Journal of Experimental Psychology: Applied
2005, Vol. 11, No. 3, 200–209

1076-898X/05/$12.00 DOI: 10.1037/1076-898X.11.3.200

Social Recognition Memory: The Effect of Other People's Responses for Previously Seen and Unseen Items

Daniel B. Wright, Sorcha A. Mathews, and Elin M. Skagerberg
University of Sussex

When people discuss their memories, what one person says can influence what another person reports. In 3 studies, participants were shown sets of stimuli and then given recognition memory tests to measure the effect of one person's response on another's. The 1st study ($n = 24$) used word recognition with participant–confederate pairs and found that the effect of confederate responses on participant responses was larger for previously unseen items than for previously seen items ($\omega_p = .23$). This finding was replicated in the 2nd study, which used photographs of cars ($n = 24$). In the 3rd study ($n = 54$), which used photographs of faces with participant pairs, the effect was also larger for unseen items. Results indicate that people rely more on other people's memories for unremembered objects than for remembered objects. This is important for both theories of memory and applications (e.g., witnesses talking, students studying together).

Keywords: memory, misinformation, eyewitness, conformity

When two people talk about a past event, what one person says can influence what the other person reports about the event. As an analogy to Asch's (1955) work on influencing responses on perceptual judgment tasks, this is often called *memory conformity* (Gabbert, Memon, & Allan, 2003; Gabbert, Memon, Allan, & Wright, 2004; Wright, Self, & Justice, 2000). It is also referred to as *social contagion of memory* (Meade & Roediger, 2002; Roediger, Meade, & Bergman, 2001), and as an effect of *cowitness information* (Shaw, Garven, & Wood, 1997). However, none of these phrases perfectly captures the phenomenon we explore. We explore conformity of memory responses and more particularly recognition memory. We use the phrase *social recognition memory*, although many of the processes involved are also relevant to free-recall tasks.

Understanding social recognition memory is important for both theories and applications of memory. This research explicitly uses explanations from both social and cognitive psychology. Social psychological explanations are becoming increasingly important in memory research (Conway & Pleydell-Pearce, 2000; Kihlstrom, 2002; Weldon, 2000). Regarding applications, in education, eyewitness testimony, and much everyday reminiscence, people discuss past events with other people. When students, eyewitnesses, and friends describe their memories, this influences what they later report, though these effects depend on how they discuss the event. For example, Marsh, Tversky, and Hutson (2005) showed people a 7-min film clip. Some of the participants were placed in pairs and told either to talk about the factual elements of the film or to talk about their emotional reaction to the film. Another group of participants did not discuss the event. Later, all participants recalled the film individually. They found that the people in groups recalled approximately 10% more of the events. More interesting however were the differences between the groups of pairs. The factual group had less than half the number of self-references and approximately half the number of what they classified as "major" errors compared with the emotion pair.

Two different types of research have examined the negative aspects of discussing an event with others. The first type involves presenting people with the same set of stimuli and then having them free recall the items either individually or in groups (e.g., Basden, Basden, & Henry, 2000; Clark, Hori, Putnam, & Martin, 2000; Weldon & Bellinger, 1997; Wright & Klumpp, 2004). The finding is that groups generally recall less together than they would if the individual members of the groups had recalled separately. For example, in Wright and Klumpp (2004) the pairs recalled an average of 47% of the words they were previously shown. Though individuals, on their own, recalled an average of only 41%, when they are put into pairs and only nonredundant recollections are counted, the average of these nominal pairs (63%) is higher than the average for the actual pairs. This is called *collaborative inhibition*. Because free recall is used and what people recall is generally accurate, researchers in this area tend to look at the amount of the correct recall and tend not to focus on errors (exceptions include Basden, Reysen, & Basden, 2002). Collaborative inhibition occurs with accurate information. The second type of research, which is the focus of this article, focuses on what happens when somebody reports both accurate and inaccurate information. To make sure enough errant information is included, researchers use either confederates or recognition tasks, which

---

Daniel B. Wright, Sorcha A. Mathews, Elin M. Skagerberg, Psychology Department, University of Sussex, Brighton, United Kingdom.

This research was partially supported by Grant RES-000-23-0016 from the Economic and Social Research Council of the United Kingdom. A version of this article was presented at the Bethschrift conference in honor of Elizabeth Loftus, January, 2005, Wellington, New Zealand. The article has benefited from the many useful comments from the audience at that conference.

Correspondence concerning this article should be addressed to Daniel B. Wright, Psychology Department, University of Sussex, Brighton BN1 9QH, United Kingdom. E-mail: danw@sussex.ac.uk

have higher error rates than free-recall tasks (Koriat, Goldsmith, & Pansky, 2000).

A real-world example of how one person's report can affect another person's report occurred with the Oklahoma City bombing (Memon & Wright, 1999; Schacter, 2001). When Timothy McVeigh arrived at Elliot's Body Shop to rent the truck used in the explosion, CCTV showed that he arrived alone. When questioned after the explosion, the owner of the shop, who showed McVeigh the truck, and the secretary, who dealt with the paperwork, did not remember anyone with McVeigh. However, a mechanic, who saw part of the interaction between McVeigh and the secretary, recalled an accomplice with McVeigh. After speaking with his coworkers, his report led the owner and the secretary to report that McVeigh was not alone and, on the basis of their statements, the FBI began a massive hunt for John Doe 2. The FBI now believes that the mechanic was recalling a different customer and that his false memory of an accomplice spread to the other people's memory.

Wright et al. (2000, Experiment 2) showed that it was possible in a laboratory situation both to add an accomplice to somebody's report of an event and to make it less likely that somebody reports an accomplice. Participants were shown a sequence of photographs of somebody going to a pool hall. Half saw a single culprit take a person's wallet; half saw the culprit with an accomplice. Participants were initially questioned individually, and 39 of the 40 participants correctly said whether there was an accomplice. Participants were then allocated into pairs in which one person saw the accomplice and the other did not. They were told to discuss the event and the principal characters. For 19 of the 20 pairs, the people within the pair reported different memories about the accomplice. Of these, 15 pairs came to agree about the presence or absence of an accomplice. Eight agreed that there was no accomplice, and 7 agreed there was an accomplice. With only 19 disputed memories, it was not possible to show whether it is more likely to add a nonexistent accomplice or to make less accessible a previously seen accomplice. The power was too low for this comparison. One aim of the present studies is to see which of these effects is larger.

Talking with other people about an event is a particular type of postevent information. This research, pioneered by Loftus in the 1970s (e.g., Loftus, Miller, & Burns, 1978), shows that it is possible to distort people's memories. Wright, Loftus, and Hall (2001) showed that it is possible both to add scenes into people's memory reports and to make scenes less likely to be reported, but their study lacked the power to show which effect was larger. Our aims are to determine (a) whether what one person says can affect both items that the other person has seen and items that he or she has not seen and (b) whether these effects are of a similar size.

There has been much research and theoretical speculation about how people recognize items as new. The current view of recognition memory is that it is based on two processes. One is familiarity (sometimes referred to as *memory strength* or *gist*), and, on its own, this can account for the results of many studies (e.g., Donaldson, 1996; Dunn, 2004). The process is rapid and automatic. The results of other studies are more difficult to account for with a single-process model. Because of this, most researchers now assume that there is at least one additional process underlying recognition memory. The second process is usually described as a *recollective* or a *verbatim* process. It takes more time than the familiarity process and there is a higher degree of conscious control (Tulving, 1985; Yonelinas, 2001). Several different models for recognition memory assume two processes exist. For example, Rotello, Macmillan, and Reeder's (2004) model assumes presenting an item shifts the memory representation of that item in two dimensions. They then apply a two-dimensional model to account for the data from a large set of studies. Limiting the models to two processes is good for parsimony, but it is of course possible that more processes will in time be necessary to account for all the memory recognition data.

In social recognition memory research, it is unclear which, if either, of the two processes of traditional recognition memory models is influenced when hearing somebody else report about a shared experience. In traditional postevent information research it is usually assumed that the misinformation is more likely to affect the familiarity–gist process (Brainerd & Reyna, 2002). Therefore, the errant responses to misinformation can be rapid (Loftus, Donders, Hoffman, & Schooler, 1989) but can lack the perceptual details that accompany "remember" responses (Wright & Stroud, 1998). This happens because the information is not stored as an episodic event, but as information related to another event and often the participant is unaware that this information is being learned. As Loftus (1993, p. 530) elegantly describes: "the new information invades us, like a Trojan horse, precisely because we do not detect its influence."

Most memory recognition models focus on people recognizing items as being previously shown. Less research has examined how people come to reject an item as being previously shown. For example, Ghetti (2003) has argued that people can use metacognitive strategies to help with rejections. If during the test people are presented with an item, even if it seems familiar, they may reject the item if they think they would have vividly remembered if it had been presented. For example, if at test in a face recognition study the participant is shown a photograph of someone with bright green hair, and he or she does not have a vivid recollection of this person from earlier in the study, the participant should reject this photograph on the basis of a recollection failure. This hypothesis predicts that it should be more difficult to create memories for atypical items. Ghetti (2003) and Rotello (1999) have examined recollection failure by item typicality, but their data are inconsistent. In our first two studies, we vary the frequency and typicality of the items to explore this hypothesis, but given the inconclusive findings in the literature we make no predictions.

One of the most developed models for rejecting items is Brainerd and Reyna's (2002) fuzzy trace theory. In a series of studies they show how a recollection of one item can lead to rejection of another, what they call recollection rejection (Brainerd, Reyna, Wright & Mojardin, 2003). They give the example of being presented with the word *Houston* during the initial stages of a memory study. At test, if the participants are presented with the word *Phoenix*, they argue that some people may reject it because they realize any familiarity is due to the presentation of the semantically similar word *Houston*. The estimated proportion of times that recollection rejection occurs varies with several factors (e.g., age, similarity of distracters), but across nine studies they estimate the proportion to be between 0% to 50% (Brainerd et al., 2003, Tables 2 and 3).

Our predictions can be described in a similar manner. In the nomenclature of fuzzy trace theory, *gist representations* refer to familiar items and *verbatim representations* refer to the perceptual items that are recollected. In fuzzy trace theory, it is assumed that old items can have gist and verbatim representations, and new items may have gist representations but will usually not have verbatim representations. If one person reports that an item can affect the gist but not the verbatim representation, then we would expect that what that person said would have a stronger effect on new items than on old items.

In studies examining how one person's memory responses are affected by other people's memory responses, participants are first presented with information and then they are allowed to interact with each other. There are several design choices. The first choice concerns whether to use a confederate within each participant group. The use of confederates allows control over the interactions, but it is difficult to ensure that the confederates' behavior will be similar to that of real participants. We use confederates in Experiments 1 and 2 and participants only in Experiment 3.

The second choice concerns whether to allow people to discuss the information and then report individually or to include the interaction as part of testing. Given that we wanted to test several items, we combined testing the participant with the presentation of information from the other person (see Schneider & Watkins, 1996). Reysen (2005) has shown that this procedure does produce effects on both how people respond during the group test and, if tested individually, how they respond afterward.

The final choice concerns whether to show a complex event sequence, like a story, or a set of unrelated stimuli. Because the individual elements of any story can relate in complex ways (Schank & Abelson, 1977), only a small amount of independent data can be gathered for any participant, thus lessening the statistical power. The drawback of using multiple trials of unrelated stimuli is that the condition is dissimilar to most actual forensic situations in which people must recall a complex event sequence. The multiple trials approach is more like that which would occur for students studying for a multiple choice exam. We chose to show sets of individual stimuli because of the need for power to evaluate our hypotheses, but we are aware of the need for caution in generalizing with regard to memory for complex event sequences.

We used three different types of stimuli to test whether the results are restricted to only a certain type of stimuli. In the first study, we presented people with words, which are the most common stimulus used to explore false memories because of the popularity of the Deese–Roediger–McDermott (DRM) procedure (Roediger & McDermott, 1995). They are also important within an educational context. We used high- and low-frequency words and nonwords. We expected to find the *word frequency effect* (i.e., better memory for low-frequency words, Arndt & Reder, 2002), but, on the basis of the different findings of Ghetti (2003) and Rotello (1999), we had no a priori view of how this might relate to social recognition memory. Three levels of frequency allowed an internal replication of any observed findings. In the second study, we presented people with photographs of cars because (a) they are common everyday objects and (b) memory for cars is of forensic importance (Wright & Davies, 1999). On the basis of data from a pilot study, cars were classified as high, medium, and low typicality. This allowed further replication of any findings. In the third study, we presented photographs of faces because they are (a) one of the most used stimuli in memory studies and (b) the most important in a forensic setting.

Our hypotheses are as follows:

> *Hypothesis 1:* For new items, we predict an effect of what one person says on the memory responses of the other person. The mean difference for the d' values for accurate and inaccurate information from Schneider and Watkins (1996) and Wright et al. (2000, Experiment 1, Table 1) provides an approximate size for our prediction ($d' = .89$).
>
> *Hypothesis 2:* For old items, we predict an effect of what one person says on the memory responses of the other person. The predicted size is also $d' = .89$.
>
> *Hypothesis 3:* On the basis of our effect size predictions for hypotheses 1 and 2 and the findings of research in related fields, we expect the effects to be larger for new items; however, insufficient data exist to make a more precise prediction.

Despite that these hypotheses all have a predicted direction, we used the more conservative two-tailed hypothesis tests.

We conducted power analyses to determine sufficient sample sizes in order to detect the effects associated with the above hypotheses. We used the traditional alpha of .05. According to Baguley (2004), power analysis should be based on the minimum effects of interest, rather than solely on past effect sizes. To estimate the minimum effects of interest we used Cohen's (1977) U_3 statistic as a measure of effect size, as it is particularly well suited for studies that use methods from signal detection theory, and then we translated it into the familiar d' statistic. For the first two studies, we set the minimum U_3 to be 80% for the three hypotheses. According to the terms from signal detection theory, this means 80% of the target distribution is above the midpoint for the noise distribution. A U_3 of 80% is equivalent to a difference in means of the two distributions of $d' = .84$. Because we tested these hypotheses in multiple studies, we aimed for a power of approximately 90% for the first two studies to lessen the chance of a Type II error. The normal.sample.size function from S-Plus was used for all calculations. For the first two hypotheses, we compare the d' value to 0; this is a one sample case. In this case, d' is equal to Cohen's d because the standard deviation is 1. To achieve a power of 90%, only 15 participants are needed. However, we were aware that the third hypothesis would have less power because we were comparing two measurements. Assuming a medium size correlation ($r = .30$) between these, the d' needs to be divided by 1.18 [$\sqrt{2(1 - r)}$; Cohen, 1977, Equation 2.3.7] to get d. To achieve a minimum power of 90%, 21 participants are needed. We used 24 participants. This produced a power of 98% for the first two hypotheses and a power of 94% for the third hypothesis.

In the third study, we focused on a more specific hypothesis. The critical comparison is comparing a within-participants control condition with the conditions in which the other person has responded. The minimum effect size that we wanted to detect was smaller than those in the first two studies. We set the minimum U_3 to be 70%, which is $d = .44$ (assuming $r = .30$ between measurements). For a power of 80% a sample size of 40 is needed. We also included a between-participants control group that we compared with the within-participants control group. To allow an effect size of U_3 = .80 to be detected, we used 14 participants in the individual condition, which produced a power of 77%.

Experiments 1 and 2

Method

The first two studies had very similar designs and are reported together. The difference between the studies is that participants in the first study had to remember words and those in the second study had to remember photographs of cars.

Participants. In the first study with words, 24 participants (age: M = 21.75 years, SD = 2.75; 13 women, 11 men) volunteered. In the second study, 24 different participants (age: 21.25 years, SD = 4.50; 11 women, 13 men) volunteered. All participants were told that they could withdraw from the study at any time. They were recruited by posters and by word-of-mouth and were entered into a prize drawing. All participants were University of Sussex students and none were used in any more than one study, including the pilot research.

We used 10 confederates to reduce the possibility that any particular confederate would exhibit peculiarities that could greatly affect the results. The confederates were (a) recruited on a voluntary basis, (b) given training on the response coding scheme, and (c) instructed to act as naturally as possible. All confederates were University of Sussex students, and none knew any of the participants they worked with.

Materials. The first study used words (and nonwords) as stimuli; the second used pictures of cars. For the first study, we used 60 two-syllable words. This included 20 high-frequency words (HF; e.g., engine, object, flower), 20 low-frequency words (LF; e.g., stylus, scuba, secrete), and 20 nonwords (NW; e.g., prindel, bannow, glistow). The HF words were from Kilgarriff's corpus (1997) of those words most commonly used in everyday writings. The LF words were from the British National Corpus (BNC). The nonwords were from a study of children's nonword repetition (Gathercole, Willis, Baddeley, & Emslie, 1994). We chose these sources in part because they were all based on U.K. norms. The words are matched on length but were otherwise different.

For the second study, 100 color pictures of a variety of cars were from Adams and Wright (2001). A short pilot study was conducted with these stimuli. Ten participants (university undergraduates) volunteered to take part. They were shown the 100 cars and asked to rate the uniqueness and memorability of each car on a 7-point scale ranging from 1 (*very average*) to 7 (*very unique*). This activity took approximately 10 min to complete.

Means were calculated for each of the 100 cars and 70 were chosen to be used in the study. Of those 70 cars, 20 had a mean above 4.5 and were classified as HF, 20 had a mean below 2.5 and were classified as LF, and 30 had means between 2.9 and 4.1 and were classified as average frequency (AF). The remaining 30 were between these categories (between 2.5 and 2.9, or between 4.1 and 4.5) and were not used in the main study.

Design and procedure. Each study has a 3 × 2 × 2 within-participants design. The variables were the frequency of the stimuli (three levels: HF, LF, and NW for words; HF, LF, and AF for cars), whether the stimulus was presented in the study phase (two levels: presented, not presented), and what the confederate said (two levels: said presented, said not presented). The dependent variable was how often the participant reported having previously seen the item.

Participants and confederates arrived together. After reading through the procedure and consent form, they sat side by side in front of a computer screen. The experimenter explained that they were about to see some words or cars on the screen. The words or cars were presented in a random order for 3 s each. For the first study, they saw 28 words, with approximately equal number of HF words, LF words, and nonwords. They were supposed to have seen 30 words, but there was an error in the computer program. However, this does not affect any of the interpretations. For the second study, 40 cars were shown: 10 HF, 10 LF, and 20 AF. The experimenter told them not to talk with each other; she then started the program and left the room.

The experimenter returned and told them that for the next part of the study the pair would need to work separately on a 5-min word search task. The experimenter told the participant to work on this filler task in the room and escorted the confederate out, ostensibly so he or she could also complete the word search task individually. No reason was given to the participant for the task. This period was necessary to remind confederates how they were supposed to respond to each stimulus. In hindsight, the use of a word search task when words were the stimuli for one of the studies was unfortunate, but none of the words in the study were in the filler task. After 5 min, the experimenter and the confederate returned to the room.

For the final part of the studies, the confederate and participant sat in front of the computer screen. They were told that they would be presented with 60 words or 70 cars, and they were also told that they had previously seen some but not all of these. They were to look at the stimulus and to answer whether they had seen it before. The participant–confederate pairs were told that they would take turns responding, first one person and then the other. They were presented with two envelopes, which were shuffled in front of them, and told that if their envelope contained a *1* they would respond first and if it contained a *2* they would respond second. Both envelopes contained a *2*, although the confederate always said that his or her envelope contained a *1*. Thus, the confederate was told to reply "yes" or "no" to each stimulus and then the participant would answer "yes" or "no." The experimenter would then move on to the next word or car. In all cases, the confederates responded as instructed. The confederates gave erroneous information for approximately one third of the stimuli presented. This proportion was chosen so that there were enough data where confederates were providing errant information without making it appear that they were performing near chance levels. For the word study, this was 20 of 60 responses; for the car study, this was 22 of 70 responses. The stimuli were presented in random order.

At the end of the testing, participants were asked about the study and none thought that the confederate was not an actual participant. Next, they were debriefed and thanked for their participation. In total, each session took approximately 15–20 min. All studies received ethics clearance from the Cognitive and Computing Sciences Ethics Committee of the University of Sussex.

Results

There are two main approaches that can be used to analyze these data: logistic regression and signal detection theory. At one level, both of these approaches are generalized linear models (DeCarlo, 1998) and they usually produce similar results. The logistic regression procedure is more flexible than models from standard signal detection theory but is less well-known among memory recognition researchers. The particular type of model needed is called a *multilevel logistic regression* (Wright, 1998). In order to facilitate communication, we report results from signal detection theory here. However, both methods lead to the same conclusions.

Signal detection theory is usually used with memory recognition research by showing how well a participant is able to discriminate between previously seen stimuli and previously unseen stimuli. In other words, these methods use whether the item is old or new to predict the participant's response. Here, the hypotheses are whether the confederate's responses affect the participant's responses for old and new items. According to the terms of signal detection theory, *hits* refer to instances in which participants respond "old" when the confederate responds "old," and *misses* refer to instances in which participants respond "new" when the confederate responds "old," *correct rejections* refer to instances in which participants respond "old" when the confederate responds "new," and *false alarms* refer to instances in which participants respond "old" when the confederate responds "new." The difference between the probit (i.e., the normal deviate) of the hit rate and the probit of the false alarms rate, d', measures the effect of the confederate response. This allows us to calculate whether the effects for the confederate response, the d' values, are different for new and old items. We also calculate C, which measures how often the participant responds "old." We calculate these separately for new and old items, and the difference between the C scores for new and old items is a measure of accuracy. In calculating the hit and false alarm rates, we added a flattening constant, .5, to each cell so that their normal deviates would not equal either negative or positive infinity for any cases. This is standard practice with signal detection theory.

For the first two studies, we used three different levels for the stimuli. Thus, d' and C were calculated for each of these. For each participant, six d' scores were calculated (for old and new items, for the three frequency levels). Therefore, 2 × 3 within-participants analyses of variance (ANOVAs) are appropriate. When reporting the effect sizes, we use partial omega (ω_p). This is a measure of the amount of variance accounted for by the effect adjusted to estimate the population value (see Vacha-Haase & Thompson, 2004). Following Rosenthal, Rosnow, and Rubin (2000), we used ω_p rather than ω_p^2 because, as they argue, the unsquared version better conveys the size of the effect. When $F < 1$, we report $\omega_p = 0$. The Geisser–Greenhouse correction is used if Mauchly's test of sphericity is statistically significant.

Table 1 shows the percentage of trials that participants reported having previously seen a word partitioned by whether they had previously seen the word, the confederate response, and the word frequency. The d' statistic and the 95% between-participants confidence interval (CI) for each condition are also shown. The d' values are all above 0 showing that participants were influenced by the confederate response for both new and old items at each frequency level. In each case, the d' statistics were larger for new words than for old words. A 2 × 3 within-participants ANOVA was conducted, and the relevant statistics are reported at the top of Table 2. There were statistically significant main effects for word frequency and for new and old items. The interaction was not statistically significant.

We also compared the C statistics across the different conditions (see bottom of Table 1). A 2 × 3 within-participants ANOVA yielded a statistically significant effect for whether the item was new or old, with the old items having higher C values. This shows that the participants were able to discriminate old from new items. There was no significant main effect for word frequency and no significant interaction.

The second study examined people's recognition memory for cars. The data were analyzed in the same way as in the first study. Table 3 presents the percentages for how often participants reported that they had seen the item previously, with the mean d' values, C values, and 95% CIs for each condition. The d' values are all above 0 showing that participants were affected by the confederate response in each condition, for new and old items, and for each level of frequency. The results from the 2 × 3 within-participants ANOVAs for d' and C values are reported in Table 2. For d' values, there was a statistically significant main effect for whether the item was new or old, with d' being higher for new items. Neither the main effect of car frequency nor the interaction was statistically significant for d'. For C, there were statistically significant main effects for whether the item was new or old and for frequency, and their interaction was statistically significant. The main effect for item was due to C being higher for old items than for new items. The main effect for item shows that participants were able to discriminate old from new items. The effect for frequency shows that there were more "old" responses for high-frequency cars than for the other categories. The interaction occurred because the difference between new and old items was greatest for the low-frequency cars. This means that participants were most accurate in discriminating old from new cars for the low-frequency cars.

Discussion

One person's report about a past event can affect other people's reports. In two studies, each with 12 conditions, the mean d' was positive and significantly different from 0. This was true for new and old items. However, in all six comparisons the d' scores were higher for the new stimuli than for the old stimuli. In both studies, this was statistically significant and the effect sizes were substantial ($\omega_p = .23$ and $\omega_p = .31$, respectively). To return to the hypotheses, we found that participants were influenced on both new and old items, but the effect was larger for new items. We also reported values for C, allowing us to measure accuracy. In each of the six comparisons, the

Table 1
Percentages of Participant "Old" Responses and Test Statistics for New and Old Items From Experiment 1

Confederate response and statistic	HF words		LF words		Nonwords	
	New	Old	New	Old	New	Old
New	14.29	59.72	10.67	74.20	10.12	79.17
Old	41.86	72.73	25.00	86.11	25.00	81.55
d'	1.80	1.43	1.64	1.51	1.21	0.82
CI	1.54, 2.05	1.14, 1.71	1.29, 2.00	1.14, 1.88	0.86, 1.57	0.54, 1.10
C	−0.24	0.11	−0.33	0.18	−0.39	0.21
CI	−0.42, −0.07	−0.05, 0.27	−0.49, −0.16	0.02, 0.33	−0.63, −0.14	0.01, 0.41

Note. The d' values provide an index of the magnitude of the confederate's influence on the memory report; a d' of 0 indicates that there was no influence. The C values measure how often participants responded "old." The difference between these scores for new and old items measures how well participants discriminated between new and old items. HF = high frequency; LF = low frequency; CI = 95% confidence interval.

Table 2
Test Statistics and Effect Sizes for Experiments 1 and 2

Effect	*df*	*MSE*	*F*	ω_p
Experiment 1				
d', word frequency	2, 46	0.45	11.94**	.36
d', old or new	1, 23	0.36	8.85**	.23
d', interaction	2, 46	0.56	0.45	.00
C, word frequency	1.42, 32.69[a]	0.18	0.05	.00
C, old or new	1, 23	0.19	44.53**	.48
C, interaction	1.35, 30.99[a]	0.17	1.61	.09
Experiment 2				
d', typicality	2, 46	0.53	1.03	.02
d', old or new	1, 23	0.35	16.74**	.31
d', interaction	2, 46	0.35	0.97	.00
C, typicality	2, 46	0.09	4.61*	.22
C, old or new	1, 23	0.55	44.37**	.48
C, interaction	2, 46	0.18	5.05*	.23

Note. For Experiment 1 and tests with *C,* the test statistics for Mauchly's test of sphericity were $\chi^2(2) = 11.50$, $p < .01$, for the main effect of word frequency, and $\chi^2(2) = 14.57$, $p < .01$, for interaction of word frequency and whether the item was new or old. Therefore, the Geisser–Greenhouse correction is used for these tests. For other applicable comparisons in all of the experiments, the test of sphericity was nonsignificant.
[a] Mauchly's test of sphericity was significant in these comparisons.
* $p < .05$. ** $p < .01$.

C value was higher for old items than for new items, which means that the participant responses were accurate.

The word and car frequency were varied in these studies to explore whether any systematic patterns would arise for the effects of what the confederate said. There was only one statistically significant effect: The effect is larger for nonwords than for low- and high-frequency words. Although this effect is interesting, we focus on the overall pattern of results, which showed that all the conditions yielded significant effects and that for all six different levels of frequency the effect is larger for new items than for old items. Therefore, we did not differentiate the stimuli by frequency in our third study.

In the third study, we made three changes to explore the generality of these findings. The first change for generality was to use only pairs of participants rather than pairs consisting of a confederate and a participant. Although none of the participants in Experiments 1 and 2 said that they thought the confederate was not a participant, confederates still could have been behaving differently than the participants. The second change was to allow participants to use a 4-point rating scale rather than simply saying "old" or "new." This means that the response variable is more sensitive and also that the perceived credibility of each participant's response would vary along this scale (Schneider & Watkins, 1996). The third change was that we used photographs of faces as stimuli.

The main conclusion from the first two studies is that the effect of what other people say is larger for unseen than for seen items. There are two different situations for unseen and seen items. For unseen items, the other person could say that the item has been seen before (*errant information*) or that it has not been seen before (*accurate information*). Similarly for seen items, the other person could provide accurate or inaccurate information. In the final study, the effects in each of these four situations were estimated by comparing them with

Table 3
Percentage of Time That Participants Respond That They Had Previously Seen the Car and Therefore Respond "Old" in the Second Study

Confederate response and statistic	HF cars		MF cars		LF cars	
	New	Old	New	Old	New	Old
New	10.00	53.70	19.89	50.00	14.95	52.08
Old	40.28	77.08	66.67	71.11	47.92	74.40
d'	1.21	0.61	0.84	0.56	0.92	0.60
CI	0.83, 1.58	0.33, 0.90	0.59, 1.10	0.26, 0.84	0.61, 1.23	0.25, 0.95
C	−0.24	0.30	−0.46	0.32	−0.69	0.39
CI	−0.40, −0.07	0.09, 0.50	−0.64, −0.27	0.08, 0.56	−0.87, −0.51	0.15, 0.63

Note. The d' values provide an index of the magnitude of the confederate's influence on the memory report; a d' of 0 indicates that there was no influence. The *C* values measure how often participants responded "old." The difference between these scores for new and old items measures how well participants discriminated between new and old items. HF = high frequency; MF = medium frequency; LF = low frequency; CI = confidence interval.

one of two control conditions: The between-participants control condition consisted of participants who took part in the study individually, and the within-participants condition measured the responses for the person in the pair when responding first.

Experiment 3

Method

Participants. Fifty-four participants were recruited from posters placed around the University of Sussex campus and through a participant e-mail list. Participants were undergraduate and postgraduate students. No further personal information was recorded. Fourteen participants were allocated into an individual control condition. Forty participants were allocated into a pair condition. All were paid £5 (~$US9) for participating in this study.

Materials. Photographs of White male faces were taken from the Psychological Image Collection at Stirling (PICS; http://pics.psych.stir.ac.uk/, retrieved November 8, 2004) and from previous research (Wright, Boyd, & Tredoux, 2003) conducted at other universities. Only White male faces were used to make the stimulus set relatively homogeneous.

Design and procedure. Participants arrived at the laboratory. Those in the individual condition were taken to a laboratory room with a computer and told to sit in front of the computer. They were shown 50 faces, in random order, on a computer screen for 2 s per face. After being shown the 50 faces, they were given a response sheet with 100 numbered lines with a single rating scale that ranged from 1 (*not seen before*) to 4 (*seen before*). They were told that they would be shown 100 faces, 1 at a time, on the computer screen. They were told that half of these were previously shown. They were told that if they were confident that the face had not been shown previously, they were to circle *1*; if they were confident that it had been shown, they were told to circle *4*; and they were told to circle *2* or *3* if they were less confident. After they had responded for a face, they pressed the return key and the next face appeared. The instructions for this phase took about 2–3 min. There were no time constraints during this part of the study. After completion of this task, participants were debriefed and paid.

Participants in the pair condition were taken in pairs to a laboratory room with a computer and two chairs equally distant from the computer. They saw the presentation of the original 50 faces and were then given the instructions for the second phase of the study. The pair was given a single response sheet with two columns of 100 numbered lines. They were told they would be shown 100 faces, half of which they had previously seen. One of the participants was randomly allocated to respond first for the first 50 faces, and then the order was reversed. For each face, the first participant responded in the first column using the 4-point scale. The sheet was handed to the second person who responded in the second column, and then one of the pair pressed the return key to start the next test trial. They were instructed not to talk with each other during the study. After the 50th face was shown, participants were reminded to switch—the person who originally responded first would now respond second, and vice versa. Afterward they were debriefed and paid.

Results and Discussion

For each participant, there were 50 trials for previously seen items and 50 trials for previously unseen items. For those in the individual condition (between-participants control), there was no other information presented. For those in the pair condition, each participant was the first responder for half of the trials (25 seen and 25 unseen items; within-participants control). For half of the trials in this condition, participants saw the other participant's rating on the 1–4 scale. Sometimes the other participant responded accurately, sometimes he or she did not. These data can be analyzed in several different ways. After comparing the between-participants and within-participants control groups, we used two techniques to investigate the hypotheses. First, we applied the same techniques (from signal detection theory) that were used in the first two studies. Next we used the rating scale responses and within-participants ANOVAs.

The first question is whether the two control conditions produced similar response patterns. Means were calculated for the individual control condition participants for the 50 previously unseen items and for the 50 previously seen items, and for the participants in the pair condition when they were the first responder for the 25 previously unseen and 25 previously seen items. The mean response on the 1–4 scale for previously unseen items was 1.82 both for the individual condition (SE = .08) and for the pair condition when responding first (SE = .07). For seen items, the means were also very similar. For those in the individual condition, the mean was 2.96 (SE = .08) compared with 2.82 (SE = .07) in the pair condition, $t(52) = 1.11$, *ns*. We used the responses for the pair condition in the remaining analyses because they are based on the responses of a greater number of participants ($n = 40$ vs. $n = 14$), and because this allows the more powerful within-participants analyses to be used.

To allow comparison with the first two studies, we begin by reporting data based on recoding the responses of both participants in each pair into either recognizing or not recognizing the face (*1* and *2* on the scale were coded as "new" responses, *3* and *4* were coded as "old" responses). Although in general it is not advisable to dichotomize variables (MacCallum, Zhang, Preacher, & Rucker, 2002; Wright, 2003), here it allows direct comparison with the earlier studies. Table 4 shows the percentages of "new" and "old" responses for the different conditions. We calculated d' and C in the same manner as in the first two studies. The mean d' for new items was .61 (95% CI, .38, .84) and for old items the mean d' was .37 (95% CI, .18, .56). The fact that both of these are above 0 shows that participants' responses were influenced by the other

Table 4
Percentages of Frequencies and Responses for the Different Conditions in Experiment 3

	Between-participants control		Within-participants control		Other participant says "new"		Other participant says "old"	
Frequency and response	New	Old	New	Old	New	Old	New	Old
Total trials "New"	543	232	773	378	643	182	152	190
%	77.57	33.14	77.30	37.80	83.18	48.15	66.96	30.55
Total trials "Old"	157	468	227	622	130	196	75	432
%	22.43	66.86	22.70	62.20	16.82	51.85	33.04	69.45

Note. Rating scale data have been collapsed into "new" and "old" responses to allow comparisons with data from Experiments 1 and 2.

participants' responses, for both new and old items. The difference between these d' values was statistically significant, $t(39) = 2.08$, $p < .05$, $\omega_p = .20$; the social recognition effect was greater for new items than for old items. The corresponding effect sizes in the first two studies were $\omega_p = .23$ and $\omega_p = .31$, respectively. The C statistics also differed between the sets of items, $t(39) = 10.50$, $p < .01$, $\omega_p = .76$, with the higher mean for the old items ($M = .31$, 95% CI = .20, .43) than for the new items ($M = -.67$, 95% CI = $-.80$, $-.53$). These results indicate that participants were responding accurately. The corresponding effect sizes in the first two studies were both $\omega_p = .48$.

The second method of analysis used data from the whole 1–4 scale. Figure 1 shows the mean values ("old" responses being higher) for participants whose responses were dependent on the other participant's response. One approach to analyzing data in this form is to calculate and compare means for each individual for the 10 conditions shown in Figure 1. The problem with this approach was that there were an unequal number of trials in these conditions because the number of trials was based on participant response. In fact, some conditions have no trials because one or the other of the participants did not use the entire confidence scale. We used the expectation-maximization algorithm (Dempster, Laird & Rubin, 1977) to estimate these missing values, and we then conducted a within-participants ANOVA (without weighting the means). A second approach to analyzing these data is to use multilevel modeling in which the individual trial is nested within the participant (Wright, 1998; Goldstein, 2003). Both approaches produce the same basic results. The former is simpler to communicate so we report those results here.

The ANOVAs compared ratings for each participant when he or she responded first (the within-participants control condition) with their ratings after the other participant responded. This was done separately for items to which the other participant responded "new" and for items to which the other participant responded "old." Thus, each ANOVA had a 2 (old or new) × 3 (control, low-confidence, high-confidence) design. The test statistics are shown in Table 5. When the other participant responded "new" there were statistically significant main effects for item and the other participant's response, but there was no interaction. The effect for item shows that participants gave higher ratings for old items, as expected (i.e., participants responded accurately).

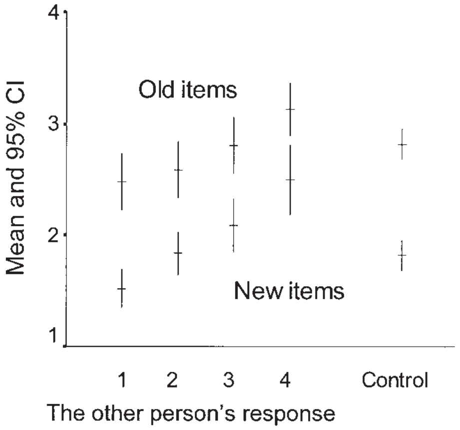

Figure 1. The means and 95% confidence intervals (CI) for participants whose responses were dependent on the other participant's response. The rating scale ranges from 1 (*confident that the item is new*) to 4 (*confident that the item is old*). The means for control conditions are for participants responding first. The confidence intervals are between-participants intervals and were calculated with MLwiN 2.00 (Rasbash, Browne, Healy, Cameroon, & Charlton, 2004).

Comparing the control condition with the other participant's "old" responses produced a different pattern of results. As with the previous analysis, there were statistically significant main effects for item and the other participant's response. The first effect shows that participants were accurate, and the second effect shows that when the other participant responded "old" confidently the ratings increased. However, these effects must be viewed in light of a statistically significant interaction. Visual inspection of Figure 1 shows that this is due to the effect of the other participant's response being stronger for new items than for old items.

In summary, the first two studies found that the influence of the other participant's response was largest for new items. This study replicates this finding by using pairs of participants, rating scales, and a different type of stimuli, and it isolates the effect. The effect is largest when the other participant responds "old" to a previously unseen item. The largest social recognition memory effects were found when false reports were created.

General Discussion

Three hypotheses were examined. The first two concerned whether one participant's response affects the other participant's response for previously unseen items (new) and previously seen items (old). We found strong effects for both new and old items. This was true for three different types of stimuli and for different levels of frequency. This confirms that one participant's response affects how the other participant responds. The third and main hypothesis concerned whether the effects for new and old items were of a similar size. For all three studies, we found significantly larger effects for new items than for old items. The finding was consistent across the studies (all effect sizes over $\omega_p = .20$). Further, the strongest effects were found when it was suggested that a previously unseen item had been seen.

This suggests that it is more likely that a person can be more readily influenced to report a false memory for a nonexistent event than to not report a memory for an actual event. Returning to the Oklahoma City bombing example, it would be more likely that the mechanic, who reported an accomplice with McVeigh, would be able to convince his coworkers that there was an accomplice than it would be for his coworkers to convince him that there was no accomplice. There are other studies supporting the idea that another person's response is more likely to create a memory than to "remove" one. For example, Ost, Costall, and Bull (2002) questioned 20 people who had recovered memories of sexual abuse and who later came to believe that the memories were false. There is much debate about how to treat the veracity of these retractors' memories, but if we take their beliefs at the time of recovery and retracting as a guide, it appears that they began with no memory and had one added, and then they had a memory which was "removed." On several measures, Ost and colleagues found that there was less effort necessary for the memories to be added than for them to be removed.

Our findings can be applied to several areas, for example eyewitness testimony, education, and reminiscences. In all of these situations, people often try to remember information about past events in groups. Our results suggest that the impact of what one person reports

Table 5
Test Statistics and Effect Sizes From the ANOVAs in Experiment 3

Other participant response and effect	*df*	*MSE*	*F*	ω_p
"New"				
Old or new	1, 39	0.54	104.02**	0.75
Control, low, high confidence	2, 78	0.19	7.08**	0.36
Interaction	2, 78	0.18	0.63	0.00
"Old"				
Old or new	1, 39	0.37	103.02**	0.75
Control, low, high confidence	2, 78	0.31	30.29**	0.65
Interaction	2, 78	0.24	3.88*	0.26

Note. ANOVA = analysis of variance.
* $p < .05$. ** $p < .01$.

is greater for events that have not occurred than for events that have occurred. From an applied perspective, it is important to realize that in most situations when people freely recall information, their memories are accurate (Koriat, Goldsmith, & Pansky, 2000). There are more errors of omission than of commission in free recall, and therefore trusting other people's memories for events tends to increase the probability of an accurate belief.

It is important to be cautious in generalizing our results to memory for extended and complex event sequences. Three types of stimuli were used to help generality, but each study used the same basic protocol in which several stimuli were shown and then after a brief delay participants were tested in a controlled setting with a recognition task. These conditions were necessary to allow for comparisons of the size of the different effects, but remembering words, photographs of cars, and photographs of faces are different situations than remembering extended autobiographical memories, as was investigated in Ost et al. (2002). The techniques used in this article are more akin to what students might encounter when studying for an exam, a situation in which what may seem like large amounts of unrelated facts need to be learned.

We also urge caution in generalizing our results to extremely unusual stimuli. Although we did vary typicality in our first two studies, we did not use highly unusual exemplars, such as including a photograph of a military tank in the study with cars or a profane word in the word study. It would be valuable to test whether the same pattern of results would occur for those instances. Future research should expand our findings to include more types of stimuli and extended event sequences.

Finally, it is important to relate the findings to the standard models of recognition memory. The finding that people are influenced by other people's responses shows that information from social communication can affect people's responses. Assuming recognition memory is based on two processes, called *familiarity* and *recollection*, it is worth asking which of these may be influenced by information from others. A useful distinction in some postevent information research is whether the information is blatant or subtle (Loftus, 1979). As described earlier, subtle information, like the Trojan horse, can enter into a person's beliefs. It can make the postevent information seem familiar but provides no contextual information about the source of the information. The characteristics of reported misinformation suggest this is true. Responses can be rapid, confidently held, and be characterized by a "know" response rather than a "remember" response (Loftus et al., 1989; Wright & Stroud, 1998). Over time, the familiarity could create a belief and from that a full-blown recollection (Mazzoni, Loftus & Kirsh, 2001), but in the short term it appears that when postevent information is reported it is because it seems familiar.

In our studies, when the participant responded right after another person, whether he or she was a confederate or another participant, the contextual information was present. Given the old–new recognition format of our tests, it is difficult to argue that familiarity is being affected. Similarly, it is difficult to argue that new recollective traces were established. What seems more likely is that the participant used the other person's response as an additional type of information. This information guided their responses. As such, it is likely that participants used a meta-cognitive strategy when responding. The main finding of our research, that there is a larger effect for new items than for old items, suggests that this meta-cognitive strategy includes the rule that it is better to trust other people's memories of items that are not remembered than to trust other people's lack of a memory of items that are remembered.

References

Adams, E. J., & Wright, D. B. (2001, July 16–20). "Memory for vehicles." Paper presented at the Third International Conference of Memory (ICOM 3), Valencia, Spain.

Arndt, J., & Reder, L. A. (2002). Word frequency and receiver operating characteristic curves in recognition memory: Evidence for a dual-process interpretation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 830–842.

Asch, S. E. (1955). Opinions and social pressure. *Scientific American, 193,* 31–35.

Baguley, T. (2004). Understanding statistical power in the context of applied research. *Applied Ergonomics, 35,* 73–80.

Basden, B. H., Basden, D. R., & Henry, S. (2000). Costs and benefits of collaborative remembering. *Applied Cognitive Psychology, 14,* 497–507.

Basden, B. H., Reysen, M. B., & Basden, D. R. (2002). Transmitting false memories in social groups. *American Journal of Psychology, 115,* 211–231.

Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science, 11,* 164–169.

Brainerd, C. J., Reyna, V. F., Wright, R., & Mojardin, A. H. (2003). Recollection rejection: False-memory editing in children and adults. *Psychological Review, 110,* 762–784.

Clark, S. E., Hori, A., Putnam, A., & Martin, T. P. (2000). Group collaboration in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1578–1588.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (rev. ed.). New York: Academic Press.

Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological Review, 107,* 261–288.

DeCarlo, L. T. (1998). Signal detection theory and generalized linear models. *Psychological Methods, 3,* 186–205.

Dempster, A., Laird, N., & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: B, 39,* 1–38.

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition, 24,* 523–533.

Dunn, J. C. (2004). Remember-know: A matter of confidence. *Psychological Review, 111,* 524–542.

Gabbert, F., Memon, A., & Allan, K. (2003). Memory conformity: Can eyewitnesses influence each other's memories for an event? *Applied Cognitive Psychology, 17,* 533–543.

Gabbert, F., Memon, A., Allan, K., & Wright, D. B. (2004). Say it to my face: Examining the effects of socially encountered misinformation. *Legal and Criminological Psychology, 9,* 215–227.

Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The Children's Test of Phonological Working Memory. *Memory, 2,* 103–127.

Ghetti, S. (2003). Memory for nonoccurrences: The role of metacognition. *Journal of Memory and Language, 48,* 722–739.

Goldstein, H. (2003). *Multilevel statistical models* (3rd ed.). London: Edward Arnold.

Kihlstrom, J. F. (2002). Memory, autobiography, history. *Proteus: A Journal of Ideas,19*(2), 1–6.

Kilgarriff, A. (1997). Putting frequencies in the dictionary. *International Journal of Lexicography, 10,* 135–155.

Koriat, A., Goldsmith, M., & Pansky, A. (2000). Toward a psychology of memory accuracy. *Annual Review of Psychology, 51,* 483–531.

Loftus, E. F. (1979). *Eyewitness testimony.* Cambridge, MA: Harvard University Press.

Loftus, E. F. (1993). The reality of repressed memories. *American Psychologist, 48,* 518–537.

Loftus, E. F., Donders, K., Hoffman, H. G., & Schooler, J. W. (1989). Creating new memories that are quickly accessed and confidently held. *Memory & Cognition, 17,* 607–616.

Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 19–31.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7,* 19–40.

Marsh, E. J., Tversky, B., & Hutson, M. (2005). How eyewitnesses talk about events: Implications for memory. *Applied Cognitive Psychology, 19,* 531–544.

Mazzoni, G. A. L., Loftus, E. F., & Kirsh, I. (2001). Changing beliefs about implausible autobiographical events. *Journal of Experimental Psychology: Applied, 7,* 51–59.

Meade, M. L., & Roediger, H. L., III (2002). Explorations in the social contagion of memory. *Memory & Cognition, 30,* 995–1009.

Memon, A., & Wright, D. B. (1999). The search for John Doe 2: Eyewitness testimony and the Oklahoma bombing. *The Psychologist, 12,* 292–295.

Ost, J., Costall, A., & Bull, R. (2002). A perfect symmetry? A study of retractors' experiences of making and then repudiating claims of early sexual abuse. *Psychology, Crime & Law, 8,* 155–181.

Rasbash, J., Browne, W., Healy, M., Cameroon, B., & Charlton, C. (2004). *MLwiN 2.00.* London: Institute of Education, Multilevel Models Project.

Reysen, M. B. (2005). The effects of conformity on recognition judgments. *Memory, 13,* 87–94.

Roediger, H. L., III, Meade, M. L., & Bergman, E. T. (2001). Social contagion of memory. *Psychonomic Bulletin & Review, 8,* 365–371.

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in list. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 803–814.

Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (2000). *Contrasts and effect sizes in behavioural research: A correlational approach.* Cambridge, U.K.: Cambridge University Press.

Rotello, C. M. (1999). Metacognition and memory for non-occurrence. *Memory, 7,* 43–63.

Rotello, C. M., Macmillan, N. A., & Reeder, J. A. (2004). Sum–difference theory of remembering and knowing: A two-dimensional signal-detection model. *Psychological Review, 111,* 588–616.

Schacter, D. L. (2001). *The seven sins of memory: How the mind forgets and remembers.* Boston: Houghton Mifflin.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals and understanding: An inquiry into human knowledge structures.* Hillsdale, NJ: Erlbaum.

Schneider, D. M., & Watkins, M. J. (1996). Response conformity in recognition testing. *Psychonomic Bulletin & Review, 3,* 481–485.

Shaw, J. S., III, Garven, S., & Wood, J. M. (1997). Co-witness information can have immediate effects on eyewitness memory reports. *Law and Human Behavior, 21,* 503–523.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26,* 1–12.

Vacha-Haase, T., & Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counseling Psychology, 51,* 473–481.

Weldon, M. S. (2000). Remembering as a social process. In D. L. Medin (Ed.), *Psychology of Learning and Motivation.* San Diego, CA.: Academic Press.

Weldon, M. S., & Bellinger, K. D. (1997). Collective memory: Collaborative and individual processes in remembering. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1160–1175.

Wright, D. B. (1998). Modelling clustered data in autobiographical memory research: The multilevel approach. *Applied Cognitive Psychology, 12,* 339–357.

Wright, D. B. (2003). Making friends with your data: Improving how statistics are conducted and reported. *British Journal of Educational Psychology, 73,* 123–136.

Wright, D. B., Boyd, C. E., & Tredoux, C. G. (2003). Inter-racial contact and the own race bias for face recognition in South Africa and England. *Applied Cognitive Psychology, 17,* 365–373.

Wright, D. B., & Davies, G. M. (1999). Eyewitness testimony. In F. T. Durso, Nickerson, R. S., Schvaneveldt, R. W., Dumais, S. T., Lindsay, D. S., & Chi, M. T. H. (Eds.), *Handbook of applied cognition* (pp. 789–818). New York: Wiley.

Wright, D. B., & Klumpp, A. (2004). Collaborative inhibition is due to the product, not the process, of recalling in groups. *Psychonomic Bulletin & Review, 11,* 1080–1083.

Wright, D. B., Loftus, E. F., & Hall, M. (2001). Now you see it; now you don't: Inhibiting recall and recognition of scenes. *Applied Cognitive Psychology, 15,* 471–482.

Wright, D. B., Self, G., & Justice, C. (2000). Memory conformity: Exploring misinformation effects when presented by another person. *British Journal of Psychology, 91,* 189–202.

Wright, D. B., & Stroud, J. N. (1998). Memory quality and misinformation for peripheral and central objects. *Legal and Criminological Psychology, 3,* 273–286.

Yonelinas, A. P. (2001). Consciousness, control, and confidence: The 3 Cs of recognition memory. *Journal of Experimental Psychology: General, 130,* 361–379.

Received December 6, 2004
Revision received May 27, 2005
Accepted May 27, 2005 ■