthermore, this negativity bias significantly interacted with the sex of the target, $F(1,38) = 10.94$, $p < .005$. Whereas there was almost no difference in the recall of positive ($M = .54$, $SE = .04$) and negative ($M = .53$, $SE = .04$) behaviors for male targets, there was a large difference in the recall of positive ($M = .39$, $SE = .04$) and negative ($M = .64$, $SE = .04$) behaviors for female targets.

Unlike in Experiments 1a and 1b, where the negativity bias was approximately the same for male and female targets, the occurrence of this strong interaction in Experiment 2 allowed us to examine the predictions regarding retrieval-induced forgetting and valence. Because stronger items presumably need to be suppressed more than do weaker items, we might expect negative behaviors associated with a female target to suffer more retrieval-induced forgetting than their positive counterparts. More specifically, we might expect negative behaviors (which we assume to be stronger based on their greater recall in the Nrp condition when associated with female targets) to suffer more retrieval inhibition than positive behaviors during retrieval practice of neutral behaviors for female targets. In contrast, because there was no evidence of a negativity bias for the male targets (given the equivalent recall of positive and negative traits associated to males in the Nrp condition), there should therefore be no difference in the extent of the retrieval-induced forgetting.

The relevant data with respect to this prediction are presented in the first four rows of Table 4, and the apparent interaction between behavior valence, item type, and sex of the target indicated there was found to

be significant, $F(1,38) = 4.79$, $p < .05$. When recalling behaviors about females, negative descriptions (M for Rp− = .43 vs. M for Nrp = .64) were inhibited to a much greater extent than positive behaviors (M for Rp− = .41 vs. M for Nrp = .39). When recalling behaviors about males, on the other hand, the inhibitory effect was actually stronger for the positive behaviors (M for Rp− = .44 vs. M for Nrp = .54) than for the negative behaviors (M for Rp− = .49 vs. M for Nrp = .53).

This interaction is evident in terms of proportion reduction scores as well. For female targets whose neutral traits were practiced compared to female targets whose neutral traits were not practiced, negative descriptors were recalled 33% less while positive descriptors were actually recalled 5% more. For male targets whose neutral traits were practiced compared to male targets whose neutral traits were not practiced, negative descriptors were recalled 8% less and positive descriptors were recalled 19% less. We return to this finding and offer a possible explanation in General discussion.

Impression ratings

The mean likeability ratings, which were taken before the cued-recall test, are shown in Table 5 in relation to whether the target was male or female, a positive or negative target, and the target's neutral behavioral descriptions had or had not been given retrieval practice. The data summarized in Table 5 were analyzed using a 2 (male vs. female target) × 2 (positive vs. negative target) × 2 (Rp− vs. Nrp) mixed-design ANOVA, with sex of target being the only between-subjects variable.

Importantly, a main effect for valence confirmed that participants formed different impressions of individuals depending on whether they were associated with five positive or five negative behavioral descriptions, with positive targets receiving overall likeability ratings of 11.0 cm ($SE = .3$ cm) and negative targets receiving overall likeability ratings of 5.5 cm ($SE = .3$ cm), $F(1,38) = 164.47$, $p < .001$. Furthermore, the similarity of these mean likeability ratings for positive vs. negative targets to those obtained in Experiment 1a indicates that we were successful in our attempt to create behavioral descriptions that approximated the valance of the traits. No interaction was found between target valence and sex of the target, $F(1,38) < 1$.

Table 5
Mean (and standard errors) of likeability ratings (in cm on the rating scale) in relation to sex of target, behavior valence, and item type (Rp− or Nrp) in Experiment 2

Sex of the target	Behavior valence	
Item type	Positive	Negative
Male		
Rp−	11.2 (.5)	5.8 (.6)
Nrp	10.7 (.5)	5.4 (.6)
Female		
Rp−	10.7 (.6)	5.5 (.6)
Nrp	11.3 (.6)	5.5 (.6)
Combined		
Rp−	11.0 (.6)	5.6 (.6)
Nrp	11.0 (.6)	5.5 (.6)

Note. Maximum rating = 15.4.

For the ratings task, the mean likeability ratings of the positive targets were identical in the Nrp and Rp conditions (11.0 cm), and the mean likeability ratings of the negative targets were similar in the Nrp (5.5 cm) and Rp conditions (5.6 cm) as well. The ANOVA, as well as planned t test comparisons, confirmed that neither of the above differences, nor their interactions with the sex of the target, was statistically significant.

Despite our use of what we thought should be a more realistic and compelling type of information about individuals (descriptive behaviors as opposed to abstract traits), we still found no evidence in support of the hypothesis that likeability impressions can be affected by inhibiting the retrieval strength of valenced items in memory. Furthermore, the failure to find an effect cannot be attributed to the participant's remembering and trying to repeat his or her previous rating, which was a possible concern in Experiments 1a and 1b. The rating tasks were only administered in Experiment 2 *after* the retrieval-practice task occurred.

An intriguing and unexpected result observed in Experiment 2, however, was the finding of a strong negativity bias for female targets, and the complete absence of one for male targets. Furthermore, this interaction created ideal conditions to examine the comparative inhibitory susceptibility of positive and negative information. Because participants' memories for negative behaviors describing female targets were apparently much stronger than their memories for positive behaviors, the need to suppress the former during the retrieval-practice phase would have been great. Because there was no such negativity bias for male targets, however, we should not expect any differences in the amount of suppression required for negative vs. positive memories. These expectations almost perfectly matched the data. Whereas retrieval-induced forgetting occurred at roughly equivalent rates for male-referent positive and negative behaviors, it was much stronger for negative female-referent behaviors than for positive female-referent behaviors.

General discussion

Two aspects of the present findings merit further discussion. First, in all three experiments, and central to the goals of this research, we observed a dissociation between the maintenance of impressions formed earlier and changes in the explicit recall of specific attributes

that were the original basis for such impressions. Making valenced traits, whether positive or negative, less recallable did not modify subsequent likeability ratings of associated individuals. Second, there was a global negativity bias, which took the form of better recall of negative than of positive traits, and there were possible interactions of that bias with the gender of targets. In what follows, we first discuss negativity biases and the role they may play in retrieval-induced forgetting, and we conclude with a discussion of the apparent insensitivity of metacognitive judgments to changes in the accessibility of the information on which they were originally based.

Before proceeding with these discussions, however, it is probably important to point out that although we did not employ the cue-independent procedure in the present experiments, and thus it could be argued did not independently establish that inhibition is involved in the production of our findings, we believe—as findings by Anderson and Spellman (1995) and Ciranni and Shimamura (1999), using cue-independent procedures suggest—that retrieval-induced forgetting does involve the active suppression of certain items in memory. Regardless of why or how certain memories became less recallable, however, the finding that impaired retrieval access and changes in impression are dissociated remains.

Valence and retrieval inhibition

As cited earlier, research on impression formation has demonstrate a negativity bias in the perception of other people (e.g., Fiske, 1980; Rozin & Rozyman, 2001). Negative events tend to be perceived as more potent and tend to dominate evaluations when combined with more positive events. The present research provides additional evidence of a negativity bias in person memory: in all three experiments, participants recalled significantly more negative traits and behavioral descriptions than positive traits and behavioral descriptions. Despite this difference, negative information about other people appears to suffer from retrieval-induced forgetting to the same extent that positive information does. Furthermore, the results of the present Experiment 2 demonstrated that a stronger negativity bias could actually exacerbate the inhibition of negative information. The suppression account of retrieval-induced forgetting, however, which we favor, may help to explain these counterintuitive findings.

According to the suppression account of retrieval-induced forgetting, it is not weak items that are the most inhibited by the retrieval of competing items, but, rather, strong items. As mentioned earlier, for example, Anderson et al. (1994) found that it is the taxonomically strong exemplars of a given category that are most subject to retrieval-induced forgetting and that, in fact, weak exemplars seem spared of any such effects. Their explanation of this finding is that strong exemplars—being the items that want to intrude, so to speak, during the retrieval of other items—are the items that need to be suppressed.

From the standpoint of that argument, if it is the case that negative information is more salient and memorable than positive information (similar to taxonomically strong vs. taxonomically weak exemplars), negative information might be expected to provide more interference than positive information when neutral information is being recalled, and thus suffer more retrieval-induced forgetting. It should follow, then, that the extent to which negative items will suffer from retrieval-induced forgetting, compared to positive items, should correspond to the strength of the negativity bias.

The results of Experiment 2 provided an opportunity to test such a prediction. As it happened, participants, after learning 10 (five neutral and five negative or positive) behavioral descriptions about a target individual, remembered far more negative than positive behavioral descriptions for female targets, but showed no such bias for male targets. Given that participants were inclined to remember more negative than positive information about female targets, we should, according to the suppression hypothesis, expect the negative information to compete more during retrieval practice of the neutral information than the positive information and, thus, to suffer more retrieval-induced forgetting. In contrast, because negative and positive information were equally well remembered for male targets, no such effect should be observed.

The results support that prediction. Whereas both positive and negative behavioral descriptions about male targets suffered from retrieval-induced forgetting to roughly the same extent, only the negative behavioral descriptions about female targets were inhibited. Furthermore, the degree of retrieval-induced forgetting observed for negative behaviors attributed to female targets far exceeded that observed for negative behaviors attributed to male targets. Said differently, and paradoxically, participants' greater ability to remember negative information about female targets gave rise to a greater inhibition of that information.

It is not clear, though, *why* the above interaction was obtained—that is, why there should be such a difference in how participants remember behavioral descriptions about men and women. One possibility arises from research on person memory suggesting that we are biased to remember inconsistent information better than consistent information about others (e.g. Hastie & Kumar, 1979; Macrae, Bodenhausen, Schloerscheidt, & Milne, 1999). If, in the present situation, participants generally expected the pleasant looking female targets to be good people, the negative behaviors they were said to have performed may have, in effect, been perceived as highly inconsistent behaviors. In contrast, participants may have been more inclined to accept that the male

targets (although equally pleasant looking) had performed the given negative behaviors, thus perceiving them as relatively less inconsistent. Thus, rather than observing a negativity bias for female targets but not for male targets, our results may instead reflect a bias for remembering information inconsistent with one's expectations.

Viewed in this way, our failure to see such an extreme negativity bias in the results of Experiments 1a and 1b might also be understandable. To illustrate, in a recent study by Ter Doest et al.'s (2002), a bias for remembering inconsistent items was found when the stimuli were of low abstraction (e.g., behaviors), but not when the stimuli were of high abstraction (e.g., traits). If, similarly, in the present research, it was only when behavioral descriptions were used, as opposed to traits, that the negative information was perceived as being highly inconsistent for the female targets, then the interaction between sex of the target and valence of the information observed in Experiment 2 and the lack of such an interaction in Experiments 1a and 1b makes sense.

Impressions and retrieval inhibition

The present findings suggest that inhibiting the explicit recall of positive or negative information about others does not, at least immediately, alter our previously formed impressions of those individuals. Despite significant retrieval-induced forgetting of either positive or negative trait or behavioral information, targets were rated as likeable, or unlikable, as they would have been without such an intervention. In addition, and equally surprising, we found no correlation between the likeability rating assigned to a given target and the number of that target's valenced traits, positive or negative, participants were able to recall. The nature of the traits or behaviors used to characterize a given target heavily influenced subsequent likeability ratings of that individual, but such ratings did not vary significantly with how many and which of those traits or behaviors were later recallable.

That impressions, once formed, remain largely intact even if the specific traits upon which those impressions were formed are no longer recallable may well be beneficial. Viewed from a practical perspective, the persistence of such feelings could serve as a guide to acting appropriately towards another person with whom we have interacted in the past, without our having to recollect explicitly every positive or negative thing we have ever learned about that person. From a theoretical perspective, the persistence of our impressions or feelings about others suggests that our impressions, once formed, might be mediated by distinct memory processes. As mentioned earlier, prior research has suggested that retrieval access to specific behaviors or exemplars may be independent from more abstract trait judgments (e.g., Klein et al., 1999, 1992). Additionally, research on directed forgetting (e.g., Bjork & Bjork, 1996, 2003) has suggested that information that is no longer recallable can continue to influence judgments and behaviors.

Viewed from another perspective, however, the fact that impressions cannot be altered, at least not easily, by retrieval-induced forgetting is disappointing. Considerable research on intentional stereotype suppression has suggested that people have little ability to control or suppress their stereotypical thinking about others (e.g., Wyer, Sherman, & Stroessner, 2000). Furthermore, research by Macrae, Bodenhausen, Milne, and Ford (1997) indicates that not only is it difficult to intentionally forget information consistent with our stereotypes, even if successful, it can lead to a later heightened accessibility of such information. Dunn and Spellman's (2003) finding, however, suggested—at least prior to the present findings—that retrieval-induced forgetting might constitute a more effective mechanism for altering undesirable stereotypes. To the extent that the recall of individuating information about an individual can impair the subsequent recall of stereotypical information about that person, as demonstrated by Dunn and Spellman, one might have expected impressions of that person to be altered as well, but, unfortunately, this optimistic possibility was not upheld by the present results.

Whatever role retrieval-induced forgetting may play in updating and shaping our memories, it appears that it may not, by virtue of altering what is accessible in memory, also alter global impressions and attitudes. One might have hoped that the indirect suppression of negative memories via retrieval practice of other memories might be a more effective way of changing attitudes; unfortunately, however, that hope was not supported by the present results.

Concluding comment

It is possible, of course, that longer term and repeated retrieval of selected attributes might, eventually, alter impressions and attitudes. It is possible, too, that asking participants to engage in tasks that require actively using the accessible information about a target—such as generating descriptive sentences about the individual in question—might eventually alter impressions. In short, by inducing more frequent, active, or explicit use of the information that remains accessible, it may yet be possible to demonstrate that impressions can eventually be altered by retrieval-induced forgetting. Based on the present findings, however, our conclusion is that impressions, once formed, tend to persist—and that only new information, not selective access to old information, has the power to alter existing impressions and attitudes. It may also be the case, as suggested by the work of Mitchell and his colleagues (e.g., Mitchell, Macrae, & Banaji, 2004; Mitchell, Mason, Macrae, & Banaji, in press), that social judgments are special in certain ways and are not

governed by the same brain and cognitive dynamics that underlie non-social judgments.

Appendix A. Traits used in Experiment 1a

Neutral traits: forgetful, average, cautious, lucky, emotional; talkative, bold, excitable, choosy, critical; quiet, persuasive, dependent, blunt, shy; timid, proud, aggressive, moderate, bashful.

Positive traits: trustful, humorous, kind, loyal, helpful; clever, happy, friendly, honest, gentle.

Negative traits: phony, conceited, rude, nosey, jealous; mean greedy, selfish, annoying, shallow.

Appendix B. Behavioral descriptions used in Experiment 2

Neutral descriptions: always misplaces his keys, rides his bike to school, looks both ways before crossing the street, entered a raffle and won a new car, cries when he watches sad movies; spends hours on the phone, expresses his opinions in class, takes the same route to work everyday, buys only a certain type of jeans, gets nervous before every test; doesn't initiate conversation, convinced all his friends to invest in stocks, arrives to class on time, tells his friends when he thinks they're wrong, lives life for the moment; sits in the corner at parties, argues with professors about his grade, never asks other people for help, washes his car every Saturday, blushes when he receives a compliment.

Positive descriptions: runs errands for his elderly aunt, likes to tell jokes, is able to keep a secret, always defends his best friend, helps out around the house; likes to meet new people, confessed that he broke his roommate's clock, loves to play with kittens, is good at solving riddles, has a good time wherever he goes.

Negative descriptions: interrupts during a conversation, thinks that he's smarter than his friends, tells lies to impress people, eavesdrops on conversations around him, hates when his friends talk to other people; doesn't like to share his notes from class, told people lies about his old best friend, insults people he just met, talks loudly at the movies, only talks to attractive people.

References

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1063–1087.

Anderson, M. C., & McCulloch, K. C. (1999). Integration as a general boundary condition on retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 608–629.

Anderson, M. C., & Spellman, B. A. (1995). On the status of inhibitory mechanisms in cognition: Memory retrieval as a model case. *Psychological Review, 102*, 68–100.

Anderson, N. H. (1968). Likableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology, 9*, 272–279.

Attrill, A., & MacLeod, M. D. (2004). Studying goal-directed active inhibition process for self- and other-reference trait information. Paper presented at the Fifth Inter-University Graduate Conference in Social Psychology, University of Cambridge.

Barnier, A. J., Hung, L., & Conway, M. A. (2004). Retrieval-induced forgetting of emotional and unemotional autobiographical memories. *Cognition and Emotion, 18*, 457–477.

Bjork, R. A. (1975). Retrieval as a memory modifier. In R. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 123–144). Hillsdale, NJ: Erlbaum.

Bjork, R. A. (1989). Retrieval inhibition as an adaptive mechanism in human memory. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 309–330). Hillsdale, NJ: Erlbaum.

Bjork, R. A. (1998). Intentional forgetting in perspective: Comments, conjectures, and some directed remembering. In J. M. Golding & C. N. MacLeod (Eds.), *Intentional Forgetting* (pp. 453–481). Malwah, NJ: Erlbaum.

Bjork, E. L., & Bjork, R. A. (1996). Continuing influences of to-be-forgotten information. *Consciousness and Cognition, 5*, 176–196.

Bjork, E. L., & Bjork, R. A. (2003). Intentional forgetting can increase, not decrease, residual influences of to-be-forgotten information. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 29*, 524–531

Bjork, E. L., Bjork, R. A., & Anderson, M. C. (1998). Varieties of goal-directed forgetting. In J. M. Golding & C. N. MacLeod (Eds.), *Intentional forgetting* (pp. 103–137). Mahwah, NJ: Erlbaum.

Ciranni, M. A., & Shimamura, A. P. (1999). Retrieval-induced forgetting in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1403–1414.

Dunn, E. W., & Spellman, B. A. (2003). Forgetting by remembering: Stereotype inhibition through rehearsal of alternative aspects of identity. *Journal of Experimental Social Psychology, 39*, 420–433.

Fiske, S. T. (1980). Attentional and weight in person perception: The impact of negative and extreme behavior. *Journal of Personality and Social Psychology, 38*, 889–906.

Golding, J. M., & MacLeod, C. M. (Eds.). (1998). *Intentional forgetting: Interdisciplinary approaches.* Mahwah, NJ: Erlbaum.

Hastie, R., & Kumar, P. A. (1979). Person memory: Personality traits as organizing principles in memory for behaviors. *Journal of Personality and Social Psychology, 37*, 25–38.

Johnson, H. M. (1994). Processes of successful directed forgetting. *Psychological Bulletin, 116*, 274–292.

Klein, S. B., Chan, R. L., & Loftus, J. (1999). Independence of episodic and semantic self-knowledge: The case from autism. *Social Cognition, 17*, 413–436.

Klein, S. B., Loftus, J., & Kihlstrom, J. F. (1996). Self-knowledge of an amnesic patient: Toward a neuropsychology of personality and social psychology. *Journal of Experimental Psychology: General, 125*, 250–260.

Klein, S. B., Loftus, J., & Plog, A. E. (1992). Trait judgments about the self: Evidence from the encoding specificity paradigm. *Personality and Social Psychology Bulletin, 18*, 730–735.

Koriat, A. (1995). Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology: General, 124*, 311–333.

Koriat, A. (1998). Metamemory: The feeling of knowing and its vagaries. In J. G. Adair & F. I. M. Craik (Eds.), *Advances in psychological science* (Vol. 2, pp. 461–469). Hove, Sussex: Psychology Press.

MacLeod, M. D., & Roseveare, T. A. (2002). I was always on my mind: The self and temporary forgetting. *Psychonomic Bulletin & Review, 96*, 11–614.

Macrae, C. N., Bodenhausen, G. V., Milne, A. B., & Ford, R. L. (1997). On the regulation of recollection: The intentional forgetting of stereotypical memories. *Journal of Personality and Social Psychology, 72*, 709–719.

Macrae, C. N., Bodenhausen, G. V., Schloerscheidt, A. M., & Milne, A. B. (1999). Tales of the unexpected: Executive function and person perception. *Journal of Personality and Social Psychology, 76*, 200–213.

Macrae, C. N., & MacLeod, M. D. (1999). On recollections lost: When practice makes imperfect. *Journal of Personality and Social Psychology, 77*, 463–473.

Macrae, C. N., & Roseveare, T. A. (2002). I was always on my mind: The self and temporary forgetting. *Psychonomic Bulletin & Review, 9*, 611–614.

Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2004). Encoding-specific effects of social cognition on the neural correlates of subsequent memory. *Journal of Neuroscience, 24*(21), 4912–4917.

Mitchell, J. P., Mason, M. F., Macrae, C. N., & Banaji, M. R. (in press). Thinking about people: The neural substrates of social cognition. In J. T. Cacioppo (Ed.) *Social neuroscience: People thinking about people.* Cambridge, MA: MIT Press.

Rozin, P., & Rozyman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review, 5*, 296–320.

Saunders, J., & MacLeod, M. D. (2002). New evidence on the suggestibility of memory: The role of retrieval-induced forgetting in misinformation effects. *Journal of Experimental Psychology: Applied, 8*, 127–142.

Semin, G. R., & Fiedler, K. (1988). The cognitive functions of linguistic categories in describing persons: Social cognition and language. *Journal of Personality and Social Psychology, 54*, 558–568.

Semin, G. R., & Fiedler, K. (1991). The linguistic category model: Its bases, applications and range. *European Review of Social Psychology, 2*, 1–30.

Shaw, J. S., Bjork, R. A., & Handal, A. (1995). Retrieval-induced forgetting in an eyewitness-memory paradigm. *Psychonomic Bulletin & Review, 2*, 249–253.

Smith, R. E., & Hunt, R. R. (2000). The influence of distinctive processing on retrieval-induced forgetting. *Memory & Cognition, 28*, 503–508.

Ter Doest, L., Semin, G. R., & Sherman, S. J. (2002). Linguistic context and social perception: Does stimulus abstraction moderate processing style? *Journal of Language and Social Psychology, 21*, 195–229.

Wyer, N. A., Sherman, J. W., & Stroessner, S. J. (2000). The roles of motivation and ability in controlling the consequences of stereotype suppression. *Personality and Social Psychology Bulletin, 26*, 13–25.

Ybarra, O. (2001). When first impressions don't last: The role of isolation and adaptation in the revision of evaluative impressions. *Social Cognition, 19*, 491–520.