Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Experimental Social Psychology 40 (2004) 52–64
Journal of Experimental Social Psychology
www.elsevier.com/locate/jesp

Memory monitoring and the control of stereotype distortion

B. Keith Payne,[a,*] Larry L. Jacoby,[b] and Alan J. Lambert[b]

[a] *Department of Psychology, The Ohio State University, 1885 Neil Avenue Mall, Columbus, OH 43210, USA*
[b] *Washington University, USA*

Received 9 December 2002; revised 5 May 2003

Abstract

We propose a framework for investigating the role of conscious experience in regulating stereotype-based memory distortions. Memory biases are mediated by multiple memory processes, including forms of discriminability and response biases. Different psychological interpretations of these processes depend on how they relate to subjective experiences (e.g., conscious recollection vs. implicit accessibility processes). The ability to control memory distortions, in turn, depends on the psychological meaning attached to these parameters. In Experiment 1 we found that confidence was positively related to discriminability but uncorrelated with stereotype-consistent bias. In Experiment 2 participants were allowed to selectively control when they responded. Because of the asymmetry in monitoring ability, participants were able to increase overall accuracy, but were not able to reduce the stereotype bias. Discussion focuses on the conditions in which subjective experience may provide a valid basis for controlling biases, and when subjective experience may prove deceptive, exacerbating biases.

Introduction

"*Even a fool, when he holdeth his peace, is counted wise; and he that shutteth his lips is esteemed a man of understanding*" Proverbs 17:28.

"*It's good to shut up sometimes*" Marcel Marceau.

In a classic study on memory errors and rumor, Gordon Allport (1947) showed participants a scene depicting a white man and a black man arguing on a street car. The white man held a razor in his hand. After a series of re-tellings from one participant to another, Allport reports that "In over half of the experiments with this picture, at some stage in the series of reports the Negro (instead of the white man) is said to hold the razor in his hand," (p. 111). In the years since this seminal study researchers have documented multiple ways that stereotypes can distort memory (for a review see Stangor & McMillan, 1992). Although social cognition research has made considerable progress in documenting the multiple ways in which stereotypes and other sources of social expectancies might bias memory reports, much less is known about the processes by which people might regulate those biases. If Allport's experiment were replicated today, would the same race bias be evident? Although people are much less willing to endorse and express negative race stereotypes today than in 1947, the cultural stereotypes themselves remain largely intact (Devine & Elliot, 1995). The question might hinge, in large part, on how well people are able to monitor and control the effects of stereotypes on their memory reports. In short, it is a question of whether people know when to keep quiet.

This paper proposes a framework for understanding memory distortions, with implications for whether or not people will be able to control stereotype-related memory errors. There are three main parts to the framework, each of which will be developed. To preview: (a) memory performance arises from two classes of processes, measured as discriminability and response bias parameters, (b) these parameters can each represent a variety of different psychological processes, and we show how to use confidence ratings to attach psychological interpretations to the parameters, and (c) the ability to control memory errors depends on the psychological interpretation attached to the memory parameters. That is, depending on the processes

* Corresponding author.
E-mail address: payne.265@osu.edu (B.K. Payne).

0022-1031/$ - see front matter
doi:10.1016/S0022-1031(03)00069-6

underlying the memory parameters, efforts to control prejudice may result in (a) a reduction of stereotypical bias, (b) no impact, or (c) the opposite of the desired effect, creating even stronger biases.

Although many studies have documented the ways that stereotypes and social biases may distort memory, only a much smaller set have decomposed those distortions into discriminability and response bias processes. Of those that have, several different (and mutually exclusive) psychological interpretations of those parameters have been assumed. The model developed here provides a framework for unifying the extant findings, with implications for when and how such memory distortions may be controlled.

A model of distortion regulation

Memory performance arises from discriminability and response bias processes

Memory discriminability refers to the ability to distinguish between events that actually occurred and those that did not, correcting for response biases. Response biases refer to any pattern of responding that affects both correct and incorrect responses independent of memory discriminability. Several researchers have noted that stereotypes and social biases can be mediated by multiple processes such as selective encoding, selective retrieval, and judgmental biases. In order to empirically separate processes affecting discriminability and response biases, many researchers have used signal detection methods (e.g., Banaji & Greenwald, 1995; Bellezza & Bower, 1981).

Recently, multinomial process models have been proposed to separate stereotype memory distortions into various types of discriminability and response biases (Klauer & Wegener, 1998). Klauer and Wegener's model posits multiple memory discriminability parameters and multiple bias parameters because it is designed to account for relatively complex social memory situations. In the present research we use the process dissociation procedure (Jacoby, 1991) to separate discriminability from bias. All of these approaches recognize that both discriminability and bias processes feed into distorted memory reports. Our primary interest is in how subjective experiences relate to these processes and the implications of monitoring for regulation attempts.

Subjective experience informs psychological interpretations of discriminability and bias

Discriminability and bias measures can each represent multiple psychological processes. For instance, people might discriminate between events that have and have not occurred based on conscious recollection, implicit learning, or feelings of familiarity. Similarly, response biases may arise because of conscious guessing strategies, misattributions of fluency, or unconscious uses of accessibility. Separating memory performance into discriminability and response bias measures is certainly informative. However, once these measures are computed it is important to attach the appropriate psychological meanings to them. We use the relationship between confidence and these measures to help characterize the nature of these processes.

On purely logical grounds, the relationships between confidence and the processes of memory discriminability and bias can each be either positive, negative, or nil. Each possible relationship suggests different psychological interpretations of the processes. For instance, if memory discriminability is characterized as conscious recollection for an event, then we should expect a positive correlation between subjective confidence and discriminability estimates. However, if participants are discriminating on some basis other than conscious memory (e.g., perhaps implicit learning) then the relationship might be very weak, or zero. Finally, if participants are responding based on a wrong and misleading criterion, they might show a negative correlation between confidence and discriminability: the more confidently they answer, the more likely they are to be wrong.

A similar set of possibilities exists for the relationship between experience and response bias. One of the simplest and most widely invoked interpretations of response biases is that of a guessing strategy. This interpretation implies that the bias represents a deliberate decision process (although there is nothing in the mathematics or theoretical assumptions of signal detection theory or other measurement models that necessitates this interpretation). Unfortunately researchers often malign response bias as "*just* guessing," a nuisance factor to be controlled, in contrast to discriminability, which is considered "*real* memory" (Jacoby, Marsh, & Dolan, 2001). A stereotype bias characterized as a guessing strategy would suggest that when participants are not sure of the correct answer, they guess the stereotype-consistent answer. This interpretation implies that when confidence is low participants make stereotypical guesses, but they do not do so when confidence is high. The intentional guessing account thus predicts a negative correlation between confidence and the bias estimate.

A second interpretation of the bias is as a misattribution of fluency (cf. Kelley & Rhodes, 2002). According to the misattribution of fluency characterization, stereotype categories make stereotype-consistent information come to mind easily, and this fluency is misinterpreted as conscious memory. The more easily stereotype information comes to mind, the greater confidence a person feels. This interpretation implies a positive correlation between confidence and the accessibility estimate. Whereas in the strategic guessing account, low confidence causes people to guess in stereotype-consistent ways, the misattribution of fluency account suggests that accessible stereotypes cause people to be highly confident.

Finally, the accessibility bias may be characterized as an unconscious use of accessibility similar to implicit memory. By the *unconscious accessibility* account, low confidence levels do not cause guessing, nor does accessibility cause high confidence. Instead, the subjective experience of confidence and the use of accessible information are based on separate processes, and are simply independent of each other. Stereotype categories make category-consistent information highly accessible. When people make memory reports their behaviors reflect the accessible information, but their subjective experiences might not. The unconscious accessibility hypothesis predicts the relationship between confidence and accessibility bias should be flat.

Although multiple psychological processes can produce response biases, different theoretical interpretations of these biases yield very different conclusions about the nature of stereotype processes being employed. A great deal of experimental work has demonstrated that stereotypes are flexible devices that can be used in multiple ways (e.g., Hilton & von Hippel, 1996; Sherman, Lee, Bessenoff, & Frost, 1998). It is likely that all three of the characterizations discussed above may occur under some circumstances. If a person in our hypothetical replication of Allport's classic study is under the influence of stereotypic expectations, will he or she be able to avoid memory errors? In the next section, we argue that the answer to this question depends on how subjective experience is attuned to the memory processes that produce the bias.

Control over memory distortion depends on the psychological meaning of memory parameters

The relation between monitoring and control relies on the well-supported assumption that people respond when they are relatively confident, but do not spontaneously respond otherwise (Koriat & Goldsmith, 1996). The more strongly they try to control their memory reports, the more strictly they will restrict their stories to the most confident memories. However, this simple assumption leads to different consequences depending on whether we are referring to discriminability or bias.

For memory discriminability, a positive correlation between discriminability and confidence implies that control attempts will be at least partially successful. When people respond with their most confident answers but filter out their low-confidence answers, they will filter out mostly errors (to the extent that their confidence is well calibrated). The expected result is that control attempts result in higher accuracy (although less quantity).

A negative correlation between memory discriminability and confidence implies that control attempts will backfire. If the most confident responses are the least likely to be accurate, then the result will be less accuracy *and* less quantity. Fortunately, such negative correlations are rare, and may primarily develop in deceptive laboratory situations rather than everyday life.

A zero correlation between confidence and discriminability implies that control attempts will be simply ineffective. When participants choose which responses to filter out, they will be choosing arbitrarily. As many correct statements as false ones will be filtered out, resulting in no net increase in accuracy, although quantity will decrease.

The consequences for controlling biases are somewhat different. If confidence is positively correlated with response bias (as in the case of misattributions of fluency) then the most confident memories will be the most likely to be biased by the stereotype. As a result, control attempts will backfire, leading to more stereotypical memory distortions than if participants had never tried to filter out the wrong answers to begin with.

A negative correlation between confidence and bias (as in the case of intentional guessing) suggests that control attempts will effectively reduce stereotype bias. When participants selectively respond with high-confidence answers, they will be responding with those least likely to be biased by the stereotype. This pattern is not a surprising one. It amounts to the claim that if stereotypes are influencing responses through intentional guesses, then a person can eliminate the bias by not guessing.

Finally, a zero correlation between confidence and bias suggests that control attempts will be ineffective at reducing the impact of stereotypes. When participants respond selectively with high-confidence answers, those answers are no more, nor less likely to be influenced by the stereotype.

Summary of the model. The framework presented here argues that stereotypes and other sources of memory distortion are underpinned by multiple memory mechanisms, some of which may be classified as discriminability processes, and others as bias processes. The relation between subjective experience and the different memory processes is important for understanding both the psychological nature of those processes and the prospects for controlling them. Relying on subjective experience can provide an effective means of control, but only for those memory processes to which experience is well attuned. When subjective experience is unrelated or mis-aligned with the memory process, regulation attempts will result in no improvement, or even exacerbate distortion.

To clarify the impact of cognitive monitoring on the control of stereotypes, we distinguish between two kinds of memory errors: *general memory failures* and *stereotype distortion*. By general memory failure we mean the overall rate of errors without respect to whether they are stereotype-consistent or stereotype-inconsistent errors.

In contrast, stereotype distortion refers to the *kind* of errors. It is the degree to which errors are stereotype-consistent versus stereotype-inconsistent, regardless of the overall level of accuracy.

Two experiments are reported here that test the proposed Distortion Control framework. In Experiment 1 we measured participants' confidence in each response to assess how well their subjective experience tracked processes of discriminability and response bias. In Experiment 2 we allowed participants to selectively choose on which items they would respond. Together, these experiments allow us to assign psychological meaning to the discriminability and bias parameters. Consequently, we can test predictions about the conditions under which participants can control their memory accuracy versus those conditions in which distortions slip through undetected.

Experiment one: The feel of bias

Separating discriminability and bias: Process dissociation

The approach we have taken creates conditions in which stereotypes are both congruent with, and incongruent with the events that actually occurred. This arrangement allows us to separate memory for actual events from memory distortions based on stereotypical associations (Payne, 2001; Payne, Jacoby, & Lambert, in press; Payne, Lambert, & Jacoby, 2002). We used a simple paired-associates procedure adapted from Park and Banaji (2000) to measure the effect of race stereotypes on memory. During a study phase, participants read a list of names that were stereotypical of black (e.g., Jamal, Tyrone) or white (e.g., Greg, Howard) males. Each name was paired with one of two occupations stereotypically associated with blacks (e.g., basketball player) or whites (e.g., politician). Later at test, participants saw the names as cues, and were asked to remember whether that name referred to a basketball player or a politician.

This procedure creates conditions in which stereotypes and memory for actual events are both congruent with, and incongruent with each other. For example, when a typically black name is described as a basketball player or a white name is described as a politician, responding either on the basis of memory for these facts *or* based on the stereotype category will lead to the correct answer. We characterize the influence of stereotypes here as an "accessibility bias" because the stereotype category is assumed to activate stereotype-consistent knowledge. In the case of a stereotype-based accessibility bias, information comes to mind not because of prior exposure but because it is related to other information organized within a stereotype schema. In the absence of discriminability, such accessible information may be used to guide memory reports.

The probability of a correct response in a congruent condition can be expressed as the probability of discriminability (D), plus the probability of a stereotypical accessibility bias (A) in the absence of discriminability: $P(\text{correct}\,|\,\text{congruent}) = D + A(1 - D)$. The pattern of responses expected in the incongruent conditions (e.g., black-politician, white-basketball player) is different. Participants are expected to falsely identify a black name as a basketball player or a white name as a politician to the extent that they cannot retrieve the true facts $(1 - D)$, and are influenced by a stereotypical accessibility bias (A). This is because if participants retrieve the actual event, they have a basis for choosing the correct answer. Only if discriminability fails will stereotypical biases drive responses. The probability of a false alarm in these incongruent conditions can be expressed as follows: $P(\text{stereotypic false alarm}\,|\,\text{incongruent}) = A(1 - D)$. Using these two equations, the terms D and A can be algebraically solved: $D = P(\text{correct}\,|\,\text{congruent}) - P(\text{stereotypic false alarm}\,|\,\text{incongruent})$. Given this estimate of discriminability, accessibility bias can be estimated as follows: $A = P(\text{stereotypic false alarm}\,|\,\text{incongruent})/(1 - D)$.

Our analysis suggests that stereotypes will distort memory reports, resulting in an interaction between the race of the names and the occupation studied. We expect participants to mis-report white names as politicians more than black names, and to mis-report black names as basketball players more than white names.[1] The relationship observed between subjective confidence and estimates of discriminability and accessibility bias will help shed light on how these processes should be interpreted.

Method

Participants and design

Thirty undergraduate students participated in return for course credit. All participants described themselves as either European American or Asian American.

[1] The literature relating schematic expectancies to memory shows that under some conditions, expectancy-inconsistent information is remembered better than expectancy-consistent information, with irrelevant information least remembered. A meta-analysis by Stangor and McMillan (1992) showed that the pattern of results depends on whether one is focusing on discriminability or response bias. Stereotypes tend to produce stereotype-consistent patterns in response bias and recall measures, but may produce stereotype-inconsistent advantages in discriminability. The inconsistency advantage in discriminability is related to perceivers' attempts to resolve inconsistency between expectations and actual events. In the present study, the occupation information was not so strongly tied to stereotypes that counter-stereotypical pairings required extensive inconsistency resolution. That is, it is not difficult to understand when a black person is a politician, or when a white person is a basketball player. The minimal need for inconsistency resolution in the present paradigm led us to expect no inconsistency advantage in these studies.

Participants studied both black-typical and white-typical names described as basketball players or politicians. The design, then, was a 2 (Name Race: black, white) $\times$ 2 (Occupation: basketball player, politician) factorial with both factors manipulated within participants. The primary dependent variables were the number of occupations incorrectly attributed to names typical of each race (false alarms), and the confidence judgments made for each item.

Procedure

Study phase. Participants were asked to study a list of name–occupation pairs, and were told, "It is important when meeting other people to be able to remember names and occupations, etc. In this experiment we would like you to read a list of names paired with occupations and we will ask you to match the names and occupations later." Participants then studied a list that included 24 black-typical names and 24 white-typical names, each paired with the words "basketball player" or "politician." The items were presented on a computer monitor at a rate of 3 s each. The name-occupation pairs were counterbalanced so that across participants, each white and black name was paired with both "basketball player" and "politician." The order of item presentation was randomized for each participant. Following the study phase participants worked on a filler task consisting of logical puzzles for 10 min, and then completed the memory test.

Test phase. The memory test presented the names from study, and instructed participants to indicate whether each name referred to a basketball player or politician by pressing one of two keys. After each response, a confidence probe appeared, asking "percent certainty?" and participants responded on a 6-point scale. The scale was anchored by the numbers 50 and 100%. Participants were given specific instructions on how to use this scale. They were told:

> Each of these numbers represents the number of times you expect to be correct if you made a hundred judgments at that same degree of certainty. For example, if your 'chances that your answer is correct' is 70% then you expect to be correct 70 out of 100 judgments that you gave that degree (70%) of certainty. Because there are only two alternatives, you have at least a 50% chance of being correct.

The order in which the items were tested was randomized. After completing the memory test and some ancillary measures participants were debriefed.

Results

Three sets of dependent variables were of interest. First, memory errors in each condition were examined to measure the impact of stereotypes in distorting memory reports. Second, process dissociation estimates of discriminability and accessibility bias were computed to test the mechanism by which stereotypes affected memory performance. Finally, the relationships between confidence and memory estimates were examined to test the calibration between subjective experience and each process.

Memory distortions. Table 1 displays the probability of errors in each name race and occupation condition. As predicted, participants displayed a stereotype-consistent pattern of errors, revealed in a Name Race $\times$ Occupation interaction, $F(1,29) = 4.48$, $p < .05$. Simple effects tests showed that participants falsely identified white names as politicians more frequently than black names, $F(1,29) = 4.20$, $p < .05$. They also tended to falsely identify black names as basketball players more frequently than white names, $F(1,29) = 3.18$, $p = .09$.

In sum, participants' memory performance was distorted in a stereotype-consistent direction. We have argued that this distortion arises because stereotypes make category-consistent information highly accessible, and this information is used as a basis for constructing memory reports. However, simply analyzing the error responses themselves is not able to address this question. We next report process dissociation analyses aimed at testing whether the pattern of mistaken reports reflected race differences in discriminability or accessibility bias.

Memory process estimates. Estimates of discriminability and accessibility bias were calculated for each race condition. Analyses showed no difference in discriminability for white names ($M = .23$) versus black names ($M = .20$), $F(1,29) = .32$, $p > .50$. Thus, the stereotypical pattern of errors observed cannot be explained by poorer discriminability for one group. Next accessibility bias estimates were examined. We expected the stereotype-consistent pattern to be wholly accounted for by bias differences.

In order to compare the accessibility bias for white and black names, the estimates were scored so that higher numbers indicate a greater bias toward responding "basketball player." A stereotypical accessibility bias would be revealed if the bias toward responding "basketball player" were higher for black

Table 1
Probability of correct responses and errors for basketball players and politicians described by black-typical and white-typical names in Experiment 1

	Actual basketball player		Actual politician	
	Black	White	Black	White
Correct	.66 (.16)	.56 (.20)	.55 (.17)	.67 (.22)
Errors	.34 (.16)	.44 (.20)	.45 (.17)	.33 (.22)

Note. Responses for this task were two-alternative choices. Consequently, errors for actual basketball players correspond to false "politician" responses. Errors for actual politicians correspond to false "basketball player" responses. Standard deviations are in parentheses.

names than white names. This was, in fact, what the data showed. Accessibility estimates were higher for black names ($M = .57$) than for white names ($M = .41$), $F(1, 29) = 6.07$, $p < .05$. The selective impact of race on the accessibility bias estimate suggests that the stereotypic distortions observed operated through a bias process independent of discriminability. Next we turn to the question of how well-subjective experiences of confidence related to memory accuracy, and to the estimates of discriminability and accessibility bias.

Monitoring effectiveness. The relationship between memory process estimates and subjective experience offers insights into the psychological nature of those processes. For example, our substantive interpretation of the discriminability estimate is that it represents conscious memory. If this form of memory is consciously available, then one would expect subjective experience to be well attuned to the process. We therefore expected a positive correlation between subjective confidence and discriminability.

To assess the degree of association between confidence and each memory measure we first computed performance at each level of confidence. Fig. 1 shows the calibration curves relating subjective confidence to (A) correct memory performance (hits), (B) discriminability estimates, and (C) accessibility bias estimates. To quantify the relationships we report two statistics: the Kruskal–Goodman *gamma* coefficient, and the Pearson *r*. The *gamma* coefficient has been recommended as the most appropriate index of memory monitoring performance (Nelson, 1984). The *gamma* coefficient has several desirable properties for the present purposes, including the fact that it is not dependent on the overall level of memory performance. The *gamma* coefficient may be interpreted as the degree of association between two variables, similar to other correlation coefficients. However, because some readers may be unfamiliar with this statistic, the more familiar Pearson *r* is also reported.

Panel A of Fig. 1 shows that confidence was reasonably well-related to accuracy, $gamma = .41$, $p < .0001$, Pearson $r = .48$, $p < .0001$. As we have noted, memory performance reflects both discriminability and response bias. The critical question for our present purposes is: To which memory processes people are consciously attuned? The next analyses test how well participants' subjective confidence was related to the underlying processes of discriminability and accessibility bias.

Panel B shows that, as predicted, confidence was positively correlated with discriminability, $gamma = .40$, $p < .0001$, Pearson $r = .51$, $p < .0001$. Subjective feelings of confidence were good predictors of when participants were correctly remembering an item. The size of the relationship was equal to the relationship between confidence and total accuracy, suggesting that the relationship between confidence and accuracy can be fully accounted for by the relationship between confidence

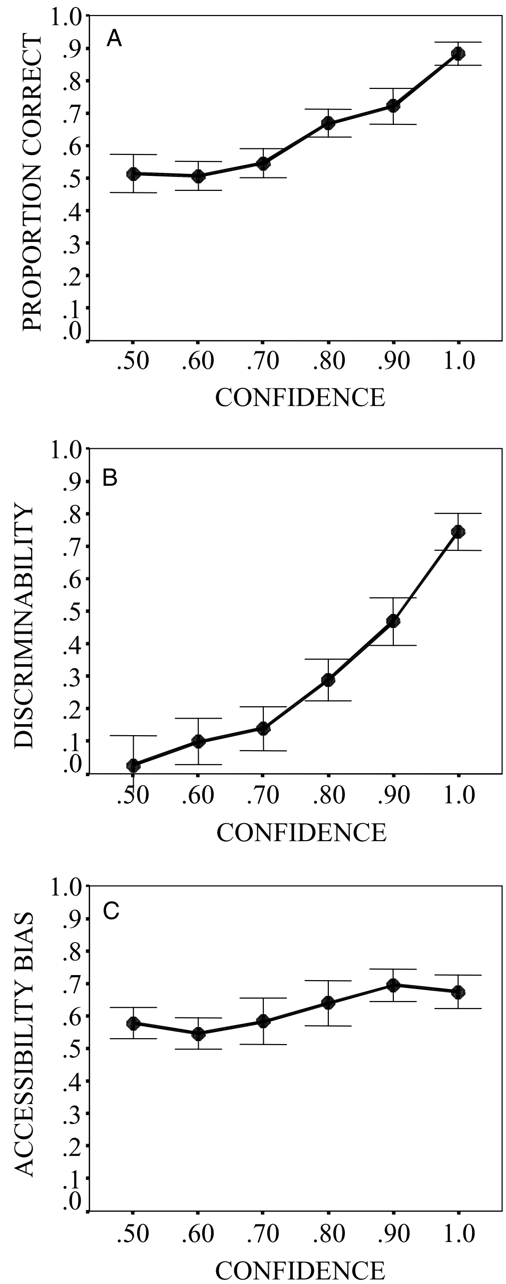

Fig. 1. Calibration curves relating confidence to (A) correct recognition, (B) discriminability estimates, and (C) accessibility bias estimates.

and discriminability. So where does this leave accessibility bias?

For the purposes of testing the relationship between confidence and accessibility bias, the bias estimates for the white and black name conditions were combined into a single score so that higher values indicate greater stereotypical bias regardless of race. In other words, a higher estimate on this scale represents a greater tendency to call black names basketball players *and* to call white names politicians. Note that .50 is a meaningful value in that it represents the point of unbiased responding—no more likely to respond in a stereotypical way than in a counter-stereotypical way.

Panel C shows that the curve relating confidence and accessibility bias is relatively flat. Tests of association showed that the relationship was not reliable, $gamma = .09$, *n.s.*, Pearson $r = .09$, *n.s.* The Pearson correlation coefficients relating confidence to discriminability and to accessibility bias were significantly different from each other, $z = 2.46$, $p < .01$. There appears to be very little relationship between confidence and accessibility bias in this study. Although this finding in itself represents a null effect, it is a theoretically informative null relationship.

Nonetheless, it is important to show that the null relationship observed was not simply the result of low statistical power. Because each participant contributed 48 observations and all variables were manipulated within participants, the design was relatively powerful. A formal power analysis was conducted, assuming an alpha level of .05. For this sample, if one assumes a true correlation of $r = .25$, the probability of detecting such a relationship was $(1 - \beta) = .88$. If one assumes the true correlation was $r = .30$, the power to detect such a relationship was $(1 - \beta) = .97$. Thus, the lack of a relation here does not seem to be the result of low statistical power.

Finally, one could argue that the flat slope depicted in Panel C of Fig. 1 could result from some participants displaying a positive correlation, and other participants displaying a negative correlation. If the numbers in each group were equal, then the mean correlation would be flat. However, a careful inspection of Panel C shows that this explanation is not supported by the data. If this were the case, then for one group the accessibility bias would be high when confidence was high and low when confidence was low. For the other group accessibility bias would be high when confidence was low, and low when confidence was high. Importantly, the lines for both groups would pass through the center of the graph, with moderate levels of accessibility bias at moderate levels of confidence. This pattern implies that for the calibration curve displayed in Panel C, the error bars would be small toward the center of the line, but larger toward the extremes. This is because the end-point values would represent the means of widely ranging values, whereas the center would represent the mean of a much smaller range of values. Inspection of Panel C shows that the error bars are not wider at the extremes. If anything, they are smaller. Thus, the data do not support the idea that the flat calibration curve reflects the combination of two different systematic patterns of relationships.

Discussion

Study 1 demonstrated that race stereotypes affected memory through accessibility bias, but not discriminability. In contrast, subjective confidence was related to discriminability, but not bias. Interpretations of the response bias as strategic guessing or as misattributions of accessibility cannot easily account for these results. Instead, the results are consistent with an interpretation of the bias as an unconscious influence of stereotype accessibility that is independent of subjective confidence. The striking aspect of this finding is that participants were just as likely to be biased by stereotypes when they expressed 100% certainty (bias = .60) as when they expressed no certainty at all (bias = .59).

The memory monitoring results have important implications for the ability to control the distorting effects of stereotypes. People who are trying to avoid stereotypic memory distortions are placed in a dilemma. On one hand, they can only regulate their behavior based on subjective experience, because they do not have direct access to the veracity of a given memory. On the other hand, stereotypes in this case appeared to operate through accessibility processes, to which subjective experience was insensitive. We explore the consequences of this dilemma for avoiding memory errors in Experiment 2.

Before introducing the second experiment we note some caveats and limitations of the present study. First, as noted in the introduction we do not assume that stereotypes always function through unconscious accessibility bias processes. Instead, in some cases people may intentionally use stereotype information to selectively encode or retrieve information, resulting in differential memory discriminability. They might also use stereotype cues intentionally to guide guesses, or they might misattribute the accessibility caused by stereotypes as conscious memory. Our point is not that unconscious accessibility bias is the only route by which stereotypes distort memory, but that it is one important and widely applicable way. Importantly, the framework developed here offers a means to detect which kind of process is mediating stereotype distortions, in a way that has direct implications for attempts to regulate those distortions. Identifying situations in which stereotypes operate through the other means detailed above may also help document the range of control outcomes posited by the model.

A second and related issue is to note strengths and weaknesses of the memory procedure used here. We selected this procedure because it allows one to cleanly and simply separate the influence of stereotypes on the processes of discriminability and bias. One limitation of such a procedure is that simple pairings do not capture the complexity of real social interactions. Other paradigms commonly used in studying social memory sometimes include sets of behaviors ostensibly performed by various individuals. As an example, the multinomial model of Klauer and colleages described earlier is designed for the "*who said what*" paradigm (Taylor, Fiske, Etcoff, & Ruderman, 1978) in which multiple characters belonging to different social groups make various statements. However, simpler lists of items such as those used here have also contributed important insights into social memory processes. For example, Banaji and Greenwald (1995) used name lists to examine implicit stereotype biases. Park and Banaji (2000) used lists similar to the present ones to examine the different processes by which mood influences stereotyping. And Lenton, Blair, and Hasite (2001) recently used word lists to study the role of stereotypic associations in producing false memories. Whether simple versus more complex and realistic procedures are more appropriate depends, of course, on one's goals. In some cases researchers wish to adapt laboratory paradigms to mirror the complexity of the "real world." On the other hand, it is often useful to scale down that complexity to more clearly isolate the processes of interest. The basic processes isolated using this laboratory paradigm can easily be generalized to more complex and realistic situations. In the next experiment we turn more directly to how well people can control the memory errors resulting from different memory processes.

Experiment two: When to keep quiet

By monitoring their feelings of confidence, participants may be able to reduce their *general* inaccuracy by "keeping quiet" when they cannot recollect the true answer. However, when stereotypes function through unconscious accessibility bias, participants may be unable to reduce the stereotypic nature of their memory errors. According to the results from the first experiment, low confidence was *not* a good indicator that one's errors were likely to be stereotypical. As a result, we expect control attempts to benefit general memory accuracy but not to guard against stereotype-consistent memory distortion. In Experiment 2, we tested these predictions by allowing participants to choose whether or not to respond.

Method

Fifty-seven undergraduate students participated in return for course credit (36 women and 21 men). Five African American participants were not included in the analyses reported, leaving a total of 52 participants. The design for the study phase was identical to Experiment 1. The test phase was similar, but with the following changes. First, at test participants were required to answer on half of the trials (forced response condition) and allowed to avoid responding whenever they chose on the other half by pressing a "pass" key (free response condition). The response option was indicated on each trial by the words "must answer" for forced response trials and "may answer or pass" for free response trials. Forced and free response trials were randomly intermixed at test. The stimulus materials and response options were counterbalanced across name and occupation conditions, so that across the experiment, all names served as basketball players and politicians, and in free-response and forced-response conditions. The design was a 2 (Name Race: black, white) $\times$ 2 (Occupation: basketball player, politician) $\times$ 2 (Response format: forced, free) factorial with all factors manipulated within participants.

Results

Our main interests in this study were whether participants could regulate their memory performance when given choices about when to respond or refrain. We examine memory performance both in terms of quantity and accuracy, for both general memory and stereotype-specific memory performance. *Memory quantity* concerns the number of correct answers out of the total number of items studied. *Memory accuracy* concerns the number of correct answers out of the number of items actually answered. In the forced response condition, these two are identical because all items are answered. However, they are different in the free response condition when the number of items answered varies.

Memory quantity

General memory failures. Table 2 displays the quantity of hits and false alarms in each Name Race, Occupation, and Response option condition. As expected, participants made fewer errors in the free response condition ($M = .19$) than the forced response condition, ($M = .30$), $F(1, 51) = 35.55$, $p < .001$. At the same time, because participants simply responded less often, they also made fewer correct responses in the free response ($M = .53$) than in the forced response condition

Table 2
Memory quantity for basketball players and politicians described by black-typical and white-typical names in Experiment 2

	Actual basketball player		Actual politician	
	Black	White	Black	White
Forced response				
Correct	.70 (.23)	.57 (.24)	.68 (.22)	.74 (.23)
Errors	.27 (.24)	.41 (.25)	.28 (.21)	.23 (.20)
Free response				
Correct	.59 (.25)	.48 (.25)	.50 (.28)	.55 (.27)
Errors	.16 (.16)	.21 (.22)	.22 (.25)	.16 (.18)

Note. Memory quantity is computed by dividing the number of correct or error responses by the total number of items studied. Responses for this task were two-alternative choices. Consequently, errors for actual basketball players correspond to false "politician" responses. Errors for actual politicians correspond to false "basketball player" responses. Standard deviations are in parentheses.

($M = .67$), $F(1,51) = 43.74$, $p < .001$. For some purposes, the absolute number of errors or correct responses alone is important. For example, a person might be nervous about embarrassing him or herself by saying the wrong thing. If the goal is to avoid saying anything incorrect, then this performance may be seen as an improvement compared to the forced response condition, even at the cost of saying fewer of the "right" things. This kind of concern may be particularly likely in situations where there is a social cost to errors, as in the case of stereotype-based errors.

Stereotype distortions. To test the impact of race stereotypes on memory errors, the false alarms from Experiment 2 were analyzed using a 2 (Name Race) × 2 (Occupation) × 2 (Response Format) ANOVA. Replicating the results of Experiment 1, participants displayed a stereotype-consistent pattern of errors. As shown in Table 2, participants falsely identified white names as politicians more frequently than black names, $F(1,51) = 9.32$, $p < .01$. They also tended to falsely identify black names as basketball players more frequently than white names, although this simple effect was not significant, $F(1,51) = 2.46$, $p = .12$. This replicable Name Race × Occupation interaction, $F(1,51) = 7.50$, $p < .01$ is the signature of stereotype use.

The stereotype-consistent nature of the errors revealed in this interaction did not differ as a function of whether responses were forced or voluntary. As previously discussed, the overall error rate was lower in the free response condition than the forced response condition, but the stereotypical *pattern* of errors was clearly present in both conditions. This was confirmed statistically by the lack of a three-way Name Race × Occupation × Response Format interaction, $F(1,51) = 1.55$, $p > .20$.

An analysis of the hits supported the same conclusions, as displayed in Table 2. Participants were more likely to correctly identify black basketball players and white politicians than black politicians and white basketball players, $F(1,51) = 3.74$, $p < .06$ for the Name race × Occupation interaction. Again, this stereotype effect was not moderated by response format, $F(1,51) = 2.69$, $p > .10$.

In addition to these predicted findings, there was an Occupation × Response format interaction, $F(1,51) = 8.65$, $p < .01$. This interaction shows that regardless of race, participants were more likely to mistake a basketball player for a politician than to mistake a politician for a basketball player, but this difference only emerged in the forced response condition. Although this effect was not predicted, it does not qualify our predicted and obtained results concerning racial bias.

Was the voluntary response option useful at all in avoiding stereotype-based errors? It was in one sense. As illustrated with general memory performance, the overall frequency of stereotype-consistent errors was lower in the free response condition than in the forced response condition. But this lower frequency of stereotype-consistent errors was accompanied by fewer stereotype-inconsistent hits. In fact, the free response option led to fewer responses of all kinds. If one's primary goal were to avoid stereotypical errors at all cost, then withholding responses might be useful, even at the cost of saying fewer correct things. Like a shy student in a seminar, if one wanted to avoid any possibility of error, the surest way to do so would be to never speak up at all.

For most purposes, however, we are concerned not with absolute numbers of true or false statements, but with how well a person's statements correspond to what actually happened. That is, we want to investigate the accuracy of memory. In essence, people in everyday life have a natural "pass" option. Accuracy is more relevant to most realistic situations outside the laboratory in which the main concern is how often people are correct out of those times when they choose to speak up.

Memory accuracy

General memory failures. As described in the sections above, participants gave fewer correct and fewer incorrect responses when allowed to withhold answers. Were participants blindly conservative in the free-response condition, so that they held silent as many correct responses as incorrect ones? Or were they strategic in their control efforts, withholding mostly errors while responding when likely to be correct? To answer these questions we analyzed participants' accuracy scores, computed as the number of errors divided by the number answered. Table 3 displays the accuracy of responses. Because discriminability was related to subjective confidence, we expected accuracy to improve when participants were allowed to decline responses.

As predicted, participants were more accurate in the free-response condition ($M = .74$) than the forced

Table 3
Memory accuracy for basketball players and politicians described by black-typical and white-typical names in Experiment 2

	Actual basketball player		Actual politician	
	Black	White	Black	White
Forced response				
Correct	.72 (.23)	.58 (.26)	.71 (.22)	.76 (.21)
Errors	.28 (.23)	.42 (.25)	.29 (.30)	.24 (.20)
% Answered	.97	.98	.96	.97
Free response				
Correct	.79 (.22)	.70 (.29)	.71 (.30)	.76 (.27)
Errors	.21 (.22)	.30 (.29)	.29 (.30)	.24 (.27)
% Answered	.75	.69	.77	.66

Note. Memory accuracy is computed by dividing the number of correct or error responses by the number of items volunteered. Errors for actual basketball players correspond to false "politician" responses. Errors for actual politicians correspond to false "basketball player" responses. Standard deviations are in parentheses.

response condition ($M = .69$), $F(1, 49) = 3.59$, $p = .03$, one tailed. Whereas they gave fewer correct and incorrect answers, they managed to be somewhat selective in their refusals. A critical question is whether this selectivity allowed participants to eliminate or reduce the pattern of errors caused by stereotypes. If the stereotype bias in the forced report condition were caused by strategic guessing, then participants should be able to eliminate the bias by simply "passing" when unsure, rather than guessing. However, stereotype bias driven by implicit memory might leave participants vulnerable to stereotype-consistent biases even in the free response condition.

Stereotype distortion. When participants' false alarms were analyzed as accuracy rather than quantity, the stereotype-consistent pattern of errors still held. In particular, participants made more stereotype-consistent errors than stereotype-inconsistent errors, as revealed by a Name Race × Occupation interaction, $F(1, 49) = 7.18$, $p < .01$. Critically, the stereotypical Name Race × Occupation interaction was not qualified by free versus forced responding, $F(1, 49) = .93$, $p > .30$. Free responding aided overall accuracy but did not eliminate the stereotyping effect.

Summary. Exercising free choice over response options provided participants some benefits, but these benefits were selective. We distinguished between general memory performance, which referred to the overall level of memory, and stereotype distortion, which referred to the pattern of memory errors. We also distinguished between memory quantity and memory accuracy. Free responding produced a benefit in general memory accuracy. In the free response condition participants made fewer hits and fewer false alarms (less quantity), but the reduction in false alarms out-paced the reduction in hits. This resulted in an increase in overall accuracy.

The benefits for stereotype-consistent distortions were more circumscribed. As just described, participants made fewer errors and fewer hits in the free-response condition, but the stereotype-consistent pattern of responses did not differ between forced and free response conditions. The fact that refusing to respond produced fewer errors and fewer hits may be seen as successful regulation if the main goal is to avoid stereotype-consistent errors, regardless of correct responses. But if the goal is to maximize the correspondence between reports and actual events rather than stereotypes, this strategy clearly failed. The stereotype-consistent pattern of responses was present regardless of whether memory was evaluated as quantity or accuracy. As a final set of analyses, we report the process dissociation estimates of accessibility bias and discriminability to show convergence with Experiment 1.

Process dissociation analyses. Experiment 1 showed that the stereotype-consistent pattern of memory errors was due to an unconscious accessibility bias. If this bias operated in the free response condition by "evading detection" by monitoring processes, then we would expect the race difference in accessibility bias also to be present under free responding. The accessibility bias estimate was scaled such that higher numbers refer to a greater bias toward responding "basketball player" rather than "politician." Replicating Experiment 1, estimates of accessibility bias were higher for black names ($M = .40$) than for white names ($M = .28$), $F(1, 51) = 4.44$, $p < .05$. This main effect was not qualified by the response option (Name Race × Response Format $F(1, 51) = 1.36$, $p = .25$), indicating that the race difference was uniform across the forced response (Ms = .47 vs. .32) and free response (Ms = .30 vs. .23) conditions.

Discriminability estimates were not affected by free responding ($M = .33$) versus forced responding ($M = .38$), $F(1, 51) = 1.83$, $p = .18$. There was, however, an effect of name race, with higher discriminability for black names ($M = .39$) than white names ($M = .31$), $F(1, 51) = 6.11$, $p < .05$. This result was not found in Experiment 1, and may be driven by the fact that there were more responses to black names than white names in the free response condition (see Table 3). At any rate, the finding of superior discriminability for black names would work against the tendency to misreport black names as basketball players. As such, this difference does not qualify the conclusions drawn about the processes involved in regulating stereotype biases.

To more formally verify that stereotype distortions of memory performance were mediated by accessibility bias rather than discriminability, we conducted a pair of ANCOVA's. First, the differences between each of the estimates for white names and black names were computed for each participant. If the stereotypic errors were driven by race-specific accessibility bias, then controlling for accessibility differences should reduce or eliminate the Name Race × Occupation interaction. Controlling for differences in discriminability should not have any effect.

When the accessibility bias estimate was entered as a covariate, the Name Race × Occupation interaction was reduced to non-significance for both forced report, $F(1, 50) = 2.13$, $p = .15$, and free report conditions, $F(1, 50) = .05$, $p = .82$. In contrast, when the discriminability estimates were entered as a covariate, the Name Race × Occupation interaction remained significant for both forced response, $F(1, 50) = 8.62$, $p < .005$, and free response conditions, $F(1, 50) = 6.48$, $p < .01$. Thus, the stereotypic memory distortions observed in Study 2 were mediated by race-specific accessibility bias but not race differences in discriminability.

Discussion

Based on the results of Experiment 1, we argued that the stereotypical pattern of memory errors caused by unconscious stereotype accessibility would be difficult to control, whereas the general level of memory performance based on discriminability would be more easily regulated. Experiment 2 provided support for both of these predictions. The novel finding in this study is the asymmetry between regulation success for general performance and stereotype-consistent patterns of bias. The regulation strategy of refusing to respond was effective for modifying general performance levels, but not for stereotype-consistent error patterns. At the same time, voluntary responding did produce fewer errors in an absolute sense. These trade-offs highlight the importance of exploring both the costs and benefits of control strategies.

General discussion

These studies shed light on the nature of stereotype memory biases, their underlying processes, and people's ability to control them. We found that in the present study the commonly observed response bias toward stereotype-consistent memories was the result of an unconscious accessibility process. The subjective experience of confidence was reliably related to discriminability, suggesting conscious recollection, but confidence was not related to accessibility bias. Critically, race stereotypes affected accessibility bias rather than discriminability. This dissociation has implications for the ability to monitor and control stereotype biases.

Participants given voluntary control over when they "speak up" or "keep quiet" met with predictably mixed success. By not responding when uncertain, they managed to reduce their general inaccuracy. However, a substantial number of errors still occurred, indicating some "leakage" through the filtering process. Because subjective confidence was uninformative about accessibility bias, those remaining errors were just as stereotypical as in the forced response condition. Errors caused by stereotypes appeared to be particularly stealthy, resulting in stereotype distortions even in volunteered behavior. In short, when it came to stereotypes, our participants did not know when to keep quiet.

Other studies have provided evidence that stereotypes may bias memory through implicit processes (e.g., Hense, Penner, & Nelson, 1995; Macrae, Schloerscheidt, Bodenhausen, & Milne, 2002). However, in the present paper we have gone beyond the issues raised in those studies in several ways. We empirically tested the relationships between subjective experience and objective estimates of discriminability and bias. The relationships we observed allowed us to derive predictions for how well participants would succeed at controlling memory distortions.

In the present study we have focused on cognitive factors in stereotyping, treating stereotypes much like other non-social categories. However, other considerations such as motivations are important in studying the regulation of *social* biases, in particular. The model we have outlined holds potential for investigating the mechanisms through which social motivations have their effects on stereotype regulation. For instance, participants highly motivated to avoid stereotypes might process the material carefully, leading to higher recollection. Such motivation might also lead to differing criteria for willingness to respond on stereotype-consistent versus stereotype-inconsistent trials. That is, participants highly motivated to avoid stereotypes might be reluctant to answer at all when they are at most risk of displaying stereotype bias. As with other bases for regulation attempts, the outcome will depend largely on the relation between subjective experience and the memory processes underlying the bias. Our model does not specify the exact ways that social motivations should influence regulation; motivated individuals may take many different routes to the same goal. Instead, the model offers a framework for analyzing exactly which mechanisms are involved when such motivations are invoked.

The present research was conducted in the context of a long history of studies examining the effects of stereotypes on memory. In the sections below, we discuss the commonalities and the points of difference between the present work and other research in this tradition.

Stereotype bias as strategic guessing or implicit process?

Research that has investigated the influence of stereotypes on response bias has often portrayed stereotype-consistent response bias as an intentional guessing strategy. For example, an influential study by Snyder and Uranowitz (1978) showed that stereotype-consistent information about a woman who was described as lesbian was preferentially remembered over stereotype-neutral information. Bellezza and Bower (1981) conducted an experiment following up on this study, applying signal detection analyses. They concluded, "Activation of the stereotype appeared to bias guessing in a congruent manner," (Bellezza & Bower, 1981, p. 864). Of course, guessing is one way to produce differences detected by bias parameters. But examining a bias parameter doesn't necessarily reveal the unintentional versus strategic, or conscious versus unconscious nature of the underlying process. Sometimes, response bias reflects implicit processes (e.g., Banaji & Greenwald, 1995; Hense et al., 1995).

Both the guessing account of Bellezza and Bower (1981) and the "implicit memory" account of Hense et al., 1995 show that stereotype-congruent memory bias can represent a kind of response bias. However, the psychological interpretations of the bias in these two accounts are quite different. Both the "misattribution of accessibility" and the "unconscious accessibility" interpretations refer to implicit processes of different kinds, as opposed to deliberate guessing. These very different psychological interpretations arising from (mathematically) similar bias parameters highlight the importance of considering subjective experiences. Our model, invoking the relationship between subjective experience and different memory parameters, provides a way of understanding those psychological processes.

Single and dual process accounts of bias

Not all researchers take response bias to necessarily reflect a conscious decision strategy. For example, Banaji and Greenwald (1995) have used signal detection methods to analyze stereotypic biases. Using a modified false fame procedure (Jacoby, Kelley, Brown, & Jasechko, 1989), they found that male names were more likely than female names to be judged famous. This effect was interpreted as a reflection of the stereotype that males have higher achievement and prestige than females. They interpreted stereotype-consistent differences in response bias as a form of implicit attribution rather than conscious decision processes. In this respect, the approach taken by Banaji and Greenwald is in agreement with our own. However, there are also important differences between these approaches, which lead to different characterizations of stereotype biases and the ways in which they might be controlled. In the following we discuss some of the differences between signal detection and dual process models in how memory biases are understood.

First, signal detection approaches assume a single continuum of evidence strength, which is the basis for both discriminability and bias. Bias refers to the placement of a criterion along the strength dimension, marking off how much evidence is needed to respond one way rather than the other. When applied to the false fame procedure, the criterion shift would reflect the idea that less evidence of fame was needed to classify male names as famous, compared to female names. Despite the fact that two parameters are computed, only a single dimension of information (e.g., memory strength) serves as the basis for both discriminability and bias.

In contrast to the single process account, our dual process model assumes that discriminability and accessibility bias reflect *qualitatively different* bases for responding. As applied to the present research, discriminability may reflect conscious memory (recollection) for events that actually happened, whereas accessibility bias is based on associations independent of the actual episode. Thus a bias does not reflect the need for less recollective information to classify a black-typical name as a basketball player. Instead, the source of information is found in stereotypic associations that are distinct from memory for actually experienced events.

A second area of difference between single dimension models such as signal detection and dual process models is in the assumed relationship between confidence and memory. Signal detection theory assumes that memory strength drives both recognition judgments and confidence judgments (MacMillan & Creelman, 1991). However, in many domains the actual confidence-accuracy relationship is not strong. Numerous studies have shown that confidence and accuracy are often very poorly related, especially when the questions are deceptive or misinformation has been introduced (e.g., Koriat, Lichtenstein, & Fischhoff, 1980; Sharp, Cutler, & Penrod, 1988).

Besides low confidence-accuracy correlations, signal detection theory has difficulty accounting for the fact that confidence can be based on multiple sources (Gardiner & Java, 1990). For example, one can be highly confident in a response either because of consciously remembering an event, or because it seems highly familiar without conscious memory. This distinction is difficult to accommodate within a single factor model which does not distinguish between separate forms of memory, but instead posits only a unitary continuum of memory strength. In contrast, a dual process model can easily account for differential bases for confidence. Similarly, the dual process model allows for a range of possible relationships between confidence and different memory processes, as outlined explicitly in the present model. As demonstrated in the present studies, the different possible relationships between confidence and memory processes are important for regulating memory distortions.

Although we have focused on confidence judgments, other varieties of subjective experience (e.g., remembering versus knowing) may also be used as a basis for monitoring and controlling the impact of social memory biases. In everyday life there are many sources of bias that affect memory and judgment, and they are likely to vary in how easily they are controlled. In developing a better understanding of how these biases are regulated (and dys-regulated), it will be important to examine the relationships between subjective experience and the cognitive processes that drive those biases.

Acknowledgments

This work was funded, in part, by Grant SBR 9817554 from the National Science Foundation to Alan Lambert. Thanks to Elizabeth Marsh, Klaus Fiedler, and two anonymous reviewers for comments on earlier versions of this paper.

References

Allport, G. W. (1947). *The psychology of rumor*. New York: Rinchart & Winston, Holt.

Banaji, M. R., & Greenwald, A. G. (1995). Implicit gender stereotyping in judgments of fame. *Journal of Personality and Social Psychology, 68*, 181–198.

Bellezza, F. S., & Bower, G. H. (1981). Person stereotypes and memory for people. *Journal of Personality and Social Psychology, 41*, 856–865.

Devine, P. G., & Elliot, A. J. (1995). Re racial stereotypes really fading? The Princeton trilogy revisited. *Personality and Social Psychology Bulletin, 21*, 1139–1150.

Gardiner, J. M., & Java, R. I. (1990). Recollective experience in word and nonword recognition. *Memory and Cognition, 18*, 23–30.

Hense, R. L., Penner, L. A., & Nelson, D. L. (1995). Implicit memory for age stereotypes. *Social Cognition, 13*, 399–415.

Hilton, J. L., & von Hippel, W. (1996). Stereotypes. *Annual Review of Psychology, 47*, 237–271.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*, 513–541.

Jacoby, L. L., Kelley, C., Brown, J., & Jasechko, J. (1989). Becoming famous overnight: Limits on the ability to avoid unconscious influences of the past. *Journal of Personality and Social Psychology, 56*, 326–338.

Jacoby, L. L., Marsh, E. J., & Dolan, P. O. (2001). Forms of bias: Age-related changes in memory and cognition. In M. Naveh-Benjamin, M. Moscovitch, & H. L. Roediger III (Eds.), *Perspectives on Human Memory and Cognitive Aging: Essays in Honour of Fergus Craik*. New York: Psychology Press.

Kelley, C. M., & Rhodes, M. G. (2002). Making sense and nonsense of experience: Attributions in memory and judgment. In D. L. Medin (Ed.), *The psychology of learning and motivation: Advanes in research and theory* (Vol. 41,. San Diego, CA: Academic Press.

Klauer, K. C., & Wegener, I. (1998). Unravelling social categorization in the "Who said what?" paradigm. *Journal of Personality and Social Psychology, 75*, 1155–1178.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*, 490–517.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning & Memory, 6*, 107–118.

Lenton, A. P., Blair, I. V., & Hastie, R. (2001). Illusions of gender: Stereotypes evoke false memories. *Journal of Experimental Social Psychology, 37*, 3–14.

MacMillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.

Macrae, C. N., Schloerscheidt, A. M., Bodenhausen, G. V., & Milne, A. B. (2002). Creating memory illusions: Expectancy-based processing and the generation of false memories. *Memory, 10*, 63–80.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95*, 109–133.

Park, J., & Banaji, M. R. (2000). Mood and heuristics: The influence of happy and sad states on sensitivity and bias in stereotyping. *Journal of Personality and Social Psychology, 78*, 1005–1023.

Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology, 81*, 181–192.

Payne, B. K., Jacoby, L. L., & Lambert, A. J. (in press). Attitudes as accessibility bias: Dissociating automatic and controlled processes. In J. S. Uleman, J. A. Bargh, & R. Hassin (Eds.), *The unconscious mind*.

Payne, B. K., Lambert, A. J., & Jacoby, L. L. (2002). Best laid plans: Effects of goals on accessibility bias and cognitive control in race-based misperceptions of weapons. *Journal of Experimental Social Psychology, 38*, 384–396.

Sharp, G. L., Cutler, B. L., & Penrod, S. D. (1988). Performance feedback improves the resolution of confidence judgments. *Organizational Behavior & Human Decision Processes, 42*, 271–283.

Sherman, J. W., Lee, A. Y., Bessenoff, G. R., & Frost, L. A. (1998). Stereotype efficiency reconsidered: Encoding flexibility under cognitive load. *Journal of Personality and Social Psychology, 75*, 589–606.

Snyder, M., & Uranowitz, S. W. (1978). Reconstructing the past: Some cognitive consequences of person perception. *Journal of Personality and Social Psychology, 36*, 941–950.

Stangor, C., & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and social developmental literatures. *Psychological Bulletin, 111*, 42–61.

Taylor, S. E., Fiske, S. T., Etcoff, N. J., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology, 36*, 778–793.