MEMORY, 2008, 16 (2), 137–148

Psychology Press
Taylor & Francis Group

Changing the criterion for memory conformity in free recall and recognition

Daniel B. Wright
University of Sussex, Falmer, UK

Fiona Gabbert
University of Abertay, Dundee, UK

Amina Memon
University of Aberdeen, UK

Kamala London
University of Toledo, OH, USA

People's responses during memory studies are affected by what other people say. This memory conformity effect has been shown in both free recall and recognition. Here we examine whether accurate, inaccurate, and suggested answers are affected similarly when the response criterion is varied. In the first study, participants saw four pictures of detailed scenes and then discussed the content of these scenes with another participant who saw the same scenes, but with a couple of details changed. Participants were either told to recall everything they could and not to worry about making mistakes (lenient), or only to recall items if they were sure that they were accurate (strict). The strict instructions reduced the amount of inaccurate information reported that the other person suggested, but also reduced the number of accurate details recalled. In the second study, participants were shown a large set of faces and then their memory recognition was tested with a confederate on these and fillers. Here also, the criterion manipulation shifted both accurate and inaccurate responses, and those suggested by the confederate. The results are largely consistent with a shift in response criterion affecting accurate, inaccurate, and suggested information. In addition we varied the level of secrecy in the participants' responses. The effects of secrecy were complex and depended on the level of response criterion. Implications for interviewing eyewitnesses and line-ups are discussed.

When two people talk about a previously seen event, what one person says can influence what the other person reports. This has been shown in several studies using both free recall and recognition of complex sequences (e.g., Dalton & Daneman, 2006; Gabbert, Memon & Allan, 2003; Gabbert, Memon, Allan & Wright, 2004; Gabbert, Memon & Wright, 2006, 2007; Garry, French, Kinzett, & Mori, in press; Hope, Ost, & Gabbert, in press; Mori, 2003, 2007; Paterson, 2004; Shaw, Garven & Wood, 1997; Skagerberg, 2007; Wright, Self & Justice, 2000) and recognition of simple stimuli (e.g., Allan & Gabbert, in press; Meade & Roediger, 2002; Reysen, 2003, 2005; Roediger,

Address correspondence to: Daniel B. Wright, Psychology Department, University of Sussex, Falmer, BN1 9QH, UK. E-mail: danw@sussex.ac.uk

Funded by the Economic and Social Research Council and the Leverhulme Trust. Thanks to Kristin Price for help with the second study and to Geoff Ward for useful comments on a previous version of the paper.

http://www.psypress.com/memory DOI:10.1080/09658210701836174

Meade & Bergman, 2001; Schneider & Watkins, 1996; Skagerberg & Wright, 2008; Wright, Mathews & Skagerberg, 2005). This research is important for applied purposes including criminal investigations because eyewitnesses often view crimes with other people and often discuss these events (Paterson & Kemp, 2006; Skagerberg & Wright, in press). Further, in retelling events, people often incorporate exaggerations and other distortions (Marsh & Tversky, 2004). Together, these findings sum to a large problem for the justice system. Many eyewitness reports will be based in part on erroneous information taken from other eyewitnesses. Therefore, the instructions given to eyewitnesses are vital if they affect whether this erroneous information is reported. While memory errors within a forensic context may be the most catastrophic, similar errors are likely to exist when students study together, when family members reminiscence, and in group problem solving.

Exploring social influences on memory has a long past within memory research (Bartlett, 1932), but recently there has been increased interest (Basden, Basden, & Henry, 2000; Blank, 2005; Clark, Hori, Putnam, & Martin, 2000; Conway & Pleydell-Pearce, 2000; Finlay, Hitch & Meudell, 2000; Kihlstrom, 2002; Meudell, Hitch & Boyle, 1995; Rajaram & Pereira-Pasarin, 2007; Reysen & Adair, in press; Thompson, 2002; Weldon, 2000; Wright & Klumpp, 2004). The current studies examine how changing the task demands affects both the accurate reporting of memories and how often people incorporate what other people say into their own memory reports. Participants are either told they should report everything that they can remember, and not worry about reporting errant information, or to only report information that they are fairly certain is accurate. The aim of the instruction for the first group is to maximise the amount of correct details reported, and for the second group to minimise the amount of erroneous details reported.

The most relevant study in the literature is Experiment 1 of Meade and Roediger (2002). They showed participants six scenes and then had them recall aspects of these scenes with a confederate who the participants thought had also just viewed the scenes and was trying to recall them. The confederate purposefully recalled some accurate and some erroneous information. The participants then took part in an individual recall test. Half were warned that the other person's recall may have been erroneous and also that they should only report ''items that they were sure had actually appeared in the scenes'' (p. 998). The first part of this instruction, a warning about possible misinformation, has been shown with other procedures to reduce misinformation effects (Echterhoff, Hirst, & Hussy, 2005; Weingardt, Toland, & Loftus, 1994). The second part of the instruction is about a change in the criterion for reporting information. While Meade and Roediger's instruction reduced the memory conformity effect, it is unclear which part of the instruction caused this.

Within another area of false memory research, the Deese-Roediger-McDermott (DRM) procedure (Roediger & McDermott, 1995), Wright, Startup, and Mathews (2005) have shown that a response criterion manipulation can affect both the number of accurate free recalls from a DRM list and the number of inaccurate critical lure recalls. Response criterion manipulation is the focus of the current studies. We examine how this affects accurate and inaccurate memory for items, both when they are suggested by another person and when they are not. In the first study we examine free recall and in the second recognition.

SIGNAL DETECTION THEORY TERMINOLOGY

Our predictions can be best described using terminology from signal detection theory (SDT). SDT is a technique popular in diagnostic medicine, but is also used in some memory recognition research (Banks, 1970). It can be thought of as both a theoretical framework for understanding discrimination problems and a statistical method for analysing some of the data from these situations. We use it as a theoretical framework. Other methods are better suited for our statistical analyses. The following is a brief introduction to SDT as a theoretical framework (see Macmillan & Creelman, 2005; Zho, Obuchowski, & McClish, 2002, for further details).

For most memory tasks there are two types of accuracy: sensitivity and specificity. Sensitivity is being able to accurately say when an object was previously seen; what is called a *hit*. Specificity is being able to accurately say that previously unseen objects were not seen; what is called a *correct rejection*. Incorrectly identifying a previously unseen item is called a *false alarm*.

We manipulate the task instructions so that people are either told to report everything that they can without worrying if they report a few items that were not previously seen, or they are told to report items only if they are sure that they were previously shown. We predict that these will establish lenient and strict criteria for reporting, respectively. The lenient criterion should increase accuracy in terms of sensitivity, but lower specificity. The opposite pattern should emerge for those using the strict criterion. Figure 1 shows a simplified schematic of memories for previously unseen (new) items and previously seen (old) items. The x-axis here corresponds to some unidimensional construct like memory strength (it may be based on several constructs and be viewed as a latent variable for the likelihood of responding ''old''). It is assumed that the person has some type of conscious access to this memory strength, although it may be as discrete steps in strength rather than as a continuum. Figure 1 shows that the typical new item will have a memory strength value around 3, but that some values will be closer to 1 and some closer to 5. Old items will tend to have values of about 5 but some values will be around 3 and some above 7. Because these two distributions overlap, if basing a response decision on memory strength, then there will be some previously unseen new items and some previously seen old items that will have memory strength values around 4 and will be difficult to distinguish (if just using memory strength). Because these distributions overlap there are bound to be errors when responding on the bases of memory strength.

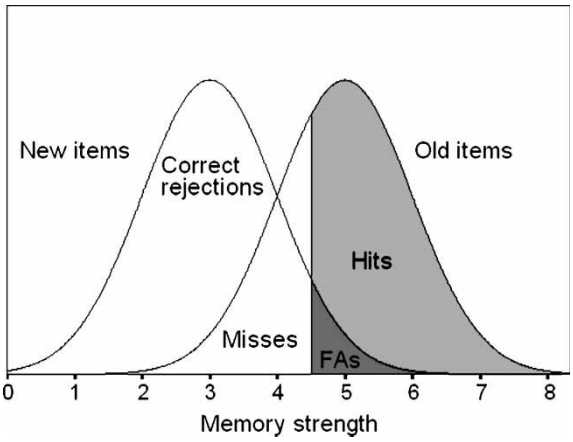

Figure 1. Distributions for new and old items with the classification of correct rejections, misses, false alarms (FAs), and hits.

According to this simplified framework, the person has to decide what level of memory strength is necessary to report an item. If the item is above this criterion then the item is reported; if not, the item is not reported. Because errors are inevitable, since the distributions overlap, the person has to decide the relative values of different kinds of errors. One error, a false alarm, is reporting an unseen item. The other error, a miss, is failing to report an item. Depending on which of these errors is worse, and how much worse, an optimal decision maker will adjust the criterion appropriately (Swets, Dawes, & Monahan, 2000). If the person thinks misses are worse than false alarms, the criterion should be set very low (lenient) so that many items from both distributions are reported. This is called being sensitive because the person is being sensitive to the possible presence of a memory (the term fits better for the most common use of SDT, medical diagnosis; for example, being sensitive to the presence of any cancer cells). This approach will produce a very high hit rate, but also a high false alarm rate. If the person decides that false alarms are much worse than misses, the criterion should be higher (strict). When the criterion is raised, the hit and false alarm rates should both decrease. Having few false alarms is called specificity because the person is trying to respond specifically to items that have been seen before and not to other items. The criterion shown in Figure 1 is strict; it has a high specificity because the proportion of false alarms is relatively small compared with the proportion of misses.

EXPERIMENT 1: FREE RECALL

There has been a large amount of research in the eyewitness literature where people have been given a prompt to recall as much as they can. Most of this has been in evaluating the cognitive interview with respect to eyewitness memory. In the cognitive interview (Fisher & Geiselman, 1992) people are asked to report everything that they can remember. This suggests a lenient response criterion and therefore you would expect sensitivity to be high, at the cost of specificity (Memon & Higham, 1999). This is supported by a meta-analysis that found an increase in the amount of accurate information with the cognitive interview, but also an increase in the amount of errant information (Koehnken, Milne, Memon, & Bull, 1999; see also Roberts & Higham, 2002).

The cognitive interview has several different components and some research has been conducted to try to tease apart the effectiveness of these different components, including the ''report everything'' component. Milne and Bull (2002) found that the adults in their sample actually recalled less accurate and less inaccurate information in this condition than those in a control condition. This is the opposite pattern from that predicted if their manipulation lowered the response criterion. However, their sample size was too small to produce any definitive comparisons (they had 34 adult participants, but spread across six conditions) and they did not observe this pattern with the children in their sample. A further consideration is that the cognitive interview is a complex procedure and its effects depend on many characteristics such as the interviewers' skills and witness characteristics (see Memon, 2006, for a review). Our first study will inform this debate.

The aim of this study is to examine the effect of a shift in response criterion on free recall of both visual material presented and critical items discussed by people. In contrast with Milne and Bull's (2002) findings, we expect from Meade and Roediger's (2002) and Wright et al.'s (2005) studies, and from Koehnken et al.'s (1999) meta-analysis, that shifting the response criterion from one that emphasises only reporting what you are sure about (strict criterion, specificity) to one that emphasises reporting everything you can (lenient criterion, sensitivity) should increase both accurate recall of presented items and increase inaccurate recall of the suggested critical items. Our central question is: Does shifting the response criterion affect suggested items in a similar manner to neutral items?

Method

A total of 88 first-year psychology students at the University of Aberdeen were tested in return for course credit and were allocated to a dyad with a previously unacquainted partner. Three dyads were excluded from analysis because their discussions were too quiet to be transcribed. Thus, data from 82 participants were included (17–56 years; $M = 19.8$; $SD = 5.2$).

Four pictures of complex scenes containing a large number of details were used as stimuli to encode (adapted from Forbes & Venneri, 2003) [these are on available on https://portal.abertay.ac.uk/portal/page/portal/SHARED/f/F513576]. Two versions of each picture (A and B) were created that were identical except for two contradicting details. For example, in a crowded street scene picture, one version shows a *man* walking a dog while the other version shows a *woman* walking a dog. We refer to these as the *critical suggested items*. Each member of a dyad saw a different version of each of the four pictures. Thus over four pictures, dyad members saw eight critical items that differed from what their partner had seen. Pilot data ($n = 19$) showed that the contradicting details of the different versions were remembered approximately as well as each other. The versions were counterbalanced across conditions.

The dyads were randomly assigned to experimental conditions so that both people in the dyad were in the same condition. They were told that they were taking part in a picture memory study and would be shown four pictures that they would be asked to recall both jointly and individually. Participants were in the same condition for each of the four picture trials. Dyad members were led to believe they were seeing identical pictures.

On arrival participants were seated at separate computer desks with their backs to one another. To begin they worked through a filler task that took approximately 5 minutes. Next, participants were shown the first picture for 30 seconds, then did a brief filler task. Dyad members then jointly recalled the picture in as much detail as possible. Then they individually completed a free recall test. Those in the *strict* condition were given the following written recall instructions:

> Please think back to the last picture you saw. Prior research has demonstrated the importance of striving for accuracy and reporting only that which you are certain you remember. In the space below please *only* report details that you remember seeing. Be sure to report only those details that you are sure about and do not attempt to guess at any particular feature. Do not worry if you fail to report details that you are unsure of.

Those in the *lenient* condition received the following instructions:

> Please think back to the last picture you saw. Prior research has demonstrated the importance of reporting everything that you can remember. In the space below please report

as *many* details as possible. Try not to leave out any details, even if you think they are not important. Do not worry if you errantly report details that were not seen, even if you start to feel that you are guessing.

No time limits were imposed for the joint or individual recall tasks. The discussions lasted about 2 minutes each. On completion, participants worked through another filler task individually. Picture 2 was then studied for 30 seconds, with each dyad member seeing either version A or B. Another filler task was then completed before the dyad discussion phase for Picture 2. The same procedure was followed for the individual free recall task, with participants receiving the same strict or lenient recall instructions as before. This procedure was repeated until all four pictures had been shown.

Experimental sessions were audio-taped. The four discussions for each dyad were transcribed and coded for whether each critical item was discussed. The individual free recall responses were coded for the number of accurate and erroneous neutral details from the pictures, and the number of accurate and erroneous critical items. Neutral items are all the non-critical items in the scenes. A critical recalled item was accurate if it was one that they had seen and was inaccurate if it was one their partner had seen but they had not. Data from the four discussions and four individual free recall tests were pooled for the analyses.

This study received ethical clearance from the University of Aberdeen's Psychology Department Ethics Committee.

Results

There were no significant differences between participants who had viewed the different photographs, for the number of correct neutral, incorrect neutral, correct critical, or incorrect critical items recalled (all *ns*). Therefore these groups are combined for the remaining analyses.

The main question is whether the task demand manipulation affected what people recalled and, if so, whether it affected items that were seen and items that were suggested in a similar and predictable manner. The prediction is that more items of each of these types will be recalled in the lenient condition than in the strict condition. Table 1 shows that the amount of accurate and inaccurate information recalled for both neutral and critical items is higher in the lenient condition than in the strict condition.

TABLE 1
Mean accurate and inaccurate neutral items and critical items recalled in the lenient and strict conditions (Experiment 1).

	Items reported at test	
	Accurate	*Inaccurate*
Neutral items		
Strict	44.64 (12.67)	5.81 (3.32)
Lenient	51.84 (13.80)	8.82 (5.37)
Critical items		
Strict	3.90 (1.57)	0.52 (0.74)
Lenient	4.45 (1.78)	1.70 (1.31)

SDs in parentheses.

The standard methods of SDT, calculating hit and false alarm rates, are not easily adopted for free recall data. This is because of the number of potential false alarms being very high due to the complexity of the pictures. Further problems occur with these data, as the responses for inaccurate recalls, for both neutral and critical items, were highly skewed, but the responses for accurate recalls were not. A large percentage of people (37%) did not falsely recall any critical items. Further, because correctly recalling any specific items will mean that the item cannot be incorrectly recalled, these tests are not independent. Different quantitative transformations were applied and led to the same results. Here we report the data based on the ranks for the four variables. Ranking is a common transformation to use when comparing scores among several variables because it makes the distributions of all these variables similar (Conover & Iman, 1981). Using Conover and Iman's classifications of rank transformations, we used an RT-2 transformation (ranking each of the variables). For the neutral items the mean rank for accurate recalls was 35.36 for the strict condition and 47.95 for the lenient condition. For the inaccurate recalls the corresponding mean ranks were 34.75 and 48.59. Because both variables are ranked there is no main effect of correct versus incorrect. There was a main effect of group with those in the lenient condition recalling more neutral items, $F(1, 80) = 9.21$, $p = .003$, $\eta_p^2 = .11$, but no interaction between recall criterion and whether the item was correct, $F(1, 80) = 0.06$, $p = .81$, $\eta_p^2 = .00$. These results are consistent with the manipulation only changing the response criterion.

For accurate recall of critical items, the mean rank for the strict condition was 37.81 and the mean rank for the lenient condition was 45.38. For errant recall the corresponding ranks were 29.73 and 53.86. The ANOVA showed a main effect for the response criterion manipulation, $F(1, 80) = 24.50$, $p < .001$, $\eta_p^2 = .24$, and also an interaction between criterion and whether the recall was correct or not, $F(1, 80) = 5.57$, $p = .02$, $\eta_p^2 = .07$. The interaction corresponds to a larger effect for erroneous recall than accurate recall. This suggests that in recalling a critical item people might be basing their decisions on something more complex than what the traditional unidimensional SDT models presume. However, we urge caution in interpreting this interaction because if somebody erroneously reports a suggested item, that person cannot also report the correct item. This is a limitation of free recall tasks like the one used. An alternative way to examine this is to compare people who reported at least one suggested item with people who did not report any suggested items. In the strict condition, 17 of the 42 people (40%) reported at least one of the suggested items compared with 35 of the 40 people (88%) in the lenient condition. This is statistically significant, $\chi^2(1) = 19.53$, $p < .001$, odds ratio = 10.29 with 95% CI: 3.35–31.60.

Finally, past research has shown that the person who originally brings up an item is more likely to convince the other person (Gabbert et al., 2006). Multilevel logistic regression analysis (Wright, 1997) was performed with ''influenced'' (yes/no) as the outcome variable and two predictors: ''condition'' (lenient/strict) and ''mentioned critical item first'' (yes/no). Item was nested within participant. There was a significant effect of condition whereby participants were more likely to be influenced in the lenient condition, $\chi^2(1) = 28.19$, $p < .001$. Participants were also more likely to be influenced if they were *not* the person who initially mentioned the item, $\chi^2(1) = 58.40$, $p < .001$. There was no significant interaction, $\chi^2(1) < 1$.

Discussion

Several studies have shown that when two people witness and discuss an event, what one person says can influence what the other person later reports. This has been called memory conformity. It is important to understand how the size of this effect can be affected by the task. An important manipulation for both theoretical and applied purposes is changing the response criterion. We used two response criteria: strict and lenient. We found that using a strict procedure lowered the number of times that people erroneously reported what the other person suggested. However, the manipulation had a similar effect on recall of things that were actually seen. From a theoretical stance this suggests that some suggested false information and actual information were at a similar level of memory strength for this experimental context. While we had predicted this effect based on some studies (Meade & Roediger, 2002; Wright et al., 2005), it does run counter to the results from other studies (Milne & Bull, 2002), so it is noteworthy.

From an applied perspective this research provides an explanation for Koehnken and colleagues' (1999) finding that the cognitive interview, which includes the suggestion of a more lenient criterion, should have increased sensitivity, but also decreased specificity. This, of course, does not mean that those using the cognitive interview should necessarily try to prompt people to use a different response criterion. This is because the cognitive interview stresses the need for free recall as opposed to answering direct question. Free recall tends to be much higher in specificity than direct questions or recognition (Koriat & Goldsmith, 1994), so any attempts to shift the response criterion within the cognitive interview must be viewed in the context of changing the question types.

For critical items, there was an interaction between task and the accuracy of item reported. While the shift for accurate critical items was approximately the same as for the neutral items, the shift for inaccurate items was larger. In the lenient condition 88% erroneously recalled at least one of the suggested items, but in the strict condition only 40% did. This means that using a stricter criterion may lessen the proportion of suggested items reported more than it lessens the proportion of non-suggested items. The choice of the appropriate response criterion will therefore depend on the value of accurate recalls versus the cost of inaccurate recalls.

EXPERIMENT 2: RECOGNITION

Experiment 1 showed that manipulating task demands can lessen the amount of false recall

reported by someone after a discussion, but this is at a cost of lessening the amount of accurate recall reported. This is important for eyewitness testimony because interviewers are often trained to get eyewitnesses to adopt a lenient response criterion (Fisher & Geiselman, 1992). In the second study we wanted to test if this finding also occurs with recognition. This is important for eyewitness testimony because in the main forensic recognition task, the line-up, witnesses are told to adopt a strict response criterion and only to identify a person if they are sure.

We also wanted to see whether participants' responses are affected by whether they provide these responses in secret or allow the other person to see the responses. Most of the studies that examine recognition memory in groups have each person provide their answers in the shared environment, so that the other person can see/hear whether they have given the same answer or a different answer. One of the reasons given for memory conformity is that people do not want to appear argumentative with the other person (e.g., Gabbert et al., 2007; Wright et al., 2000). It is possible that people only report what is suggested to them because they do not want the other person to know that they disagree with them. We are able to test this using a new method, which allows the participant to respond in secret.

With the free recall in the first study participants could only respond to a maximum of eight misleading items. Here, with recognition memory, we were able to increase the number to 40. This greatly increases the statistical power of the subsequent analyses.

Method

Participants ($n = 57$; 54% female; age range: 17–37, mean = 22 years) were University of Toledo undergraduates and received course credit. One of three female confederates arrived at the same time. None of the confederates knew any of the participants with whom they worked.

The participant and the confederate sat together approximately 1 metre from a computer screen and were told to pay attention because they would be shown 50 unfamiliar faces. The faces (all white males) were shown for ~ 2 seconds each using PowerPoint. The faces were those used in Exp. 3 of Wright et al. (2005).

When the presentation ended, the participant and the confederate were told that they would be shown another set of faces, and that some of these would be new and some would be old. At this point each pair was randomly assigned to one of the four between-participant conditions, with the restriction that there were approximately equal numbers in each condition. The first factor was the response criterion that they were to use. There were two levels:

- *Strict*: When responding it is important that you only say that a face is "old" if you are certain that you saw the face previously. We do not mind if there are faces that were previously shown that you fail to remember. What is most important is that you do not wrongly say that you saw a face when you did not.
- *Lenient*: When responding it is important that you say that a face is "old" even if you are not certain that you saw the face previously. If you think you might have seen the face, then respond "old". We do not mind if you say "old" to some faces that you had not previously seen. What is most important is that you do not miss any of the faces that you may have seen.

These instructions were read to the pair. We also handed a piece of paper with these on it to each of the people in the pair to stress their importance.

The second factor was the method used to respond. The pair was shown each of 100 faces (half new, half old) and told that person A would respond first, then person B. They were asked to draw one of two pieces of paper out of an envelope. Both pieces said "SECOND", but when they drew them out the confederate said her piece said "FIRST". Thus, the confederate was always person A and the participant person B.

In the *shared* condition, which is what has been used in past studies, there was a single sheet with 50 numbered lines on one side, and 50 on the back, with a column for person A to respond "old" or "new", and then a column for person B to respond "old" or "new". In this condition, person A (the confederate) provided her answer, handed the sheet to person B (the participant), and then person B handed it back to person A, before the next trial began.

In the *secret* condition there were 100 small sheets of paper each labelled with the trial number and pair number, and a column for

person A and person B to respond. Person A responded, person B responded, and then person B placed the card in an envelope. Because of the angle that they were sitting, person A could not see the response of person B. Because person A was a confederate, she was instructed not to look at person B while responding.

The confederate had instructions on how to respond for each item. This means all the participants were receiving the same information from the confederate. The confederate provided correct ''old'' responses for 30 (hits), incorrect ''old'' responses for 20 (false alarms), correct ''new'' responses for 30 (correct rejections), and incorrect ''new'' responses for 20 (misses). In pilot research we have found that if the confederate makes more than 40% errors some participants become suspicious (Wright et al., 2000). The responses were spread throughout the 100 trials with the restriction that half of the hits, false alarms, correct rejections, and misses were in the first 50, and half in the second. Thus, there were four within-participant conditions (2×2 with old or new faces, and whether the confederate said old or new) and four between-participant conditions (2×2 with strict or lenient instructions, and shared or secret responding). In total therefore there are 16 different proportions that we report in the results section.

At the end of the study, all participants were asked what they thought was the purpose of the study. They were asked about the other person, and none mentioned that they thought she was a confederate. When told that the person was a confederate, none expressed any problems associated with this. This study received ethical clearance from the University of Toledo's IRB.

Results

In total there were 5700 trials. In one trial the participant ticked both ''old'' and ''new'' and in two trials neither option was ticked. These three trials (0.05%) are excluded, leaving 5697 trials for analysis. The proportions of ''old'' responses for each of the 16 conditions are shown in Table 2. Some of the effects are clear from this table. First, the effect of the face being previously seen is large ($\sim$30% difference) showing that people did remember the faces above chance levels. Second, the effect of what the confederate says, shown in the ''Diff'' columns, is also large ($\sim$20%); showing that memory conformity occurred in each condition. Both the remembering effect and the memory conformity effect are fairly consistent across the conditions. The effect for the response criterion was in the predicted direction for all eight comparisons, although the individual effects were variable with a range from 2% to 23% with a mean of 9%. The effects of the shared responding versus secret responding are less clear. The effects are smaller than those for the other factors and only relatively large (10%) when the items are old and the confederate says new.

Traditional measures from signal detection theory can be calculated for these data: d′ is a measure of accuracy (how well people discriminate old and new faces) and C is a measure of the response criterion (whether people have a bias for saying ''old'' or ''new''). The value .5 was added to each cell prior to calculating hit and false alarm rates as is the norm. No accuracy corresponds to $d' = 0$ and no bias to $C = 0$. Table 3 shows these values for the different conditions. The mean d′ values have a small range, from 0.72 to 1.12. The mean C values show more variability, from -0.25 to 0.72, and in a systematic and predictable fashion. The C value is higher when the response criterion is strict. The empirical forms of Figure 1 for each d′ and C are shown for these values in the table.

In the past, researchers often have run ANOVAs on individual d′ and C scores. However, because this approach is equivalent to running a generalised linear model for each individual and then running ANOVAs on the resulting coefficients (DeCarlo, 1998), a more efficient (and flexible and powerful) approach is to run a multilevel logistic regression to predict responding ''old'' from the different factors. The (freeware) lme4 package (Bates & Sarkar, 2007) for R was used for these analyses and the syntax for this type of analysis is available on the first author's web page (see example 13 of http://www.sussex.ac.uk/Users/danw/modreg/codedata.htm). We began including all main effects and interactions for the four variables (whether the faces was previously seen [seen], what the confederate said [consay], if the response criterion was lenient or strict [strict], and if the responding was done in secret or not [secret]). We removed effects, one at a time, with the provision that they were not included in any higher-order effect. Therefore we would not remove a main effect of a variable if that variable was included in an interaction term that was still in the model. We began with the effect with the highest *p* value and only removed

TABLE 2
Proportion of "old" responses in the 16 conditions (Experiment 2).

	New items			*Old items*			
Confederate says:	*New*	*Old*	*Diff*	*New*	*Old*	*Diff*	*Row totals*
Shared							
Strict	.14	.36	.22	.39	.63	.24	.38
Lenient	.29	.42	.13	.62	.75	.12	.52
Secret							
Strict	.16	.36	.20	.49	.70	.20	.43
Lenient	.20	.38	.18	.52	.78	.26	.48
Column totals	.20	.38	.18	.51	.71	.21	.45

Diff = the difference between cells to show the effect of what the confederate says.

those with $p > .05$. The final model had a statistically significant interaction ($p < .001$) between secret, strict, and consay. As shown in Table 2, this is due to the confederate effect being smallest (but still 12–13%) when the lenient criterion is used with the shared response format. This interaction was not hypothesised.

The secret by strict by consay term is the only three-way interaction in the model. We removed it to allow interpretation of the remaining effects. We continued to remove effects using the same rules as before. All effects involving secret dropped out and only the main effects for strict, seen, and consay remained. They showed that the following increased the chances of an old response: using the lenient criterion ($z = 3.61$, $p < .001$), having previously seen the face ($z = 24.61$, $p < .001$), and having the confederate say that she had seen it ($z = 15.67$, $p < .001$). All are in the predicted direction.

Discussion

The three main effects observed: (a) that seeing an item previously increased the number of "old" responses, (b) that the confederate saying "old" increased the number of "old" responses, and (c) that suggesting a lenient criterion increased the number of "old" responses, correspond to the three predicted effects. These are the effects for memory, memory conformity, and response criterion respectively. These have important implications, which we cover in the general discussion. Further, these data provide more evidence that a strong memory conformity effect can occur in recognition studies. The size of the memory conformity effect (around a 20% shift) is similar to those reported elsewhere using similar procedures (Wright et al., 2005).

With one exception, no significant interactions were found. We included the response mode manipulation, thinking it might interact with what the confederate said, but the only significant effect including this term was as part of a three-way interaction. That interaction was that the memory conformity effect was smallest (about 12% as opposed to about 20% in the other conditions) when the lenient response criterion was used with shared responding (see Table 2). We had not predicted this interaction, and in fact thought that the memory conformity effect could

TABLE 3
Measures of accuracy (d′) and bias (C) for the recognition data of Experiment 2.

	Confederate says new			*Confederate says old*		
	d′	*C*		*d′*	*C*	
Shared						
Strict	0.86	0.72		0.72	0.02	
Lenient	0.92	0.12		0.90	−0.23	
Secret						
Strict	1.10	0.57		0.86	−0.08	
Lenient	0.95	0.41		1.12	−0.25	

The empirical form is shown in the third column. The scales are the same for all the graphs.

have been larger in the shared conditions where participants might not want to be seen disagreeing with the confederate.

GENERAL DISCUSSION

Past research shows that people can be led to report erroneous information if another person has erroneously reported information in their presence (Gabbert et al., 2003; Wright et al., 2000). In memory research there are two main ways of conceptualising accuracy. The first is accurately reporting the previously presented information. This type of accuracy is often called sensitivity. The second is not falsely reporting information that was not previously presented. This is often called specificity. In our studies half the participants were given instructions that stressed sensitivity and half were given instructions that stressed specificity. Those in the lenient condition reported more accurate information than those in the strict condition, but also more inaccurate information suggested by another person. In the first study this was shown with free recall of complex scenes. In the second study these findings were extended to memory recognition.

The findings in both studies were that the effect of the response criterion manipulation was in the predicted direction for seen and suggested information, for both accurate and inaccurate reporting. This pattern was predicted and is consistent with people using different response criteria for reporting the information. From a theoretical standpoint this suggests the information provided by the other person, whether as part of a dialogue or during testing, is affected by response criterion in a similar way to items actually seen.

What are the implications of our findings from an applied perspective? Those in the lenient condition did accurately report more critical items, but also inaccurately reported the suggested items more than those in the strict condition. The first study provides a laboratory analogue for people witnessing a crime and discussing it together prior to a police interview. Many police forces in the US and the UK receive training in the cognitive interview for interviewing eyewitnesses (Memon, 2006). Other agencies, like the FBI and NASA, are also trained in this procedure (Wells, Memon, & Penrod, 2006). One of the main aspects of the cognitive interview is to report as much as possible. A meta-analysis (Koehnken et al., 1999) showed that this increases the amount of accurate information recalled (increased sensitivity), but with a small increase in the amount of inaccurate information recalled (decreased specificity). Experiment 1 showed that using a more lenient response criterion does increase accuracy as measured by sensitivity, but decreases accuracy as measured by specificity. This was found for both neutral information and information suggested by the other participant. This reinforces the need for police departments to be careful when interviewing eyewitnesses who have spoken with other eyewitnesses, and stresses the importance of recalling accurate information.

The second study is a laboratory analogue for the line-up identification procedure. Current procedures stress that the eyewitness should be sure in their identification (a strict criterion), but how this is done varies by jurisdiction (and probably by the individuals conducting the line-ups). Our research shows that relaxing this criterion would likely increase the number of accurate identifications, but also the number of inaccurate identifications. Further, relaxing the criterion will increase the likelihood that people will identify someone suggested to them by another eyewitness (Skagerberg, 2007).

For applied purposes it is necessary to consider the relative value of different types of memory errors. The focus in this paper has been on the forensic situation. In the forensic (cognitive) interview a lenient criterion is usually used so that eyewitnesses report everything that they can. The idea is that the police can sort out erroneous from accurate information. This is not a bad approach given that free recall tends to be mostly accurate (Koriat & Goldsmith, 1994). In the line-up, a strict criterion is suggested. The eyewitness identification usually occurs later in the police investigation than the interview, because the police have to already have a suspect. An identification is often the final piece of evidence and therefore it would be inappropriate for the police to try to sort accurate from inaccurate suspect-identifications. Therefore, stressing a strict criterion is important although, as is clear from the number of erroneous identifications (about 20% of the time a filler is chosen), there are a great number of identification errors. Our findings can be applied to other applied areas, like education. For example, some tests encourage guessing by not penalising wrong answers (a lenient criterion), while others dis-

courage guessing by penalising wrong answers (a strict criterion).

In summary, memory conformity occurs for both free recall and recognition. By suggesting different response criteria we were able to alter the amount of accurate and suggested information reported. By suggesting a lenient criterion more accurate information is reported, but so is more inaccurate information.

Manuscript received 16 June 2006
Manuscript accepted 29 November 2007
First published online 18 January 2008

REFERENCES

Allan, K., & Gabbert, F. (in press). I still think it was a banana: Memorable lies and forgettable truths. *Acta Psychologica.*

Banks, W. P. (1970). Signal detection theory and human memory. *Psychological Bulletin*, *74*, 81–99.

Bartlett, F. C. (1932). *Remembering: An experimental and social study.* Cambridge, UK: Cambridge University Press.

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

Bates, D., & Sarkar. D. (2007). *lme4: Linear mixed-effects models using S4 classes.* R package version 0.99875-2. http://cran.r-project.org

Blank, H. (2005). Another look at retroactive and proactive interference: A quantitative analysis of conversion processes. *Memory*, *13*, 200–224.

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.

Conover, W. J., & Iman, R. L. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. *American Statistician*, *35*, 124–129.

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

Dalton, A. L., & Daneman, M. (2006). Social suggestibility to central and peripheral misinformation. *Memory*, *14*, 486–501.

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

Echterhoff, G., Hirst, W., & Hussy, W. (2005). How eyewitnesses resist misinformation: Social postwarnings and the monitoring of memory characteristics. *Memory & Cognition*, *35*, 770–782.

Finlay, F., Hitch, G. J., & Meudell, P. R. (2000). Mutual inhibition in collaborative recall: evidence for a retrieval-based account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1556–1567.

Fisher, R. P., & Geiselman, R. E. (1992). *Memory-enhancing techniques for investigative interviewing: The cognitive interview.* Springfield, IL: C. C. Thomas.

Forbes, K. E., & Venneri, A. (2003). A case for case: Handling letter case selection in written spelling. *Neuropsychologia*, *41*, 16–24.

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 & Criminological Psychology*, *9*, 215–227.

Gabbert, F., Memon, A., & Wright, D. B. (2006). Memory conformity: Disentangling the steps towards influence during a discussion. *Psychonomic Bulletin & Review*, *13*, 480–485.

Gabbert, F., Memon, A., & Wright, D. B. (2007). I saw it for longer than you: The relationship between perceived encoding duration and memory conformity. *Acta Psychologica*, *124*, 319–331.

Garry, M., French, L., Kinzett, T., & Mori, K. (in press). Eyewitness memory following discussion: Using the MORI technique with a Western sample. *Applied Cognitive Psychology.*

Hope, L., Ost, J., & Gabbert, F. (in press). ''With a little help from my friends . . .'': The role of co-witness relationship in susceptibility to misinformation. *Acta Psychologica.*

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

Koehnken, G., Milne, R., Memon, A., & Bull, R. (1999). A metaanalysis of the effects of the cognitive interview. *Psychology, Crime & Law*, *5*, 101–116.

Koriat, A., & Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: Distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of Experimental Psychology: General*, *123*, 297–315.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide (2nd ed.).* Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Marsh, E. J., & Tversky, B. (2004). Spinning the stories of our lives. *Applied Cognitive Psychology*, *18*, 491–503.

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

Memon, A. (2006). The cognitive interview. In O. Hargie (Ed.), *The handbook of communication skills* (3rd ed., pp. 531–550). London & New York: Routledge.

Memon, A., & Higham, P. (1999). A review of the cognitive interview. *Psychology, Crime and Law*, *5*, 177–196.

Meudell, P. R., Hitch, G. J., & Boyle, M. M. (1995). Collaboration in recall: Do pairs of people cross cue each other to produce new memories? *Quarterly Journal of Experimental Psychology*, *48A*, 141–152.

Milne, R., & Bull, R. (2002). Back to basics: A componential analysis of the original cognitive

interview mnemonics with three age groups. *Applied Cognitive Psychology*, *16*, 743–753.

Mori, K. (2003). Surreptitiously projecting different movies to two subsets of viewers. *Behavior Research Methods, Instruments, and Computers*, *35*, 599–604.

Mori, K. (2007). A revised method for projecting two different movies to two groups of viewers without their noticing the duality. *Behavior Research Methods*, *39*, 574–578.

Paterson, H. M. (2004). *Co-witnesses and the effects of discussion on eyewitness memory.* PhD thesis, University of New South Wales. Available from: http://www.library.unsw.edu.au/∼ thesis/

Paterson, H. M., & Kemp, R. I. (2006). Co-witnesses talk: A survey of eyewitness discussion. *Psychology, Crime and Law*, *12*, 181–191.

Rajaram, S., & Pereira-Pasarin, L. P. (2007). Collaboration can improve individual recognition memory: Evidence from immediate and delayed tests. *Psychonomic Bulletin & Review*, *14*, 95–100.

Reysen, M. B. (2003). The effects of social pressure on group recall. *Memory & Cognition*, *31*, 1163–1168.

Reysen, M. B. (2005). The effects of social pressure on group recall. *Memory & Cognition*, *31*, 1163–1168.

Reysen, M. B. & Adair, S. A. (in press). Social processing improves recall performance. *Psychonomic Bulletin & Review.*

Roberts, W. T., & Higham, P.A. (2002). Selecting accurate statements from the cognitive interview using confidence ratings. *Journal of Experimental Psychology: Applied*, *8*, 33–43.

Roediger, H. L, 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 lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 803–814.

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.

Skagerberg, E. M. (2007). Co-witness feedback in lineups. *Applied Cognitive Psychology*, *21*, 489–497.

Skagerberg, E. M., & Wright, D. B. (2008). Manipulating power can affect memory conformity. *Applied Cognitive Psychology*, *22*, 146–155.

Skagerberg, E. M., & Wright, D. B. (in press). The prevalence of co-witnesses and co-witness discussions in real eyewitnesses. *Psychology, Crime, & Law.*

Swets, J. A., Dawes, R.M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest*, *1*, 1–26.

Thompson, R. (2002). Are two heads better than one? *Psychologist*, *15*, 616–619.

Weingardt, K. R., Toland, H. K., & Loftus, E. F. (1994). Reports of suggested memories: Do people truly believe them? In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 3–26). New York: Cambridge University Press.

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

Wells, G. L., Memon, A., & Penrod, S. D. (2006). Eyewitness evidence: Improving its probative value. *Psychological Science in the Public Interest*, *7*, 45–75.

Wright, D. B. (1997). Extra-binomial variation in multilevel logistic models with sparse structures. *British Journal of Mathematical and Statistical Psychology*, *50*, 21–29.

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., Mathews, S. A., & Skagerberg, E. M. (2005). Social recognition memory: The effect of other people's responses for previously seen and unseen items. *Journal of Experimental Psychology: Applied*, *17*, 200–209.

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., Startup, H. M., & Mathews, S. A. (2005). Mood, dissociation, and false memories using the Deese-Roediger-McDermott paradigm. *British Journal of Psychology*, *96*, 283–293.

Zho, X-H., Obuchowski, N. A., & McClish, D. K. (2002). *Statistical methods in diagnostic medicine*. New York: John Wiley & Sons, Inc.