

## Gender and orientation stereotypes bias source-monitoring attributions

**Richard L. Marsh and Gabriel I. Cook**

*University of Georgia, Athens, GA, USA*

**Jason L. Hicks**

*Louisiana State University, Baton Rouge, LA, USA*

Four experiments were conducted to determine whether gender stereotypes influence source-monitoring decision processes. Statements that were consistent with a male were more often correctly attributed to a male source and less frequently correctly attributed to a female. The reverse was true for items traditionally associated with a female. Both of these biases were reversed if participants believed the speaker was either a gay male or a lesbian female. These effects persisted under divided attention during test, suggesting that they are caused by automatic influences. But these biases were partially attenuated when participants first considered the detrimental impact of stereotypes. Because these biases were absent for gender-neutral statements, the results from this study show that the content of a memory can influence judgements about the context in which something was learned. The authors argue that the data are most consistent with a heuristic, early selection process that can be influenced by a conscious, late correction process (e.g., Jacoby, Kelly, & McElree, 1999).

There are a variety of means by which to attribute a memory to the specific context in which it was originally formed (Johnson, Hashtroudi, & Lindsay, 1993). Nevertheless, the processes of source monitoring that have been investigated most extensively in recent years involve inspecting what attributes are strongest in the memory trace. These memorial attributes, termed qualitative characteristics in the source memory literature, include such features as the amount of perceptual detail, level and type of emotional content, degree of cognitive operations that are residual from organisation and elaboration, and spatio-temporal details (to name a few). For example, if one knows that an individual spoke a particular fact but cannot remember on which of two occasions that occurred, the amount of perceptual detail or

spatio-temporal detail can be used to discriminate in what context the statement was originally made.

When partial details that are retrieved are insufficient for determining the origin of a memory (e.g., Hicks, Marsh, & Ritschel, 2002; Qin, Raye, Johnson, & Mitchell, 2001), there are alternative ways of making a source attribution. In these cases, people can use categorical knowledge, schemas, stereotypes, or plausibility information (e.g., Mather, Johnson, & De Leonardis, 1999; Spaniol & Bayen, 2002). For example, one might forget the person who uttered a ribald remark, but one may only know a few people who are good candidates for such witticisms. Based on this general knowledge of personality traits, a source attribution can be made or, at the very least, the

---

Correspondence should be addressed to Richard L. Marsh, Department of Psychology, University of Georgia, Athens, GA 30602-3013, USA. Email: rlmars@uga.edu

We thank Betsy Spinks, Kyle McCranie, Meredith Thompson, and Danny Proto for their help in collecting the data. We also appreciate Jaime Durley's assistance in earlier stages of the project.

candidate pool of potential sources can be reduced. Our goal in this article is to explore one basis for source attributions that are based on such information—gender stereotype information. More specifically, we investigated how people might use gender stereotypes (as well as information about sexual orientation) to attribute statements to male versus female sources. Before detailing the rationale for our study, we first summarise several of the very few reports that have investigated this alternative basis for source attributions.

Mather et al. (1999) presented participants with statements from two individuals that were consistent with either a Democratic versus a Republican political point of view. They found that older adults as compared with younger adults tended to rely on the stereotype more heavily in making their source judgements. However, both age groups who considered their own emotional reactions to the statements during encoding tended to rely on the stereotype more heavily than if they considered the speaker's emotional focus. Thus, source attributions can be influenced, at least in some cases, by considering schematic information. Sherman and Bessenoff (1999) found that younger adults would use stereotype information (i.e., a skinhead versus a priest) in a list discrimination paradigm, but only when their attention was divided during the testing session. Under conditions of full attention, the influence of stereotype was completely absent. Bayen, Nakamura, Dupuis, and Yang (2000) have argued that such schematic information provided after learning biases source decisions in the absence of source-specifying details (i.e., true source memory). They found that the guessing parameters were changed in multinomial modelling examinations of schematic sources such as statements made by a doctor versus a lawyer, as well as about objects typically associated with a bathroom versus a bedroom. In other words, statements semantically consistent with a doctor, but lacking adequate source-specifying information, were attributed to the doctor source more often than to the lawyer.

Hicks and Cockman (2003) investigated the doctor versus lawyer schemas in more detail. They provided the schema either before learning or before the test as Bayen et al. (2000) did. They discovered that providing the schema before learning improved source memory dramatically, thereby eliminating the guessing bias during the source test. Like Bayen et al., however, they found

that the schema provided after encoding and before the test resulted in biased guessing in favour of the schema. By way of summary, younger adults may use schematic and stereotypical information, but they have only been shown to do so under the very limited conditions of (1) cognitive distraction at test, (2) when they considered their own emotional reactions to statements during encoding, and (3) with doctor-lawyer or bedroom-bathroom schemas. Although it makes perfect sense that people might base their source attributions on such general knowledge, the empirical evidence supporting this claim is currently neither very broad nor very deep. Given that divided attention in younger adults often mimics the performance of older adults (e.g., Jacoby, Debnar, & Hay, 2001; Jennings & Jacoby, 1993), the Sherman and Bessenoff (1999) results follow quite naturally from the work conducted by Mather et al. (1999).

Our goal in the present study was to explore whether source attributions to male versus female sources would be biased by gender stereotypes. Gender stereotypes should be quite strong because they are introduced quite early in childhood and they pervade our culture (e.g., Bem, 1981). Thus, younger adults may be quite prone to attribute statements that are consistent with a masculine stereotype (e.g., fixing cars) to a male source as opposed to a female source. Likewise, statements consistent with a female stereotype (e.g., baking cookies for a friend) may be more likely to be attributed to a female than a male. In the course of everyday life, using schematic information about the content of a statement might be very useful in narrowing down the candidate sources (e.g., Cook, Marsh, & Hicks, 2003). For example, if a person has only two or three colleagues who are avid stock market investors, then statements concerning this topic are likely to have come from just one of them. Dodson, Holland, and Shimamura (1998) found that when people could not distinguish between which of two males or which of two females uttered neutral material, they did have some partial knowledge about the source as being male versus female. Presumably, if the content of the memory trace could somehow be tied to one or the other of the two individuals, source monitoring would be successful. Of course, the strategy of first using general knowledge (schematic information) followed by more focused plausibility considerations is only useful if the content of the memory was indeed consistent with the schema. When it is not, such as

when a male claims to have fluffed pillows on a sofa, then this source-monitoring strategy will potentially fail.

Although we are not investigating a second stage in which candidate sources are narrowed further, we are investigating the initial use of gender stereotypes as a means of attributing the source to a memory trace. Given that such a gender bias may exist, it should also be possible to reverse the pattern of attributions with a different stereotype based on sexual orientation. For example, participants told that the source was a gay male may attribute more statements traditionally consistent with a female to him, as opposed to a female who is depicted as a lesbian. In this case, the gay male or lesbian stereotype may override or replace the standard gender stereotype. In a similar vein, given either two male sources or two female sources, portraying one as a gay male (or a lesbian) and the other as a heterosexual male (or heterosexual female) may bias source attributions as well. If participants use the gender and the orientation stereotypes as we have predicted, then it will demonstrate just how flexible source-monitoring decision processes can be. In the introduction to Experiments 2 and 3 we raise the issue of whether the use of stereotype information is a conscious strategy or is a more automatic process.

## EXPERIMENT 1

This first experiment was designed as an initial test of whether participants would base their source judgements on gender stereotypes. Two fictitious speakers made equal numbers of male-oriented, female-oriented, and neutral statements. Like Bayen et al. (2000) and Hicks and Cockman (2003) the stereotype information was provided after learning and before the source test. Under conditions of full attention, the gender stereotype was predicted to influence performance in so far as more male-oriented items would be attributed to the male and vice versa for the female. To ascertain if we could reverse the stereotype, a different group of participants was told that one speaker was a gay male and the other a lesbian. If the attribution bias is reversed by this manipulation, then participants must be basing their judgements on sexual orientation rather than gender.

## Method

*Participants.* A total of 56 undergraduates from the University of Georgia volunteered in exchange for partial credit towards a course requirement. Each participant was tested individually in sessions that lasted approximately 30 minutes. As described shortly, 28 participants were randomly assigned to each of the heterosexual and homosexual stereotype conditions. An additional 49 participants took part in a pilot experiment in which statements were normed for how characteristic they were of males versus females.

*Materials and procedure.* An original set of 114 statements were developed that we believed might be perceived as characteristic of a male or a female, or gender neutral. Some of these items were based on materials drawn from Bem (1981) and Bardach and Park (1996) whereas others were devised in-house. These were randomised into a single order and presented to the pilot participants as a paper and pencil task. These participants were drawn from other experiments that were ongoing in the laboratory, and were people who had extra time to fill in order to earn the amount of credit being awarded. Thus, these participants were tested individually as well. They rated each statement on a 9-point scale with an origin of 0 representing neutral. The scale was anchored on either end with 4s denoting very typical of a male and a female, respectively. From these ratings we could clearly identify 24 statements that were more associated with a male (e.g., I belittled a co-worker in a meeting; I swore at the guy who insulted me), 24 that were associated with a female (e.g., I hug old friends when I see them; I made a centrepiece for the dining table), and 24 that were clearly neutral (i.e., equally likely to come from a male or a female, for example, I am very easy going). The 72 stimuli from these three categories were used in this experiment and those that follow.

Software was written that presented 24 statements from each of two gender-neutral names (Chris and Pat), for a total of 48 statements to be learned during encoding. Each of the two sources spoke eight male-oriented statements, eight female-oriented statements, and eight neutral statements. Each item was presented in the centre of the computer monitor for 5 seconds. The speaker's name was presented two lines above the statement. For a given participant, the 8 items

from each of the three classes of items for a given speaker were drawn randomly anew from the 24 items in any given class. In this fashion, during learning each of the two sources had equal numbers of male, female, and neutral statements. This procedure left 24 items (8 in each of the three classes) to serve as distractor items on the source-monitoring test. Participants had been instructed to learn the statements and who spoke them for a later, unspecified memory test. After the statements were presented, participants were asked to work on a puzzle task for 5 minutes. Following this task, they were given two short descriptions of Chris and Pat. In the heterosexual condition, Chris was described as a heterosexual male and Pat as a heterosexual female. To add further credence to the orientation manipulation the male was described as enjoying sports and the female as enjoying keeping her apartment neat. For consistency, the gender of the sources was preserved in the homosexual condition (Chris as male, Pat as female). In this case, the gay male was depicted as enjoying the arts and the lesbian female was depicted as being rather competitive. Care was taken to ensure that statements given during the learning phase would not correspond to any of the features described in the stereotype information provided before the test. Therefore, the two conditions differed only in the sexual orientation information provided after learning but before the source-monitoring test.

We used a standard source-monitoring test. All 72 statements were randomised anew and presented individually in the centre of the computer monitor. Participants had to press one of three keys to designate that Chris made the statement, Pat had done so, or the statement was brand new.

## Results and discussion

Unless specified otherwise by a *p* value, the probability of a Type I error does not exceed chance by 5% throughout this article. In each experiment, we first consider the inferred recognition hit rate (i.e., the proportion of old items claimed to be old) and then consider the source-monitoring data. As it relates to source-monitoring performance, for each of the two conditions we could potentially report 27 means (54 total per experiment), which would be rather burdensome for readers to digest. Consequently, we have chosen to distil the data by reporting conditionalised source-monitoring scores (CSIMs) in

most cases (cf. Bayen, Murnane, & Erdfelder, 1996). These scores represent for a given item type and for a given source (e.g., masculine items spoken by the male) the average proportion of items that were correctly attributed to the source, conditionalised on the total numbers of items of that type correctly identified as old. Table 1 summarises the recognition hit rates and Table 2 summarises the average CSIMs. Readers who are knowledgeable about the use of CSIM scores will know that they are often used when the rates of identifying old items as old differ between sources and classes of items.<sup>1</sup>

In Table 1 the first two columns correspond to the average hit rate of masculine items spoken by the male and female speakers, respectively. Thus, 90% of the masculine items spoken by the male were called old and 88% of the masculine items spoken by the female were called old. Similarly, the second two columns correspond to the average hit rate of feminine items spoken by the male and female, respectively. The data were analysed with a 2 (condition: heterosexual vs homosexual)  $\times$  2 (item type: masculine vs feminine)  $\times$  2 (speaker: male vs female) mixed-model analysis of variance (ANOVA) with repeated measures on the last two factors. Except for an approximately 2.5% higher hit rate for the feminine items as compared to the masculine items, all main effects and interactions were not statistically significant. To anticipate the other experiments, this one effect is not present in the subsequent three experiments and probably represents a small, spurious outcome. Consequently, the equivalent hit rates alleviate any concerns about the interpretation of the source-monitoring data, which is discussed next.

Table 2 is arranged identically to Table 1. The first two columns correspond to the average rate at which masculine items were correctly attributed to the male and female speakers, respectively. Thus, in the heterosexual condition 83% of the masculine items spoken by the male were correctly attributed to him, whereas only 65% of the masculine statements spoken by the female were correctly attributed to her. Similarly, the second two columns correspond to the rate at which the feminine items spoken by the male and female

<sup>1</sup> Although this approach can be questioned, in none of the important comparisons was that the case in the present experiment. Had we presented the full complement of data in all of its minutiae, the conclusions drawn would have been the same.

**TABLE 1**  
Inferred recognition hit rates by condition in Experiments 1–4

<i>Experiment and condition</i>	<i>Masculine items</i>		<i>Feminine items</i>	
	<i>Male</i>	<i>Female</i>	<i>Male</i>	<i>Female</i>
<i>Experiment 1</i>				
Heterosexual	.90	.88	.92	.92
Homosexual	.87	.87	.88	.92
	<i>Homosexual</i>	<i>Heterosexual</i>	<i>Homosexual</i>	<i>Heterosexual</i>
<i>Experiments 2 and 3</i>				
Two men				
Full	.85	.87	.87	.89
Divided	.83	.88	.89	.84
Two women				
Full	.81	.86	.88	.88
Divided	.80	.83	.85	.87
	<i>Male</i>	<i>Female</i>	<i>Male</i>	<i>Female</i>
<i>Experiment 4</i>				
Control	.82	.84	.89	.84
Negation	.82	.87	.89	.87

speakers were correctly attributed to their speaker. In an identical ANOVA model to that described for the recognition hit rates, the three-way interaction was statistically significant,  $F(1, 64) = 31.42$ . When depicted as heterosexual, masculine items were more often correctly attributed to the male and fewer masculine

statements spoken by the female were correctly attributed to her,  $t(27) = 2.61$ . The reverse was true for the feminine items. Feminine items spoken by the male were correctly attributed to him less often than those spoken by the female,  $t(27) = 4.24$ . This pattern of results strongly suggests that people will use standard gender stereotypes in

**TABLE 2**  
Average conditionalised source-monitoring score by condition in Experiments 1–4

<i>Experiment and condition</i>	<i>Masculine items</i>		<i>Feminine items</i>	
	<i>Male</i>	<i>Female</i>	<i>Male</i>	<i>Female</i>
<i>Experiment 1</i>				
Heterosexual	.83	.65	.69	.86
Homosexual	.61	.82	.85	.60
	<i>Homosexual</i>	<i>Heterosexual</i>	<i>Homosexual</i>	<i>Heterosexual</i>
<i>Experiments 2 and 3</i>				
Two men				
Full	.67	.80	.79	.62
Divided	.53	.65	.72	.49
Two women				
Full	.81	.66	.64	.83
Divided	.68	.60	.52	.76
	<i>Male</i>	<i>Female</i>	<i>Male</i>	<i>Female</i>
<i>Experiment 4</i>				
Control	.54	.84	.74	.60
Negation	.72	.89	.87	.69

determining the origin of their memories, at least when the content of that memory is gender oriented.

When the speakers were depicted as homosexual, the pattern of results was just the opposite. Masculine items spoken by the lesbian female were more frequently correctly attributed to her (.82) than were the masculine items spoken by the gay male (.61),  $t(28) = 2.92$ . By contrast, the feminine items spoken by the gay male were more often correctly attributed to him than these same items were correctly attributed to the lesbian female,  $t(27) = 4.51$ . Therefore, the pattern of results in this condition suggests that participants used an alternative set of stereotypes when judging who spoke which statements.

A similar pattern was observed when new items were incorrectly claimed to have been studied, but the effects were attenuated by very low false alarm rates. In the heterosexual condition, masculine items were more often erroneously attributed to the male source than to the female (.10 vs .05),  $t(27) = 1.79$ ,  $p = .08$ . By contrast, feminine items were more often erroneously attributed to the female than to the male (.07 vs .02),  $t(27) = 1.91$ ,  $p = .06$ . The reverse was true with a homosexual depiction of the speakers. Masculine items were more often falsely attributed to the female than to the male (.12 vs .04) whereas feminine items were falsely attributed to the male more frequently than to the female (.13 vs .05), both  $t(27) > 2.75$ . Therefore, both sets of stereotypes affected old and new items similarly. These same patterns of bias were not generally observed for neutral items. In the heterosexual condition, CSIM scores were equivalent for the male and for the female (.75 vs .76 respectively). There was one odd asymmetry in the homosexual condition for these neutral items. Source monitoring was marginally better for the items spoken by the male (.77) than by the female (.66),  $t(27) = 2.94$ ,  $p = .06$ . We are unsure exactly why this anomaly occurred, but it has been suggested to us that what may be considered entirely neutral for a heterosexual orientation (as normed) could be construed differently under the homosexual orientation.

Because gender was not known at encoding, the results from this experiment clearly demonstrate that gender and orientation stereotypes biased source-monitoring decision processes. Statements that were consistent with gender orientation were more often correctly attributed to the original speaker, and statements that were inconsistent were less frequently attributed to the

original speaker. As we noted earlier, the detection of old items as old was equivalent across sources, and therefore when a source misattribution was made it was attributed to the other speaker. In contrast to Sherman and Bessenoff (1999), divided attention was not required to observe biased source monitoring based on stereotypes in younger adults.

## EXPERIMENTS 2 AND 3

One important issue is whether the biases observed in Experiment 1 are caused by automatic influences or are more consciously controlled (i.e., strategic). Theoretically, the change from using a natural gender stereotype when knowledge about sexual orientation is available could suggest that the application of schematic and stereotypic knowledge is a conscious process requiring attention. This point is nontrivial because many source-monitoring decisions are depicted as being made effortlessly and heuristically (Johnson et al., 1993). However, systematic plausibility checks have been depicted as conscious and effortful (see Chaiken, Lieberman, & Eagly, 1989; Johnson & Raye, 1981). On the other hand, Sherman and Bessenoff (1999) claimed that the skinhead versus priest schemas were only used when cognitive resources were depleted with a divided attention manipulation. That result implies that schematic information is used more automatically in that case.

More generally there is no agreement on whether the influence of stereotypes is an automatic or conscious process. Hense, Penner, and Nelson (1995) investigated age-related stereotypes and concluded mainly that they were applied more automatically. More recent work by Sherman and his colleagues argues that the application of stereotypes is a relatively automatic process, especially when participants are placed under conditions of divided attention which reduces their capacity to use recollection (Sherman, Groom, Ehrenberg, & Klauer, 2003). In contrast, Bayen et al. (2000) believe that the application of stereotypes is a conscious strategy that affects memory judgements via guessing strategies. In the absence of veridical source information, participants guess consistently in favour of the stereotype. These two opposing views of how stereotypes influence source judgements may be a consequence of derived estimates of performance. Sherman et al. used process dissociation models

and inferred that reliance on familiarity indicates an automatic influence on memory. By contrast, the model used by Bayen et al. yields only memory and guessing parameters but does not indicate whether guessing is influenced by automatic or strategic influences. Because there appears to be some debate about when younger adults will most often use schematic information, the next two experiments manipulated full versus divided attention at test. If the pattern of results is the same under full and divided attention (or even exacerbated), then gender and orientation stereotypes are used more automatically to make source judgements. If the biases are attenuated or absent under divided attention, then application of the stereotypes requires conscious processing.

Another purpose of these two experiments was to explore the generality of the bias caused by the gender and orientation stereotypes. If these stereotypes are applied automatically, then participants should be able to use both sets of stereotypes concurrently, and do so even under divided attention. This situation might arise quite naturally when the source discrimination required is between two males (or two females) who differ only in their sexual orientation. Based on the results of Experiment 1, masculine items may be attributed to the heterosexual male (and to the lesbian female) whereas the feminine items may be attributed more often to the gay male (and to the heterosexual female). Whereas participants in Experiment 1 had to use either the gender or sexual orientation stereotypes, participants in this experiment were required to consider both. Experiment 2 contrasted source monitoring for two men (one heterosexual male and one gay male) under either full or divided attention. Experiment 3 did the same for a heterosexual versus lesbian female as the two sources. Because the data were collected at slightly different times, we analysed the two experiments separately, but report them together for brevity.

## Method

*Participants.* A total of 57 University of Georgia undergraduates volunteered in exchange for partial credit towards a course requirement in Experiment 2, and 54 from the same pool were tested in Experiment 3. Each participant was tested individually in sessions that lasted approximately 30 minutes. In each experiment, equal numbers of participants were assigned to the full

versus divided attention conditions (29 and 27, respectively for Experiments 2 and 3).

*Procedure.* This experiment was identical in all of its essential properties to Experiment 1, with the exception of how the speakers were depicted before the source-monitoring test and the use of a divided attention manipulation at test for half of the participants. In Experiment 2, participants were told that Chris and Pat were both males. Like the homosexual condition of Experiment 1, Chris was depicted as the gay male whereas Pat was described as a heterosexual male. The brief depictions used in Experiment 1 were used again. In Experiment 3, participants were told that Chris and Pat were both females. In this case, Pat was depicted as a lesbian female and Chris as a heterosexual female. After learning, a 5-minute distractor ensued in the full attention conditions and then the person descriptions were introduced. There was no distractor task in the divided attention condition, rather the participants learned how to perform the divided attention task.

For the divided attention task, random number generation (RNG) was used because it can have a rather deleterious effect on memory (e.g., Hicks & Marsh, 2000; Marsh & Hicks, 1998). Participants were informed that a metronome would be played from a cassette tape, and at each beep they should issue out loud one random number between 1 and 10, inclusive. The metronome speed for this experiment was 1 tone every second, which in our experience is a very demanding pace. They were given detailed instructions on what it means to be random, including avoiding stereotypic sequences such as ascending or descending sequences (7, 8, 9 or 6, 4, 2, etc.). They were also instructed that to be truly random, they had to use each number about equally often in the long run. They were then given practice, during which the experimenter recorded their performance for a complete set of at least 100 random digits (i.e., a baseline) to be compared with their performance when performed concurrently with the source-monitoring task. After practice, the descriptions of the two sources were administered in the divided attention conditions. At test, the cassette tape was started 20 seconds prior to initiating the test itself. The experimenter recorded another 100 digits while the participant was concurrently performing the source test. The options on the source test were identical to Experiment 1.

## Results and discussion

The inferred recognition hit rates are displayed in the middle portion of Table 1. Note that the left two columns still reflect performance on the masculine items and the right two columns reflect source monitoring for the feminine items. However, the subcolumn headings have changed to reflect whether the speaker was depicted as heterosexual or homosexual. For example, under full attention in Experiment 2 (two male speakers) the hit rate for masculine items spoken by the gay male was .85 whereas masculine items spoken by the heterosexual male had an average hit rate of .87. In neither experiment (separately or pooled) were there any statistical differences as function of item, condition, speaker, or condition. The absence of hit rate differences under divided attention versus full attention conditions has been assumed by many to be the usual outcome (but see Hicks & Marsh, 2000).

The source-monitoring results are summarised as average CSIM scores in the middle portion of Table 2. In the 2 (full vs divided)  $\times$  2 (masculine vs feminine items)  $\times$  2 (orientation) ANOVA for Experiment 2, there was a main effect of dividing attention, reflecting worse source monitoring with concurrent RNG,  $F(1, 56) = 8.81$ . There was also an orientation by type of item interaction,  $F(1, 56) = 17.88$ . No other terms in the model were statistically significant. The interaction indicates that the biases found in Experiment 1 are preserved in this experiment, and that they survive even divided attention. Masculine statements were correctly attributed to the heterosexual male more often than they were correctly attributed to the gay male, full:  $t(28) = 2.12$ , divided:  $t(28) = 2.19$ . By contrast, more feminine statements spoken by the gay male were attributed to him than were correctly attributed to the heterosexual male, full  $t(28) = 2.69$ , divided  $t(28) = 3.13$ .

In Experiment 3, participants were told that the speakers were two females who differed in their sexual orientation. The outcome of the three-way ANOVA model was identical to Experiment 2. Dividing attention reduced source-monitoring performance,  $F(1, 52) = 4.16$ . The orientation by type of item interaction was the only other term that was statistically significant,  $F(1, 52) = 24.62$ . Masculine statements were correctly attributed to the lesbian female at a higher rate than to the heterosexual female, full  $t(26) = 2.38$ , but the effect was only numerically present under divided attention  $t(26) = 1.56$ ,  $p = .11$ . The reverse was true

for feminine statements, full  $t(26) = 3.78$ , divided  $t(26) = 4.27$ . Therefore, across the two experiments, divided attention did not generally change performance, although it did reduce it to a different, lower range in the response scale. Consequently, we conclude that these biases are driven by early, automatic selection of a source based on the semantic content of a memory trace. Had they been more consciously driven, a substantial disruption in source monitoring based on stereotypes would have been observed under conditions of divided attention. A measure of randomness called RNG can be calculated (see Hicks & Marsh, 2000). Higher magnitudes indicate less random performance. Baselines measures of RNG were lower (Exp. 2: .29; Exp. 3: .28) than when the task was performed concurrently with the source test (Exp. 2: .34; Exp. 3: .36), smaller of the two  $t$  values 2.76. Thus, participants were not only less random, their source monitoring suffered as well.

Under full attention in Experiment 2, there were virtually no false alarms to feminine new items (.04) but numerically they were more often attributed to the gay male (.03). False alarms to the masculine statements were more frequent (.11) and they were mainly attributed to the heterosexual male (.07). Under divided attention, the false alarm rate was greater. For feminine items, numerically more were attributed to the gay male (.19) than the heterosexual male (.15). For masculine new items under divided attention, more were attributed to the heterosexual male (.29) than the gay male (.14). Only this last comparison of the lot was statistically significant,  $t(28) = 3.10$ . For Experiment 3, again no substantive biases were observed in erroneously claiming that new items were spoken by someone in these conditions. Masculine items were numerically more often associated with the lesbian female as opposed to the heterosexual female (.09 vs .06, respectively) whereas the feminine new items were more often associated with the heterosexual rather than the lesbian female (.06 vs .04, respectively). Under divided attention, equivalent false alarms were attributed to each source for masculine items (.14 vs .17), but more feminine items were attributed to the heterosexual female than the lesbian female (.19 vs .11), albeit at a marginal level of significance,  $t(26) = 1.86$ ,  $p = .07$ . Across the two experiments, false alarms were rare, except under divided attention. When attention was divided and two men were being considered, false alarms to masculine items differed; and when

two women were being considered, false alarms to feminine items differed (albeit marginally).

Correct attributions for neutral items in both experiments were statistically equivalent for the heterosexual and the homosexual sources under full attention (Exp. 2: .76 vs .77; Exp. 3: .72 vs .68). Under divided attention, however, neutral items were more often correct when they were associated with the gay male as compared with the heterosexual male (.66 vs .54),  $t(28) = 2.10$ . A similar asymmetry occurred in Experiment 3 where neutral items associated with the heterosexual female were more accurate than the lesbian female's neutral items (.74 vs .62), albeit a marginal result,  $t(26) = 1.93$ ,  $p = .06$ . Given that these effects under divided attention are not consistent with one another, and one of the two results is statistically marginal, we conclude that the biases are not necessarily found with memories for neutral material. Rather, biases are much stronger for those statements whose semantic content participants associate with gender and orientation stereotypes.

#### EXPERIMENT 4

The purpose of this last experiment was to ascertain whether we could change participants' strategies on the source-monitoring test. In particular, we wanted to change their decision criteria to rely less on the stereotypes that participants were using in the previous experiments. In order to negate the stereotypes, we asked participants in one condition to write down all the reasons why they believed that stereotypes were harmful in today's society. Our intuition was that this task would act as an adequate debiasing manipulation in order to alert participants that stereotypes are not always reliable or informative. As a control comparison, we had a different group of participants write down all the ways they could devise to improve problems at the university. Both conditions performed their respective writing tasks after learning and before the stereotypes were provided. If the negation manipulation has the intended effect of changing source-monitoring decision criteria, then correct source attributions should increase as compared to the control condition, where participants might be relying automatically on the stereotype information to a greater degree.

From a theoretical standpoint, this experiment serves as a test of whether late correction can

overcome early selection (Jacoby, Kelly, & McElree, 1999). The results of Experiments 1–3 suggest an automatic selection of source based on the semantic content of memory. Jacoby's logic of opposition argues very strongly that only recollective processes that come late, or at least run in parallel, can overcome more automatic influences of memory. The early selection of a source more automatically in Experiments 1–3 can potentially be corrected in the debiasing condition through a late correction of searching for more recollective details in memory before issuing a source-monitoring judgement. Therefore, the underlying approach used in this experiment is similar to the theory advocated by Jacoby and Hollingshead (1990); namely, consciously controlled processing can be used to edit or correct the output of more automatically driven influences of memory. Banaji and Greenwald (1995) attempted a similar manipulation and found it very difficult to remove the effects of stereotypes, and consequently we did anticipate that the size of the effect may be slightly attenuated. We consider this issue more deeply in the General Discussion section after reporting only the important findings from the present experiment.

#### Method

*Participants.* A total of 45 undergraduates at the University of Georgia volunteered in exchange for partial credit towards a course research requirement. Each participant was tested individually in sessions that lasted approximately 30 minutes. A total of 22 volunteers were tested in the control condition and 23 were tested in the negation condition.

*Procedure.* This experiment was identical to the previous experiments (under full attention) except in the following two respects. First, Chris was depicted as a gay male and Pat was depicted as a lesbian female (i.e., the homosexual condition from Experiment 1). We chose this combination of sources because we thought it might maximise observing a change in source-monitoring decision processes. Second, rather than performing a 5-minute distractor activity, participants performed the puzzle task for only 2 minutes. They were then handed a piece of paper with numbered blank lines. In the control condition they were asked to list ways to improve problems or shortcomings at the university. In the negation condition they were

asked to list reasons why they believed that stereotypes were harmful in today's society. Both groups were informed that they would have 3 minutes to perform their respective tasks. Timing was controlled by a hand-held stopwatch. After the listing task was completed, the profiles of Chris and Pat were introduced. The source-testing phase ensued just as it did in the previous experiments.

## Results and discussion

As in the previous experiments, the inferred recognition hit rates were statistically indistinguishable from another (see Table 1). The average CSIM scores are presented in the lower portion of Table 2. Note that the subheadings for this experiment are identical to those of Experiment 1. The data were analysed using a 2 (person source)  $\times$  2 (item type)  $\times$  2 (condition) ANOVA. As predicted, participants in the negation condition performed much better on the source test than those in the control condition,  $F(1, 43) = 7.31$ . This outcome indicates that participants who were asked to write about stereotypes used more stringent decision criteria and inspected memories more carefully during the test. The source by item type interaction was statistically significant, indicating that both the control and negation conditions displayed patterns of biased source-monitoring performance that would be expected based on the previous experiments,  $F(1, 43) = 20.71$ . In reduced ANOVA models for each condition separately, the control condition who wrote about solutions to university problems gave rise to the expected two-way interaction between item type (masculine vs feminine) and gender (male vs female),  $F(1, 21) = 8.92$ . The same was true in the negation condition,  $F(1, 22) = 17.80$ . There was also a source by condition interaction in the full three-way ANOVA model that emanated from better performance in assigning masculine items to the gay male after negation as compared with the control condition. Stated slightly differently, the negation manipulation attenuated the bias for the gay male, but not the lesbian female. In sum, writing about the way in which stereotypes can be harmful did not cause participants to completely abandon the strategy of consulting stereotypes as a means of determining the origin of a memory, but it did significantly attenuate the bias for the gay male.

The overall better source monitoring after the negation manipulation suggests that participants were trying to inspect memory much more closely, as would be predicted by a late correction strategy (Jacoby et al., 1999). Further evidence for this fact was observed in claims made for the masculine and feminine new items. Numerically fewer of these new statements were erroneously ascribed to a speaker after the negation manipulation (.14) than were labelled old in the control condition (.20), but this outcome did not reach statistical significance,  $t(43) = 1.26$ . The improvement in CSIM scores was also observed on the neutral statements spoken by the two sources. The average CSIM score for neutral items in the control condition was .71 whereas after the negation manipulation it was .81,  $t(43) = 2.61$ . Together these outcomes suggest that reliance on the gender and orientation stereotypes may be weakened but not completely eliminated, at least with the negation manipulation used here. The clear improvement in source memory in the negation condition is consistent with a more stringent change in source-monitoring decision criteria, because the negation manipulation was introduced after the statements were presented at encoding. Of course, our interpretation of a late correction account may be in error. One possibility is that the stereotypes become deactivated as a consequence of the negation manipulation.<sup>2</sup> One counter-argument is that the negation manipulation was for participants to write about the stereotypes, so the stereotypes should if anything be more strongly activated in the negation condition, not less activated. Nevertheless, the deactivation account remains because we have no empirical evidence that speaks to this issue.

## GENERAL DISCUSSION

Collectively, the experiments in this study are consistent with previous reports that the content of a memory can bias attributions about its origin. Statements that were consistent with a male or a female stereotype were more accurately attributed when the gender of the speaker matched the content of the statement. During learning, participants were not provided with the gender of the speakers and there were equal numbers of masculine, feminine, and neutral statements uttered by both sources. As Bayen et al. (2000) and Hicks

<sup>2</sup>We thank an anonymous reviewer for raising this point.

and Cockman (2003) have argued, therefore, the results observed here are likely to reflect a retrieval bias in the absence of a clear recollection for which source spoke a particular statement. Obviously participants had the propensity to attribute statements to the most likely candidate based on both gender and sexual orientation stereotypes. This approach worked well for statements that were indeed associated with a particular source, but detrimentally impacted source attributions when the statement originally came from the stereotype-inconsistent source. The fact that the bias could be reversed simply by activating the homosexual stereotypes lends further credence to a retrieval bias account of the results.

From a theoretical standpoint, we believe this study demonstrates that the use of stereotype and schematic information can be a relatively automatic and heuristic process, at least with the types of stereotypes used here (cf. Devine, 1989). After all, the biases persisted even under extremely demanding divided attention at test. We believe these effects are consistent with an early selection of the source that has been biased by the semantic content of the item being considered (in the absence of recollecting the actual source). By early selection, Jacoby et al. (1999) mean a response is selected that is based on habit or semantic knowledge, and they argue that these early selection processes can actually constrain what memories or memorial details actually come to mind. In the present case, gender-stereotypic semantic details of any given test item may guide early selection of the schema-consistent source because contradictory source information is not cued by the semantic content of the statement. Only after the negation manipulation was the early selection component partially opposed with a late correction of, presumably, more conscious recollective assessment of additional details associated with the memory. Although early selection is not necessarily a purely automatic process, and could be augmented by early conscious processing (because no task is process pure), the absence of any effect of divided attention in Experiments 2 and 3 strongly suggests that if an early selection is operative then it must necessarily be a largely automatic process.

The early selection, late correction model is not synonymous with a "response plus correction" model. In addition, the early selection mechanism is not an uninteresting response bias necessarily due to guessing. Early selection is a constraint on

the retrieval of information associated with a memory that has gender-biased semantic content. Obviously, our appeal to the early selection and late correction model awaits more rigorous empirical scrutiny. But other than embedded in the source-monitoring component of the process dissociation procedure that has received so much debate, we know of no other empirical report that associates schema-based source attributions to the early selection and late correct model advocated by Jacoby and his colleagues.

The early selection bias that we are arguing occurs with the gender and orientation schemas is similar to other heuristic biases that arise when the candidate sources differ greatly in their memorability (e.g., Johnson, Raye, Foley, & Foley, 1981; Marsh & Bower, 1993). For example, the it-had-to-be-you effect stems from participants believing that their false alarms are more often associated with an external source as opposed to something that they had generated themselves. This effect can be reversed to produce the it-had-to-be-me effect by making the external source much stronger and more memorable than the internally generated material (e.g., Bink, Marsh, & Hicks, 1999; Hoffman, 1997). However, the sources used in the present study did not differ in their overall memorability, and consequently the basis for biases is likely to be different. The early selection processes that bias and prevent retrieval of certain memorial details for gender-specific material appear to us to be an attractive theoretical alternative at this juncture.

Of course, a different interpretation is that our data are consistent with the conscious application of the stereotypes that bias guessing (e.g., Bayen et al., 2000). As years of work on memory show, guessing biases always increase under divided attention because recollection is compromised. If results from manipulating stereotypes were due solely to guessing, then the effects in Experiments 2 and 3 should have been greatly exacerbated, but they were not. In fact, the effect was left generally unchanged in the divided attention condition as compared to the full attention condition. Moreover, the absence of greater guessing biases was replicated across both experiments. Those results suggest that stereotypes do not merely bias guessing strategies in a deliberately conscious fashion. Although both experiments could represent Type II error, we believe that two such outcomes are rather unlikely. But if they were, then an account based on guessing would still be viable. Instead, our results are most consistent with Sherman et

al.'s (2003) conclusions that stereotypes have both conscious (Experiment 4) and automatic (Experiments 1–3) influences on source monitoring.

More generally, the attribution biases observed in the present study highlight the complexity of source-monitoring decision processes. In the present case, the content of the memory itself was used to decide in which context it was encountered. Of course, one could argue that the sources used in the present study were contextually impoverished and participants may have had little other information on which to base their source judgements. However, this is exactly the state one would be in if one had impoverished source memory in the first place. As such, we are not arguing that content of a memory will always affect source monitoring, only that it can be used as one alternative method of determining the origin of a memory. As we argued earlier, consulting schematic and stereotypic information is probably quite adaptive as a means of narrowing down potential candidate sources. However, this would only be the first step, and further plausibility checks for choosing among the candidates would then usually be required. Although we did not study such a process in the current experiments, one could imagine an extension of the Dodson et al. (1998) paradigm in which participants discriminated among two males and two females concurrently. The partial information associated with statements may be trumped by stereotypic information about sexual orientation if one male and one female were described as homosexual. Based on the results of Experiment 4, we believe that conscious decision processes can override more heuristic processes as specified in a late correction component of memory retrieval.

The present study demonstrates quite clearly that conditions of divided attention at test are not necessary to observe conformity to stereotypes in source monitoring (Sherman & Bessenoff, 1999). Why Sherman and Bessenoff required divided attention is unclear, but it may be a function of the strength of the stereotype. Contrasting a priest with a skinhead may be less natural for people than contrasting people based on gender. Weaker stereotypes based on, say, age may not evidence the strong biases demonstrated here (e.g., Hense et al., 1995). Mather et al. (1999) used subschemas of a writer and an athlete (in addition to Republican vs Democrat) and found very little bias based on this manipulation. One reason may be that memory lacks a strong pre-existing contrast

between those personality types. According to the theory put forth here, early selection may not operate with weaker stereotypes, in the same way that weak retrieval cues do not elicit information from memory very efficiently. Weak stereotypes do not guide and constrain retrieval in the same way that stronger ones do. By this analysis, there would be no need for a late correction process with weak stereotypes because no bias from early selection exists. Like strong retrieval cues, strong stereotypes can sometimes strongly cue a correct source or act as a retrieval blocker when the source is schema-inconsistent with the content of memory (cf. Smith & Tindell, 1997, on retrieval blocking). More generally, the early selection processes are efficient for determining the origin of memory only in those cases when the semantic information is consistent with general knowledge. When the two are inconsistent, source-monitoring errors are likely to occur, and performance will be better when a more effortful late correction process is engaged.

Manuscript received 22 April 2004

Manuscript accepted 8 February 2005

PrEview proof published online 21 June 2005

## REFERENCES

- Banaji, M. R., & Greenwald, A. G. (1995). Implicit gender stereotyping in judgements of fame. *Journal of Personality & Social Psychology*, *68*, 181–198.
- Bardach, L., & Park, B. (1996). The effect of in-group/out-group status on memory for consistent and inconsistent behavior of an individual. *Personality & Social Psychology Bulletin*, *22*, 169–178.
- Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 197–215.
- Bayen, U. J., Nakamura, G. V., Dupuis, S. E., & Yang, C. (2000). The use of schematic knowledge about sources in source monitoring. *Memory & Cognition*, *28*, 480–500.
- Bem, S. L. (1981). Gender schema theory: A cognitive account of sex typing. *Psychological Review*, *88*, 354–364.
- Bink, M. L., Marsh, R. L., & Hicks, J. L. (1999). An alternative conceptualization to memory “strength” in reality monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 804–809.
- Chaiken, S., Lieberman, A., & Eagly, A. H. (1989). Heuristic and systematic information processing within and beyond the persuasion context. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 212–252). New York: Guilford Press.

- Cook, G. I., Marsh, R. L., & Hicks, J. L. (2003). Halo and devil effects demonstrate valenced-based influences on source-monitoring decisions. *Consciousness & Cognition, 12*, 257–278.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality & Social Psychology, 56*, 5–18.
- Dodson, C. S., Holland, P. W., & Shimamura, A. P. (1998). On the recollection of specific- and partial-source information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1121–1136.
- Hense, R. L., Penner, L. A., & Nelson, D. L. (1995). Implicit memory for age stereotypes. *Social Cognition, 13*, 399–415.
- Hicks, J. L., & Cockman, D. W. (2003). The effect of general knowledge on source memory and decision processes. *Journal of Memory & Language, 48*, 489–501.
- Hicks, J. L., & Marsh, R. L. (2000). Toward specifying the attentional demands of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1483–1498.
- Hicks, J. L., Marsh, R. L., & Ritschel, L. (2002). The role of recollection and partial information in source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 503–508.
- Hoffman, H. G. (1997). Role of memory strength in reality monitoring decisions: Evidence from source attribution biases. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 371–383.
- Jacoby, L. L., Debner, J. A., & Hay, J. F. (2001). Proactive interference, accessibility bias, and process dissociations: Valid subject reports of memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 686–700.
- Jacoby, L. L., & Hollingshead, A. (1990). Toward a generate/recognize model of performance on direct and indirect tests of memory. *Journal of Memory & Language, 29*, 433–454.
- Jacoby, L. J., Kelly, C. M., & McElree, B. (1999). The role of cognitive control: Early selection and late correction. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology*. New York: Guilford Press.
- Jennings, J. M., & Jacoby, L. L. (1993). Automatic versus intentional uses of memory: Aging, attention, and control. *Psychology & Aging, 8*, 283–293.
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*, 3–28.
- Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88*, 67–85.
- Johnson, M. K., Raye, C. L., Foley, H. J., & Foley, M. A. (1981). Cognitive operations and decision bias in reality monitoring. *American Journal of Psychology, 94*, 37–64.
- Marsh, R. L., & Bower, G. H. (1993). Eliciting cryptomnesia: Unconscious plagiarism in a puzzle task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19* (3), 673–688.
- Marsh, R. L., & Hicks, J. L. (1998). Test formats change source-monitoring decision processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1137–1151.
- Mather, M., Johnson, M. K., & De Leonardis, D. M. (1999). Stereotype reliance in source monitoring: Age differences and neuropsychological test correlates. *Cognitive Neuropsychology, 16*, 437–458.
- Qin, J., Raye, C. L., Johnson, M. K., & Mitchell, K. J. (2001). Source ROCs are (typically) curvilinear: Comment on Yonelinas (2001). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 1110–1115.
- Sherman, J. W., & Bessenoff, G. R. (1999). Stereotypes as source-monitoring cues: On the interaction between episodic and semantic memory. *Psychological Science, 10*, 106–110.
- Sherman, J. W., Groom, C. J., Ehrenberg, K., & Klauer, K. C. (2003). Bearing false witness under pressure: Implicit and explicit components of stereotype-driven memory distortions. *Social Cognition, 21*, 213–246.
- Smith, S. M., & Tindell, D. R. (1997). Memory blocks in word fragment completion caused by involuntary retrieval of orthographically related primes. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 23*, 355–370.
- Spaniol, J., & Bayen, U. J. (2002). When is schematic knowledge used in source monitoring? *Journal of Experimental Psychology: Learning, Memory, & Cognition, 28*, 631–651.