Journal of Experimental Psychology:
Learning, Memory, and Cognition
2004, Vol. 30, No. 4, 739–755


0278-7393/04/$12.00 DOI: 10.1037/0278-7393.30.4.739

# Verbalizing Facial Memory: Criterion Effects in Verbal Overshadowing

Joseph Clare and Stephan Lewandowsky
The University of Western Australia

This article investigated the role of the recognition criterion in the verbal overshadowing effect (VOE). In 3 experiments, people witnessed an event, verbally described a perpetrator, and then attempted identification. The authors found in Experiment 1, which included a "not present" response option and both perpetrator-present (PP) and perpetrator-absent (PA) lineups, an increased reluctance to identify a person from both lineup types after verbalization. Experiment 2 incorporated a forced-choice procedure, and the authors found no effect of verbalization on identification performance. Experiment 3 replicated the essential aspects of these results. Consequently, the VOE may reflect a change in recognition criterion rather than a changed processing style or alteration of the underlying memory trace. This conclusion was confirmed by computational modeling of the data.

The veracity of eyewitness memory is of obvious importance, and the search for ways in which eyewitness accuracy can be improved has focused on the conditions at retrieval, which are under the control of police and other practitioners (Fisher & Geiselman, 1992; Gwyer & Clifford, 1997; Kebbell, Milne, & Wagstaff, 1999; Krafka & Penrod, 1985). A common task of eyewitnesses to crimes is to generate a verbal description of the perpetrator, which in turn may lead to apprehension of a suspect for subsequent identification from a lineup or photospread. Although providing a description is standard police procedure, recent research has suggested that this process can adversely affect subsequent identification performance.

Schooler and Engstler-Schooler (1990) first reported an adverse side effect of verbal descriptions. Witnesses viewed a staged crime and then either provided a verbal description of the perpetrator (verbalization condition) or completed an irrelevant filler task (control condition). Following this manipulation, witnesses attempted to identify the perpetrator from a photospread. Witnesses who described the perpetrator were found to make significantly fewer correct identifications than control participants. Schooler and Engstler-Schooler termed the negative influence of verbalization on identification the verbal overshadowing effect (VOE).

The initial studies by Schooler and Engstler-Schooler (1990) were followed by much research that has replicated the VOE numerous times (e.g., Dodson, Johnson, & Schooler, 1997; Fallshore & Schooler, 1995; Ryan & Schooler, 1998; Schooler, Ryan, & Reder, 1996) and across several domains (e.g., basic color memory; Brandimonte, Schooler, & Gabbino, 1997; wine identification, Melcher & Schooler, 1996; and insight problem solving, Schooler, Ohlsson, & Brooks, 1993). Although the effect has on occasion proved difficult to replicate, a meta-analysis of most published and unpublished VOE research by Meissner and Brigham (2001) confirmed the pervasive presence of a small but significant negative effect of verbalization on identification accuracy.

This article investigated the processes underlying the VOE. We first identified and compared the two leading explanations for the effect. This comparison revealed two unresolved empirical issues that, in turn, suggested an alternative explanation for the VOE based on a shift in people's recognition criterion; that is, an increased reluctance to choose someone from the lineup. This criterion explanation is explored in three experiments that manipulated (a) whether witnesses were able to respond "not present" to a lineup and (b) whether the perpetrator was present. When the "not present" option was available (Experiments 1 and 3), witnesses were less likely to choose someone from a lineup after verbalization, irrespective of whether the perpetrator was present. By contrast, when participants were forced to select someone (Experiments 2 and 3), accuracy of identification was unaffected by verbalization. Taken together, the experiments support the idea that verbalization leads witnesses to adopt a more stringent recognition criterion, thus reducing identification rates. Further support for this conclusion is provided by the application of a computational model, WITNESS (Clark, 2003). The model handled the results from all experiments and showed that the VOE, when present, could be captured by a shift in recognition criterion. A competing implementation that modeled the VOE by degrading the memory trace failed to accommodate the data.

## Retrieval-Based Interference

Two major theoretical explanations exist for verbal overshadowing. The first, known as retrieval-based interference (RBI), suggests that the VOE arises from an alteration of the original memory trace during verbalization. This theory was formally proposed by Meissner, Brigham, and Kelley (2001) following the findings of a study in which they manipulated the nature of

Joseph Clare and Stephan Lewandowsky, School of Psychology, The University of Western Australia, Crawley, Australia.

Portions of this research were presented at the Fifth Biennial Meeting of the Society for Applied Research in Memory and Cognition, Aberdeen, Scotland, July 2003, and the Psychology and Law International Interdisciplinary Conference, Edinburgh, Scotland, July 2003. We thank Leo Roberts for his assistance in data collection and Mat Duncan, Simon Farrell, Murray Mayberry, and Gill Rhodes for their suggestions concerning earlier versions of this article.

Correspondence concerning this article should be addressed to Joseph Clare, School of Psychology, The University of Western Australia, Crawley, Western Australia 6009, Australia. E-mail: joe@psy.uwa.edu.au

people's verbalization. When people were forced to provide a detailed and extensive description of the perpetrator, even if this required guessing, a much larger VOE resulted than when people were discouraged from guessing. This suggested that the experimentally induced inaccuracy of descriptions interfered with the earlier memory of the face, similar to the way in which experimenter-provided misinformation interferes with people's memories for earlier events (e.g., Loftus, Miller, & Burns, 1978; Roediger, 1996; Roediger, Jacoby, & McDermott, 1996).

Related findings were reported by Finger and Pezdek (1999, Experiment 1), who compared the effects of different styles of verbalization on identification accuracy. After viewing a photograph of a face, witnesses performed either an elaborate or a standard verbalization task. The elaborate condition led to significantly fewer correct identifications, with the number of correct and incorrect details produced during verbalization predicting identification errors. Finger and Pezdek interpreted these results as a retroactive interference effect caused by the increased quantity of verbalization in the elaborate condition. However, the results can also be interpreted along the same lines as those of Meissner et al. (2001), whereby interference is a function not of the quantity of verbalization per se, but of the amount of incorrect information that is being produced.

Although the studies by Finger and Pezdek (1999) and Meissner et al. (2001) showed a relationship between verbal content and identification accuracy, this relationship did not materialize in many other VOE studies (e.g., Fallshore & Schooler, 1995; Kitagami, Sato, & Yoshikawa, 2002; Schooler & Engstler-Schooler, 1990). A meta-analysis by Meissner and Brigham (2001) offered a possible reconciliation between these two distinct patterns of results based on the nature of the description provided by participants. Meissner and Brigham showed that when participants are encouraged to guess or are forced to provide a detailed description, even when uncertain (called "elaborative description" from here on), the size of the VOE is much greater than when participants are discouraged from guessing. Elaborative descriptions often include errors, which according to Meissner and Brigham, magnify the VOE, thus causing the correlation between the quality of description and identification accuracy. By contrast, more conservative instructions (called *standard* from here on) cause few errors to intrude into descriptions, meaning that the content of verbalization is unrelated to identification accuracy. The selective presence and absence of the correlation between description quality and identification accuracy raises the possibility that verbal overshadowing may arise from two different processes, depending on whether elaborative or standard descriptions are provided. On account of this possibility, all results and theorizing reported in this article involve standard verbalization only. Participants were never encouraged to generate misinformation during verbalization, and we revisit elaborative descriptions only in the General Discussion.

The RBI view has been challenged by several findings, among them the fact that the VOE can generalize beyond the specific face that has been described to other nondescribed faces (e.g., Dodson et al., 1997). In the study by Dodson et al., people viewed photographs of a man and a woman and then described either the male or the female face. Dodson et al. found that, regardless of which face was described, identification performance was degraded for both faces at test when compared to a no-description control group. This generalization of interference is difficult to reconcile with the assumption of RBI that verbalization interferes with the memory trace of the perpetrator (Schooler, 2002; Schooler, Fiore, & Brandimonte, 1997).

## Transfer-Inappropriate Retrieval

The second major theoretical explanation of the VOE, known as the transfer inappropriate retrieval (TIR) account (Schooler et al., 1997), can handle most results that create difficulty for the RBI view. According to TIR, activation of the verbal processes required for the description inhibits the subsequent application of nonverbal face recognition processes without, however, altering the memory of the perpetrator.

The TIR perspective makes at least two predictions that differentiate it from the RBI account. First, there is no expectation that the accuracy of verbalization is related to the accuracy of identification. Instead, all that is required for the VOE to occur is the act of verbalization itself, which produces the presumed processing shift to an inappropriate style. This can explain the generalization of interference to nondescribed faces (e.g., Dodson et al., 1997). Second, the TIR account assumes that the original memory trace merely becomes temporarily inaccessible, rather than being permanently altered by verbalization. In support, Finger and Pezdek (1999) found an attenuation of the VOE following a postdescription delay, and the meta-analysis by Meissner and Brigham (2001) confirmed the generality of that attenuation across existing studies.

For now, we suggest that although there is much support for both leading explanations of verbal overshadowing, each comes with its own set of limitations that cannot be fully reconciled or resolved on the basis of existing research. In particular, research to date has left open two major issues involving (a) the type of responses witnesses can make during identification and (b) the nature of the lineup. It turns out that examination of those issues gives rise to a third possible explanation of the VOE, which forms the principal theoretical contribution of this article.

## A Novel Account: Criterion Effects in Verbal Overshadowing

### *Identification Response Options*

In many VOE experiments, witnesses may reject the lineup entirely by opting for a "perpetrator not present" response (e.g., Meissner et al., 2001; Schooler & Engstler-Schooler, 1990). We call this an *optional-choice methodology*. In some other studies (e.g., Dodson et al., 1997; Ryan & Schooler, 1998), the "not present" option was withheld, thus forcing participants to select one of the lineup alternatives. We call this a *forced-choice procedure*.

Although the difference between those two methodologies may appear subtle at first glance, it gives rise to a new potential explanation of the VOE. Specifically, with an optional-choice methodology, people need to decide not only who the perpetrator is but also whether the lineup should be rejected entirely. The decision about lineup rejection necessarily requires some recognition criterion, such that witnesses say "not present" when all lineup members seem unfamiliar. Conversely, if one or more lineup members exceed a familiarity criterion, the lineup is not rejected,

and an identification decision is made instead (which in turn may be correct or incorrect). The placement of that recognition criterion clearly influences performance: With a conservative criterion, people will be reluctant to identify anyone from a lineup, whereas with a liberal criterion, identification rates will be greater. Criterion effects are pervasive and known to be large in other recognition paradigms that permit voluntary responses, especially when there is a trade-off between quantity and accuracy (e.g., Koriat & Goldsmith, 1994; Koriat, Goldsmith, & Pansky, 2000). It follows that criterion shifts may also play a role in the VOE if verbalization raises people's recognition criterion.

The specific causes of a criterion shift following verbalization remain to be identified. However, there is good reason to assume that people monitor their descriptive ability in the same way they do during other memory tasks (Brigham & Pressley, 1988; Koriat & Goldsmith, 1996; Schraw, 1998). Monitoring of memory performance, in particular when combined with expectations about performance, is known to influence people's actions. For example, Winkielman and Schwarz (2001) manipulated people's beliefs concerning the reasons they experienced difficulty during recall and found that inferences made on the basis of subjective difficulty are shaped by metamemory—that is, one's expectations about memory performance. Winkielman and Schwarz concluded that this pairing of phenomenal experience and expectation influences people's actions. In the present context, people's general inexperience with providing descriptions implies, first, that they may find the task quite challenging and second, that they are unlikely to have a suitable reference against which to compare their descriptions. The combination of those two factors may make people skeptical about their ability to describe a face, which in turn may reduce their willingness to choose someone from a lineup.

An important constraint associated with the criterion explanation is that the reduction in correct identifications must be accompanied by an increase in "not present" responses rather than an increase in false identifications of someone other than the perpetrator. This fairly strong prediction about the pattern of errors differentiates the criterion explanation from the competing RBI and TIR accounts. Specifically, the TIR view appears to consider only correct identification rates and remains mute about the type of errors and their distribution, whereas RBI suggests that most errors should consist of false identifications (Meissner et al., 2001).

Moreover, again in contrast to the RBI and TRI accounts, the criterion explanation also expects the VOE to be absent in forced-choice methodologies in which people necessarily select someone from the lineup. The forced-choice methodology reduces the decision to one of choosing among alternatives, and people's reluctance or willingness to reject the lineup is irrelevant. In research to date, the distinction between optional-choice and forced-choice methodologies has been largely ignored, and people's recognition criterion may therefore have contributed in unknown ways to existing experimental outcomes.

### *Lineup Type*

The second limitation of existing research is that most studies have only presented lineups in which the perpetrator was present (called PP lineups from here on). Few studies have included lineups in which the perpetrator was absent.[1] The importance of such perpetrator-absent (PA) lineups for eyewitness research was emphasized by Wells and Turtle (1986), who noted the parallel between eyewitness identification research and the general signal-detection paradigm in which no experiment would be conducted without "blank" noise-only trials.

The criterion explanation just introduced makes the crucial prediction that verbalization should increase the accuracy of responding to PA lineups. If people adopt a more conservative recognition criterion following verbalization, this should benefit performance with PA lineups because reduced identification rates necessarily imply increased correct rejection rates. This prediction again differentiates the criterion explanation from the RBI and TIR accounts. Although the latter proposals have not been explicitly applied to PA lineups, there is nothing to suggest that they could predict a reversal of the effects of verbalization simply because the perpetrator is not shown at test.

We report three experiments aimed at differentiating the criterion explanation from the RBI and TIR accounts. Following presentation of all results, we formalized the criterion account within a computational model and examined whether it can quantitatively capture the data.

## Experiment 1

Experiment 1 involved optional-choice identification from both PP and PA lineups. To maximize the generality of results, standard descriptions were elicited by questionnaires that targeted either holistic or featural aspects of the face. In line with traditional VOE research, both types of verbalization were compared with a control condition, in which people did not provide a description of the perpetrator.

The TIR account expects both types of verbalization to impair identification compared to the control condition, irrespective of whether or not the perpetrator is present in the lineup because verbalization engages processes that are inappropriate to identification. The RBI account would likewise expect a VOE with both types of lineup because interference occurs before lineup presentation and should not be affected by whether or not the perpetrator is shown at test.

The criterion explanation also predicts that correct identifications should decrease with PP lineups, in line with most research to date. Unlike the TIR and RBI accounts, the criterion explanation additionally predicts that the reduction in correct identifications should be accompanied by an increase in incorrect rejections (misses) rather than more false identifications. Moreover, again in contrast to existing accounts, the criterion perspective also predicts a reduction in (necessarily false) identification rates for PA lineups, thus boosting overall accuracy.

---

[1] Two published examples of VOE research incorporating a PA lineup are Yu and Geiselman (1993) and Meissner (2002, Experiment 1). However, Yu and Geiselman incorporated a 48-hr delay between verbalization and identification. Given Meissner and Brigham's (2001) meta-analytic conclusion that postdescription delays longer than 30 min produce marginal differences between description and control conditions, the relevance of the Yu and Geiselman study is unclear. The results from the Meissner study could also not be directly compared because "don't know" and "not present" responses were combined during analysis.

## Method

*Participants.* The sample consisted of 226 undergraduate students from The University of Western Australia who participated voluntarily and without remuneration.

*Overview and design.* PP lineups were shown to 144 participants, and PA lineups were shown to 82 participants. None of these participants were involved in both conditions. Data were collected across five separate test sessions, each involving a large group of different participants (three and two sessions for the PP and PA lineup, respectively, with lineup type randomly assigned to session). Within test sessions, participants were randomly assigned to one of the three conditions: control (no verbalization), holistic verbalization, or featural verbalization.

Identification decisions and confidence ratings were collected from all participants. Featural and holistic participants also rated the perceived difficulty of the verbalization task.[2]

*Materials.* The PP lineup consisted of a frontal head-and-shoulders photograph of the male perpetrator along with photographs of five similarly posed faces (*foils* from here on). The PA lineup was constructed in the same way and also contained six faces, but the perpetrator was replaced by another foil. Each six-person lineup was constructed as outlined in Brigham, Meissner, and Wasserman (1999), involving a technique recommended by Köehnken, Malpass, and Wogalter (1996) in which a pool of potential foils is first selected on the basis of objective features mentioned in the witness's description (e.g., height, weight, race, build, hair style, etc.). The final foils are then selected from that pool based on similarity ratings (Köehnken et al., 1996).

To obtain the list of objective features, five naïve participants generated descriptions of the perpetrator's face based on two head-and-shoulders photographs (a frontal view, and a 3/4 profile view). Features that were agreed on by three or more of the participants were included in a summated description of the perpetrator. This description was then used to select a pool of 13 potential foils from available face databases. The five faces that were subsequently judged (by another five naïve participants) to be most similar to the summated description of the perpetrator were selected for the lineups.

Both lineups (PP and PA) were assessed for fairness using two measures recommended by Brigham et al. (1999): effective size (ES) and functional size (FS). ES reveals any bias toward or away from the perpetrator within the lineup (Krafka & Penrod, 1985), thus identifying the extent to which foils require a witness to depend on their memory as opposed to deduction and relative comparisons (Lüüs & Wells, 1991). ES is calculated by asking mock witnesses (people who did not witness anything but are given a description of the perpetrator) to pick the person described from the lineup and examining the distribution of choices. Analysis of the responses by 18 mock witnesses for each type of lineup revealed that the ES of our PP and PA lineups was 4.67 and 3.39, respectively. Because both of these measures were greater than half the original lineup size (the criterion suggested by Brigham et al., 1999), both lineups met the ES criterion. We also computed the lineups' FS, which indicates the extent to which a lineup is comprised of implausible foils (Krafka & Penrod, 1985; Lüüs & Wells, 1991). FS is calculated by dividing the total number of mock witnesses by the number of mock witnesses who selected the perpetrator (or the suspect for PA lineups).[3] In the PP lineup, 5 of 18 chose the perpetrator (FS = 3.60), and in the PA lineup, 4 of 18 chose the suspect (FS = 4.50). Because both of these values exceeded half the original lineup size, both lineups also met the FS criterion.

*Procedure.* In all sessions, the perpetrator entered a lecture and distributed handouts throughout the classroom while the lecture continued. Lectures varied in student size from approximately 50 to 130 across the five test sessions. In each case, the lecturer had warned the audience that someone would be entering the classroom. The perpetrator remained present for approximately 4 min, distributing handouts around the room, ensuring that his face was visible to all members of the audience. During this stage, participants were unaware that they would be required to identify the perpetrator at a later time. After the handouts had been distributed, the perpetrator left the lecture venue, and the lecture continued. After a period of time ranging from 20 to 30 min (across test sessions), the lecturer made the audience aware of the nature of the experiment.

Following this announcement, test booklets containing a verbalization questionnaire and an identification form were distributed to the audience. Five minutes were given to provide the verbal description as required by the experimental condition. In the control condition, people had to list members of several different categories (based on the filler task used by Ryan & Schooler, 1998). The holistic and featural questionnaires were developed based on previous experimental methodologies (Finger & Pezdek, 1999; Wells & Hryciw, 1984). The holistic questionnaire contained 20 items requesting ratings of the "averageness" of the perpetrator's face and ratings of traits such as intelligence, friendliness, and honesty. The featural questionnaire, by contrast, contained 22 items querying the perpetrator's hair color, details about eyes, nose, mouth, ears, skin tone, and so on, in which all features were considered in isolation. Participants were presented with a number of alternatives to each question and could also leave items unanswered if they were unsure. This procedure is known as *free recognition* and has been shown to maximize quantity and accuracy of retrieved information (Koriat & Goldsmith, 1994). This style of recognition parallels the demands of the standard description tasks used in previous VOE research.

Following the verbalization phase, participants were informed that the perpetrator may be present in the lineup they were about to view and then either the PP or PA lineup was projected onto a screen. Lineups were arranged into two horizontal rows of three photographs. In the PP lineup, the perpetrator was located at the top-right position. Foils were randomly positioned. Witnesses were allowed 3 min to make an identification decision and could select any one of the six lineup members or indicate a "not present" option in the test booklet. Following identification, confidence ratings were obtained regarding the identification decision together with a self-report assessment of the difficulty of the verbalization.

## Results

*Identification accuracy.* Identification performance for the PP lineup is shown in Table 1. The pattern of responses differed significantly between conditions, as revealed by chi-square, $\chi^2(4, N = 144) = 12.57$, $p < .015$.[4] On the basis of the recommendations of Rosenthal and Rosnow (1991), the phi-coefficient ($\Phi$) was chosen as the effect size measure in this case. In this instance, the size of the effect, $\Phi = .30$, was moderately large.

In exploring these results further, the proportions of identification responses were compared between conditions (Rosenthal & Rosnow, 1991). The relative proportions of hits (correct identifications), false identifications (incorrect identifications, from here on referred to as false IDs), and misses (incorrectly selecting the "not present" option) were calculated. Initially, the two verbalization conditions were compared with the control condition (by $Z$ test). This analysis displayed two significant differences between the control and holistic verbalization groups ($Z = 2.41$, $p < .02$, $\Phi = .25$ for hits, and $Z = -3.72$, $p < .002$, $\Phi = .39$ for misses) but displayed less impressive differences between featural verbalization and the control group ($Z = 1.23$, $p > .10$, $\Phi = .12$ for hits, and $Z = -1.90$, $p = .06$, $\Phi = .19$ for misses). Further examination revealed that the two verbalization conditions did not differ on any

[2] These ratings were not collected in the first test session.

[3] The suspect was the foil who had been judged to be the most similar to the target during the mock witness evaluation of the lineup.

[4] Two cells (22.2%) had expected counts less than 5.

of the response options. Consequently, the holistic and featural conditions were combined and compared with the control group, producing significant differences for hits, $Z = 2.13$, $p < .05$, $\Phi = .18$, and for misses, $Z = -3.54$, $p < .001$, $\Phi = .30$. In all cases, with verbalization treated independently or in combination, comparisons for the false IDs turned out to be nonsignificant ($Z < 1.13$).

Turning to the PA lineup, the distribution of identification decisions is shown in Table 2. Although the table differentiates between the innocent suspect (who replaced the perpetrator) and the other foils (who were also present in the PP lineup), the analyses considered all false identifications together. The two verbalization conditions again did not statistically differ from each other. Consequently, the analysis compared the combined holistic and featural responses to the control condition, which was significant, $\chi^2(1, N = 72) = 5.43$, $p < .02$, $\Phi = .27$.

*Confidence judgments.* The effect of verbalization on postdecision confidence was examined by separate one-way analyses of variance (ANOVAs) for each lineup type. No significant effects were obtained, with $F(2, 141) = 2.29$, $p < .11$ for PP (see Table 1), and $F(2, 68) = .24$, $p < .80$ for PA (see Table 2).

In a further analysis, all participants in each condition were classified as *choosers* (if they picked someone from the lineup irrespective of whether their response was correct; i.e., hits and false IDs combined) or *nonchoosers* (i.e., miss or correct rejection of the PA lineup), and confidence ratings were compared between the two groups for each lineup type and condition separately (excluding the control condition, for which the number of nonchoosers was insufficient). None of the four *t* tests was found to be significant, using a Bonferroni adjustment to keep the significance level at .05 across tests, with the largest $t(45) = 2.27$.

*Perceived difficulty of verbalization.* The pattern of differences between self-rated difficulty scores was the same for both lineup types, and ratings were therefore aggregated across lineup type. Comparison of subjective difficulty between verbalization conditions revealed a significant difference, $t(57) = 2.21$, $p < .05$, between featural ($M = 3.55$, $SD = .93$) and holistic verbalization ($M = 3.00$, $SD = .98$).

## Discussion

Verbal overshadowing was observed with the PP lineup, expressed as a reduction in the number of correct identifications after verbalization compared with the control condition. This effect was accompanied by a difference in incorrect rejections, whereas the number of false IDs was equivalent between all conditions. The type of verbalization (holistic vs. featural), in turn, had no effect on performance.

Table 1
*Response Type (%) and Self-Report Confidence (M, SD) for the Three Verbalization Conditions for the Perpetrator-Present Lineup in Experiment 1*

| | Verbalization condition | | |
|---|---|---|---|
| Response type | Control ($n = 45$) | Holistic ($n = 47$) | Featural ($n = 52$) |
| Hit | 80.0 | 57.4 | 69.2 |
| False ID | 13.3 | 6.4 | 11.5 |
| Miss | 6.7 | 36.2 | 19.2 |
| Confidence | | | |
| *M* | 5.09 | 4.32 | 4.75 |
| *SD* | 1.84 | 1.73 | 1.62 |

*Note.* ID = identification.

Table 2
*Response Type (%) and Self-Report Confidence (M, SD) for the Three Verbalization Conditions for the Perpetrator-Absent Lineup in Experiment 1*

| | Verbalization condition | | |
|---|---|---|---|
| Response type | Control ($n = 22$) | Holistic ($n = 25$) | Featural ($n = 25$) |
| Correct rejection | 22.7 | 52.0 | 52.0 |
| False ID | 77.3 | 48.0 | 48.0 |
| Suspect | 4.5 | 20.0 | 0.0 |
| Foil | 72.8 | 28.0 | 48.0 |
| Confidence | | | |
| *M* | 3.91 | 3.92 | 4.21 |
| *SD* | 1.63 | 1.75 | 1.64 |

*Note.* ID = identification.

With the PA lineup, participants who performed either type of verbalization were less likely to identify someone than people in the control condition. Given that the lineup did not include the perpetrator, this effect translates into an increase in accuracy after verbalization. As with the PP lineup, there was no difference between the holistic and featural conditions.

What, then, caused the verbal overshadowing in Experiment 1? We suggest that it was unlikely to be related to the confidence with which identifications were made, because confidence ratings did not differ between any of the conditions for either lineup type. This result replicates previous VOE research, which has generally found confidence to be nonpredictive of verbal overshadowing (e.g., Krafka & Penrod, 1985; Schooler & Engstler-Schooler, 1990). Moreover, perceived difficulty of verbalization is also unlikely to have affected performance. The featural task was considered the harder of the two, but perceived difference did not translate into differences in identification decisions. Hence, neither people's postidentification confidence nor perceived difficulty of providing a description can account for the present effects of verbalization.

The fact that verbalization increased accuracy of performance (by reducing false identification rates) when the perpetrator was absent from the lineup is difficult to reconcile with the TIR or RBI account. Concerning the former, if verbalization causes a shift toward verbal processing that is inappropriate for subsequent face recognition, then this should impair recognition equally, whether or not the perpetrator is present in the lineup. Concerning the latter, if verbalization interferes with a memory trace, it is difficult to see how the effects of this interference could differ qualitatively, depending on whether the perpetrator is present in the lineup. Furthermore, the reduced hit rate following verbalization with the PP lineup was accompanied by an increase in misses, not in additional false IDs as RBI would expect. The results of Experiment 1, therefore, challenge both existing accounts of the VOE.

The results of Experiment 1 are, however, readily accommodated by the assumption that verbalization encourages people to

adopt a more stringent recognition criterion, such that a greater extent of resemblance is required before a person is chosen. Unlike the TIR and RBI accounts, this criterion explanation can accommodate the results obtained with the PP and PA lineups.

## Experiment 2

Although the criterion explanation can be formalized in a variety of ways, one of which we explore later within a computational model, it makes at least one immediate generic prediction. If the "not present" option is removed, and people are forced to choose someone from a lineup, the recognition criterion can no longer affect responding. It follows that if the VOE in Experiment 1 arose from a more stringent criterion, the effect should disappear with a forced-choice methodology. This possibility was examined in Experiment 2.

Forced-choice identification has been used in several experiments (e.g., Dodson et al., 1997; Fallshore & Schooler, 1995; Ryan & Schooler, 1998; Schooler & Engstler-Schooler, 1990) and some—but not all—of these studies have reported a VOE. When the effect did occur, it was often limited to the first test trial only, with the effect attenuating across multiple trials (e.g., Fallshore & Schooler, 1995, Experiment 1; Ryan & Schooler, 1998). Experiment 2 therefore used a one-trial forced-choice methodology.

Two possible outcomes can be anticipated for Experiment 2. On the one hand, on the basis of some previous research and the TIR and RBI accounts, the control condition should again outperform the verbalization conditions. On the other hand, if verbalization does not generally interfere with face recognition but instead encourages people to adopt a more stringent recognition criterion (as suggested by Experiment 1), then no verbal overshadowing should occur.

### Method

*Participants.* One hundred thirty psychology students at The University of Western Australia were tested across 14 test sessions in groups ranging in size from 5 to 19 people. All participation was voluntary.

*Overview and design.* Within test sessions, participants were randomly assigned to one of three conditions as in Experiment 1. Between test sessions, groups were randomly allocated to one of two levels of delay between presentation and verbalization: 5 min, or 30 min after viewing the perpetrator.[5] As in Experiment 1, the dependent variables were the identification decision (hits or false IDs), identification confidence, and difficulty of verbalization (the latter for holistic and featural participants only). Unlike Experiment 1, people viewed the perpetrator in a photograph rather than during a staged event.

*Materials.* The perpetrator for Experiment 2 was different from that used in Experiment 1, and new lineups were generated. The photograph of the perpetrator used during initial presentation was a 45° left-profile head-and-shoulders color picture. The test lineup consisted of six frontal head-and-shoulder color photographs, including one of the perpetrator, and five featurally similar foils. Fairness of the lineup was confirmed by an ES of 4.38 and an FS of 6.5 (when presented to $N = 26$ mock witnesses). The six lineup photos were arranged in two horizontal rows of three photos in each. The perpetrator appeared in each of the six possible positions across test sessions.

*Procedure.* A picture of the perpetrator was projected onto a screen for 5 s at the beginning of a small-group lecture, whereupon participants continued normal class activity for 5 or 30 min. Participants were unaware that a memory test would follow until after the delay had elapsed, at which point participants received a test booklet with the verbalization questionnaire and an identification form. Participants were given 3 min to work through the verbalization questionnaires. As in Experiment 1, participants in the control condition listed members of categories (such as names of countries, football teams, food types), whereas the other participants responded to free-recognition multiple-choice questions (holistic, $n = 17$; featural, $n = 19$) concerned with the perpetrator's description. After completion of the verbalization tasks, the lineup was projected onto a screen until participants had completed the forced-choice recognition test and indicated their confidence.

### Results

*Identification accuracy.* Initially, the different identification delay conditions were examined independently. However, a three-way chi-square, $\chi^2(7, N = 130) = 3.48$, $p > .10$, showed that there was no need for inclusion of the delay variable. Consequently, responses were collapsed across the delay factor. Table 3 shows the resulting classification of responses.

When considering all responses in the table, there was no effect of verbalization, $\chi^2(2, N = 130) = 0.42$, $p > .10$, $\Phi = .08$. Similarly, when comparing proportions of hits between conditions by $Z$ test, none of the effects were significant (largest $Z = .64$).

*Confidence judgments.* A one-way ANOVA of the confidence ratings failed to reveal any effect of verbalization, $F(2, 127) = 1.31$ (see Table 3). This replicated the corresponding result of Experiment 1 with a forced-choice methodology.

When the confidence of correct participants was compared with that of people who generated false IDs, a significant difference was found, $t(128) = 5.46$, $p < .001$. Participants who correctly identified the perpetrator ($M = 5.09$, $SD = 1.58$) reported greater confidence than those who produced false IDs ($M = 3.00$, $SD = 1.73$).

*Perceived difficulty of verbalization.* Difficulty ratings for the holistic and featural verbalization were compared by $t$ test, $t(81) = 2.11$, $p < .04$. As in Experiment 1, featural verbalization was judged to be the more difficult of the two. Given the identical level of identification performance for those two conditions, this shows once more that difficulty is not diagnostic of the presence or absence of a VOE.

### Discussion

When the "not present" response option was removed in Experiment 2, the effect of verbalization on identification disappeared. The proportion of correct identifications was the same for all three conditions and was similar in magnitude to the hit rate for control participants with the PP lineup in Experiment 1. The results thus failed to support the TIR or RBI accounts, which expected the VOE to remain (at least qualitatively) identical to that observed with the PP lineup in Experiment 1. Once again, as in Experiment 1, there was no difference between the two verbalization conditions.

[5] Two postencoding delays were used here to compare the delay from Experiment 1 (30 min) with the delay typically used in previous studies involving a photograph stimulus (5 min). Given Meissner and Brigham's (2001) meta-analysis found no significant relationship between postencoding delay and incidence of the VOE, no difference was expected between the identification performance of these two groups.

Table 3
*Response Type (%), Self-Report Confidence, and Difficulty for the Three Verbalization Conditions With a Perpetrator-Present Lineup and Forced Choice in Experiment 2*

| | Verbalization condition | | |
|---|---|---|---|
| Response type | Control ($n$ = 43) | Holistic ($n$ = 42) | Featural ($n$ = 45) |
| Hit | 86.0 | 81.0 | 84.4 |
| False ID | 14.0 | 19.0 | 15.6 |
| Confidence | | | |
| *M* | 5.09 | 4.69 | 4.49 |
| *SD* | 1.74 | 1.62 | 1.94 |
| Difficulty | | | |
| *M* | | 3.68 | 4.31 |
| *SD* | | 1.57 | 1.09 |

*Note.* ID = identification.

In replication of Experiment 1, confidence in the recognition decision did not differ between conditions. In further replication of the first study, perceived difficulty of the featural verbalization was greater than that of the holistic verbalization. Both self-report measures thus showed the same pattern across experiments, suggesting that they are not good predictors of identification performance, which differed considerably between the two studies.

One additional methodological difference between the two studies was the manner of stimulus presentation. Whereas Experiment 1 used a live event, a static photograph was used in Experiment 2. The meta-analysis by Meissner and Brigham (2001) considered this methodological factor as a potential moderating variable but concluded very clearly that it was unrelated to the presence or absence of the VOE. Consequently, it is unlikely that the observed variation in performance across Experiments 1 and 2 can be attributed to the differences in stimulus presentation.

Instead, we suggest that the altered pattern of identification performance is best attributed to the removal of the "not present" response option in Experiment 2. When considered together, the results of Experiments 1 and 2 support the assertion that the VOE is caused by a more stringent recognition criterion that can only affect identification rates when a "not present" response is available. Because this conclusion is at variance with the dominant explanations of verbal overshadowing, we sought to support it further by comparing both response options within a single experiment.

## Experiment 3

Experiment 3 combined the optional-choice methodology of Experiment 1 with the forced-choice procedure of Experiment 2 under otherwise identical conditions. On the basis of the findings of Experiments 1 and 2, the following predictions can be made. First, when identification is optional, it is expected that verbalization will cause people to adopt a more conservative recognition criterion (as in Experiment 1, PP condition), thus reducing the number of identifications (and increasing misses) compared with the control condition. Second, when participants are forced to choose from the lineup, no such shift is possible, and verbalization should not influence identification accuracy (replicating the findings of Experiment 2).

### Method

*Participants.* One hundred eighteen psychology students at The University of Western Australia participated voluntarily without remuneration during regularly scheduled small seminars. Participants were tested in groups ranging in size from 8 to 18 people.[6]

*Overview and design.* Experiment 3 incorporated a 2 (verbalization: control, holistic) × 2 (identification: optional, forced choice) between-subjects design. Holistic verbalization was chosen because it generated the biggest criterion shift in Experiment 1 and thus was considered the most likely to produce a VOE under forced-choice conditions.

As in both previous experiments, participants were randomly assigned to either the control or holistic condition within test sessions, with the type of identification (forced versus optional choice) manipulated between sessions. Given that the present forced-choice conditions were identical to those used in Experiment 2 (with only the delay between witnessing and verbalization varying), more participants were assigned to the optional-choice than forced-choice condition. Participants in the optional-choice condition were instructed that, "As in a real police lineup, the person you witnessed earlier may or may not be included in the lineup, and that you should keep this in mind when responding to the identification question," whereas those in the forced-choice condition were instructed to respond to the identification question as best they could.

Once again, the dependent variables collected were the identification decision (hit, false ID, or where possible, miss), and identification confidence. The holistic condition also reported the difficulty of verbalization.

*Materials.* The stimuli were the same as those used in Experiment 2. The previously reported measures of lineup fairness thus also apply here.

*Procedure.* At the beginning of a small seminar, the perpetrator's picture was projected onto a screen at the front of the class for 5 s. No information was given about the eventual memory test. After this presentation, all participants commenced normal class activity and worked for approximately 60 min. Once this delay had elapsed, participants were provided with a test booklet including the verbalization questionnaire and an identification form. Participants were given 3 min to work through the verbalization questionnaires. As in the first two studies, participants in the control condition listed members of categories (such as names of countries, football teams, food types), whereas the verbalization participants answered 17 free-recognition multiple-choice questions targeting the perpetrator's holistic appearance. Both questionnaires were identical to those used in Experiment 2. Directly following verbalization, a lineup was projected onto the screen at the front of the class, and participants then recorded the identification decision and a confidence judgment in their test booklets.

### Results

*Identification accuracy.* The results are shown in Table 4. Initially, the false IDs and misses from the optional-choice conditions were collapsed to render both choice types commensurate. Comparison of identification decisions across optional- and forced-choice lineups revealed a significant difference between the conditions, $\chi^2(3, N = 118) = 16.14$, $p < .001$, $\Phi = .37$. This comparison demonstrated that the frequency of errors varied as a result of identification type.

[6] Data were screened via a postidentification question to ensure none of the participants had completed this task previously. In addition, the location of the target face within the lineup was varied across test sessions to prevent contamination of responses between groups.

Table 4
*Response Type (%), Self-Report Confidence, and Difficulty for the Control and Holistic Verbalization Conditions Involving Both an Optional- and Forced-Choice ID Decision in Experiment 3*

| | Verbalization condition | | | |
|---|---|---|---|---|
| | Optional-choice ID | | Forced-choice ID | |
| Response type | Control ($n$ = 40) | Holistic ($n$ = 40) | Control ($n$ = 19) | Holistic ($n$ = 19) |
| Hit | 62.5 | 47.5 | 84.2 | 94.7 |
| False ID | 15.0 | 7.5 | 15.8 | 5.3 |
| Miss | 22.5 | 45.0 | | |
| Confidence | | | | |
| *M* | 4.08 | 4.67 | 4.32 | 4.53 |
| *SD* | 1.63 | 1.54 | 2.11 | 1.47 |
| Difficulty | | | | |
| *M* | | 4.05 | | 4.68 |
| *SD* | | 1.47 | | 1.38 |

*Note.* ID = identification.

Following this initial comparison, the optional-choice data were analyzed on their own, with misses and false IDs considered separately. The overall chi-square for this comparison approached but did not reach significance, $\chi^2(2, N = 80) = 4.82$, $p = .09$, $\Phi = .25$).[7] Nonetheless, the effect size here was comparable with that observed in the corresponding condition in Experiment 1 (previously, $\Phi = .30$). The specific predictions of the criterion account were further explored by planned comparisons involving misses in the control and holistic conditions. This analysis showed that there were more misses in the holistic condition than in the control group ($Z = -2.19$, $p < .03$, $\Phi = .25$). Thus, once again, verbalization resulted in reduced willingness to choose from the lineup.

In direct contrast to these findings, separate analysis of the forced-choice lineups did not show any difference between the control and holistic conditions, $\chi^2(1, N = 38) = 1.12$, $p > .10$, $\Phi = .17$.[8] Further investigation of the proportion of hits between control and holistic conditions ($Z = -1.07$) suggested that the act of verbalization, if it was having any influence at all, slightly improved performance.

*Comparing identification decisions with Experiments 1 and 2.* To gather further support for a recognition criterion explanation for the VOE, hierarchical log linear (HILOG) analyses were conducted combining the data from this study with the suitable data from the previous experiments. This technique has been used in previous VOE research (e.g., Meissner et al., 2001). Separate analyses were conducted for the optional- and forced-choice data.

Initially, the holistic and control optional-choice data from the PP condition in Experiment 1 were combined with the optional-choice data produced within this study for HILOG analysis. This resulted in a 2 × 2 × 3 model, investigating the relationship between verbalization (control vs. holistic conditions), postexposure delay (which varied between studies from 30 to 60 min), and the outcome of the identification process (hit, miss, or false ID). The results for this analysis displayed a significant Condition × Identification Decision interaction, $\chi^2(2, N = 172) = 16.23$, $p < .001$, consistent with the results of previous analyses suggesting that verbalization systematically influences the type of identification response produced. It is important to note that there was no Postexposure Delay × Identification Decision interaction, $\chi^2(2, N = 172) = 4.01$, $p > .10$, suggesting that there was no qualitative difference between the results of Experiment 1 (PP lineup only) and those produced within this study. Finally, neither the Three-Way Condition × Postexposure Delay × Identification Decision Interaction, $\chi^2(2, N = 172) = 1.64$, $p > .10$, nor the Two-Way Condition × Postexposure Delay interaction, $\chi^2(1, N = 172) = 0.43$, $p > .10$, proved to be significant. This lack of an influence of postexposure delay (i.e., Experiment 1 vs. 3) is, once again, consistent with the meta-analytic findings of Meissner and Brigham (2001).

Furthermore, a 2 × 2 × 2 HILOG analysis of the forced-choice data from the holistic and control conditions of Experiment 2 and the corresponding conditions of the present experiment also supported the recognition criterion perspective. This model included verbalization condition (control or holistic), the postexposure delay (30 or 60 min; i.e., Experiment 2 vs. 3), and the identification decisions (hit or false ID). The analysis found nonsignificant results for the Three-Way Condition × Postexposure Delay × Identification Decision interaction, $\chi^2(2, N = 123) = 1.52$, $p > .10$, and for all of the two-way interactions (all $\chi^2 < 1$), further suggesting that verbalization had no effect in either study.

In combination, the HILOG analyses provided solid support for the recognition criterion explanation of the VOE. Verbalization was consistently found to have no effect on identification accuracy under forced-choice conditions, whereas with optional choice, there was an effect of verbalization, but that effect was reflected primarily in nonidentifications (misses).

*Confidence judgments.* Confidence ratings for Experiment 3 are displayed in Table 4. The effect of verbalization condition (control or holistic) and type of identification (forced- or optional-choice) on confidence was examined by ANOVA. This analysis produced no significant interaction and no significant main effects (all $F$s < 1.53). Once again, as in both previous experiments, verbalization did not affect confidence. In contrast to Experiment 2, confidence here did not differ between participants who chose the correct person compared with those who selected a foil (all $F$s < 2.47 in ANOVA with type of response as variable).

*Perceived difficulty of verbalization.* The holistic verbalization difficulty ratings are displayed in Table 4. There was no difference between the perceived difficulty of the optional- and forced-choice conditions, $t(56) = 1.57$. This finding was not unexpected, given that conditions did not differ until after verbalization was completed.

## Discussion

This study replicated the general pattern observed in Experiments 1 and 2 within a single experiment. Varying only the opportunity for witnesses to reject the lineup, verbalization had an effect when it could shift the recognition criterion (viz. in the optional-choice case) but not when the recognition criterion was irrelevant (viz. in the forced-choice conditions). Furthermore, as

[7] Two cells (33.3%) had expected counts less than 5.
[8] Two cells (50.0%) had expected counts less than 5.

was demonstrated by the HILOG analysis, the results from Experiment 3 were consistent with the data from the first two experiments, and the additional postencoding delay involved here made no difference to the overall pattern of performance.

The pattern of confidence judgments also conformed to those produced in the preceding studies. Once again, confidence failed to distinguish between the control and verbalization groups. Furthermore, in this instance, confidence did not differ between those participants who were correct and those who chose a foil.

We now provide further support for the criterion explanation by fitting a quantitative model to the data from all three experiments. The modeling not only examined the criterion explanation at a quantitative level but also permitted a comparison of the criterion explanation to an interference-based memory process within the same architecture.

## Application of the WITNESS Model

The importance of computational modeling in cognition has been repeatedly noted (e.g., Hintzman, 1991; Lewandowsky, 1993), and in areas such as categorization (e.g., Kruschke, 1992) or short-term memory (e.g., Farrell & Lewandowsky, 2002) virtually all research is driven or guided by sophisticated computational models that have clearly contributed to the theoretical maturity of those research areas. By contrast, research in eyewitness identification and allied fields, including verbal overshadowing, has proceeded largely without any rigorous theoretical guidance. One recent exception involves the WITNESS model proposed by Clark (2003). The WITNESS model is the first theory of recognition to be applied to eyewitness identification and its methodological peculiarities and constraints.

## WITNESS: Architecture and Parameters

WITNESS is a direct-access matching model (e.g., Clark & Gronlund, 1996) in which recognition decisions are based on comparisons between the test items and the contents of memory. WITNESS is based on the following architectural principles: (a) All stimuli (i.e., faces of the perpetrator and foils) are represented as random vectors with features drawn from a uniform distribution with mean zero (range −.5 to .5). The number of features was 100 for all simulations below. (b) Encoding is imperfect, such that only a proportion $s$ ($0 < s < 1$) of features are veridically copied into a memory vector (called **M**) when the perpetrator is viewed. The remaining $1 - s$ features are stored incorrectly by sampling from the same uniform distribution. (c) At retrieval, all faces in the lineup are compared with memory by computing the dot product between the vector representing each face and **M**. The recognition decision relies entirely on the match represented by the set of dot products. (d) In Clark's (2003) version of WITNESS, recognition involves a weighted combination of an absolute-match strategy (the extent to which the best match resembles memory for the perpetrator) and a relative-match strategy (the extent to which the best match is better than the next-best match). A person is selected from the lineup if the sum of both sources of evidence exceeds an identification criterion. For the present modeling, this process was simplified and only the absolute match was considered, thus eliminating two parameters from the model. Specifically, if the best match between a lineup member and memory exceeded the recognition criterion, $c_{rec}$, the model chose the best match as its response. If all matches fell below $c_{rec}$, the model rejected the lineup and made a *not there* response. For modeling of forced-choice lineups, the recognition criterion was removed and the model identified the person with the best match, irrespective of its magnitude. (e) Much of the model's power derives from specifying a potentially complex similarity structure between the perpetrator, the foils, and an innocent suspect who replaces the perpetrator in PA lineups. For the present modeling, the similarity between the perpetrator and all foils (including the face that replaced the perpetrator in PA lineups) was assumed to be equal. That common similarity was captured by the parameter *sim* which determined the proportion of features ($0 < sim < 1$) that were identical between two vectors (i.e., the perpetrator would share a proportion *sim* of random features in common with each foil).

Thus, the version of WITNESS used here involved the following three basic parameters: the encoding strength $s$, the similarity between the perpetrator and all other lineup faces *sim*, and the recognition criterion $c_{rec}$. The recognition criterion was used only when modeling the results of Experiment 1 or the optional-choice lineups in Experiment 3 (in which a "not present" option was available). The effects of verbalization were modeled in two ways. First, as suggested by the present experiments, verbalization was mapped into a shift in the recognition criterion. Second, for comparison purposes, verbalization was modeled by reducing the quality of the memory representation of the perpetrator. The latter implementation follows the tenets of the RBI account.

## Verbalization as a Criterion Shift

To implement the criterion explanation, several criterion parameters were estimated as follows. First, a separate value of $c_{rec}$ was estimated for the control conditions in Experiment 1, $c_{rec}$(C1), and Experiment 3, $c_{rec}$(C3). The effect of verbalization was then represented by an increment of those control criteria, which was constant across experiments but differed between verbalization type. Specifically, holistic and featural verbalization incremented the criterion by a value determined by the parameters $incr_{rec}$(H) and $incr_{rec}$(F), respectively.

Altogether, this version of WITNESS thus included six free parameters that were estimated while fitting the model simultaneously to the data from all three experiments: $s$, *sim*, $c_{rec}$(C1), $c_{rec}$(C3), $incr_{rec}$(H), and $incr_{rec}$(F). Parameters were estimated using standard SIMPLX techniques by minimizing the residual mean squared deviation (RMSD) between the data and model predictions.[9] Predictions were based on 1,000 simulation replications at each step during the minimization process. The best-fitting parameter estimates, obtained with an RMSD of .057, are displayed in Table 5.

At a quantitative level, the fit of WITNESS was very good, as the minimum RMSD corresponded to an average difference between the observed and predicted response proportions of under 6%. Moreover, the model also captured the qualitative pattern in

[9] Because the model, by design, could not differentiate between conditions involving forced-choice identification, only the average of the forced-choice data was considered during minimization. This is tantamount to weighting the optional-choice data somewhat more but has no other effects on the fitting procedure.

Table 5
*Best-Fitting Parameter Values for the Application of WITNESS to Experiments 1, 2, and 3 When Only Recognition Criterion ($C_{rec}$) Was Manipulated Between Conditions*

| Parameter | Best-fitting estimate | |
|---|---|---|
| | $C_{rec}$ fit to whole data set | $C_{rec}$ fit to PP optional-choice data only |
| Encoding strength (*s*) | 0.27 | 0.26 |
| Similarity (*sim*) | 0.28 | 0.28 |
| Recognition criterion control in Exp. 1 & 2 [$c_{rec}$(C1)] | 1.20 | 1.27 |
| Recognition criterion control in Exp. 3 [$c_{rec}$(C3)] | 1.65 | 1.61 |
| Change to holistic recognition criterion [$incr_{rec}$(H)] | 0.61 | 0.66 |
| Change to featural recognition criterion [$incr_{rec}$(F)] | 0.42 | 0.39 |

*Note.* Exp. = Experiment; PP = perpetrator-present.

the data. This is underscored by Figures 1, 2, 3, 4, and 5, which summarize the data from the three experiments (using the vertical bars) and the model's predictions (solid black lines and open squares). It is clear from the figures that the model captured the results of all experiments by varying the recognition criterion, $c_{rec}$, when verbalization was required and a "not present" option was available. When that option was unavailable, as in Experiment 2 and in the forced-choice lineups in Experiment 3, the data were captured by simply removing $c_{rec}$ and leaving the remaining two parameters unchanged. This necessarily forced the model to make identical predictions for all forced-choice conditions, which implies a slight but not drastic departure from the data.

Another way in which the explanatory power of a model can be explored is by fitting it to a subset of the data and then using the best-fitting parameter estimates to make predictions for other, preferably different, aspects of the results. This approach was taken in a second simulation, in which the model was fit only to the optional-choice PP lineups of Experiments 1 and 3. Using the same six parameters as before, the minimum RMSD was .031 (see Table 5 for parameter estimates and the solid black squares in Figures 1 and 4 for model predictions). Keeping parameters unchanged, predictions were then derived from the model for PA lineups and the forced-choice conditions. As shown in Figures 2, 3, and 5, the model captured the main aspects of those results without any further data fitting. It follows that the model's success did not merely reflect some unbounded flexibility that enabled it to satisfy multiple simultaneous constraints during data fitting. Instead, the model's predictions in different experimental situations were tightly linked by its core architectural principles.

Taken together, the two simulations provide strong quantitative support for the assertion that the VOE represents a shift in the recognition criterion that occurs as a consequence of providing a verbal description and that makes people more reluctant to choose anyone from a lineup, irrespective of whether the perpetrator is present. This conclusion is further strengthened by the examination of the competing model, in which verbalization affected the quality of memory representation.

## Verbalization as a Memory Decrement

One favorable characteristic of WITNESS is its parsimony. However, that parsimony entails the drawback that the model contains no obvious analog to processes such as retrieval-based inhibition or transfer inappropriate processing that have been put forward as underlying the VOE. We therefore implemented a fairly generic alternative to the criterion view, which assumed that verbalization compromised the integrity of the memory trace. Within the model, this was represented as a reduction in the encoding parameter, *s*. (Although verbalization follows encoding, a reduction in *s* was an appropriate vehicle for modeling because the effect of a reduction is isomorphic to overwriting the memory trace with varying extents of noise.) Conceptually, this implemented a variant of the RBI explanation because verbalization affected memory representation rather than processing style.

This version of the model was parameterized as follows: The similarity parameter, *sim*, was constant across all conditions and experiments, as was the single value of the recognition criterion, $c_{rec}$. A value of the encoding parameter was estimated separately for the control conditions in Experiments 1 and 2 on the one hand, *s*(C12), and for Experiment 3 on the other, *s*(C3). The effects of verbalization were captured by two additional parameters, *decs*(H) and *decs*(F), which reduced the magnitude of the appropriate value of *s* for the holistic and featural condition, respectively. Altogether, this version of the model thus also included six parameters: *sim*, $c_{rec}$, the two baseline values of encoding strength *s*(C12) and *s*(C3), and their reduction through the two types of verbalization, *decs*(H) and *decs*(F).

Because there is every expectation that this model would not be able simultaneously to capture the presence and absence of the VOE, parameters were estimated using the data from the optional-choice conditions in Experiments 1 and 3 only. The forced-choice and PA lineups were excluded during the fitting process, thus maximizing the model's chances of handling at least a subset of the data.

The results are shown as the broken black lines with open triangles in Figures 1, 2, and 4 (with best-fitting parameter estimates shown in Table 6). Numerically, the goodness-of-fit index (RMSD = .078) was 2–3 times greater than that obtained when the criterion model was fit to the same subset of the data. Moreover, the figures clarify that the memory-decrement version of WITNESS failed to capture the main aspects of the data, namely the presence of verbal overshadowing. The reason for this failure lies in the fact that in the data, the reduced hit rate associated with verbalization is accompanied not by an increase in false identifi-

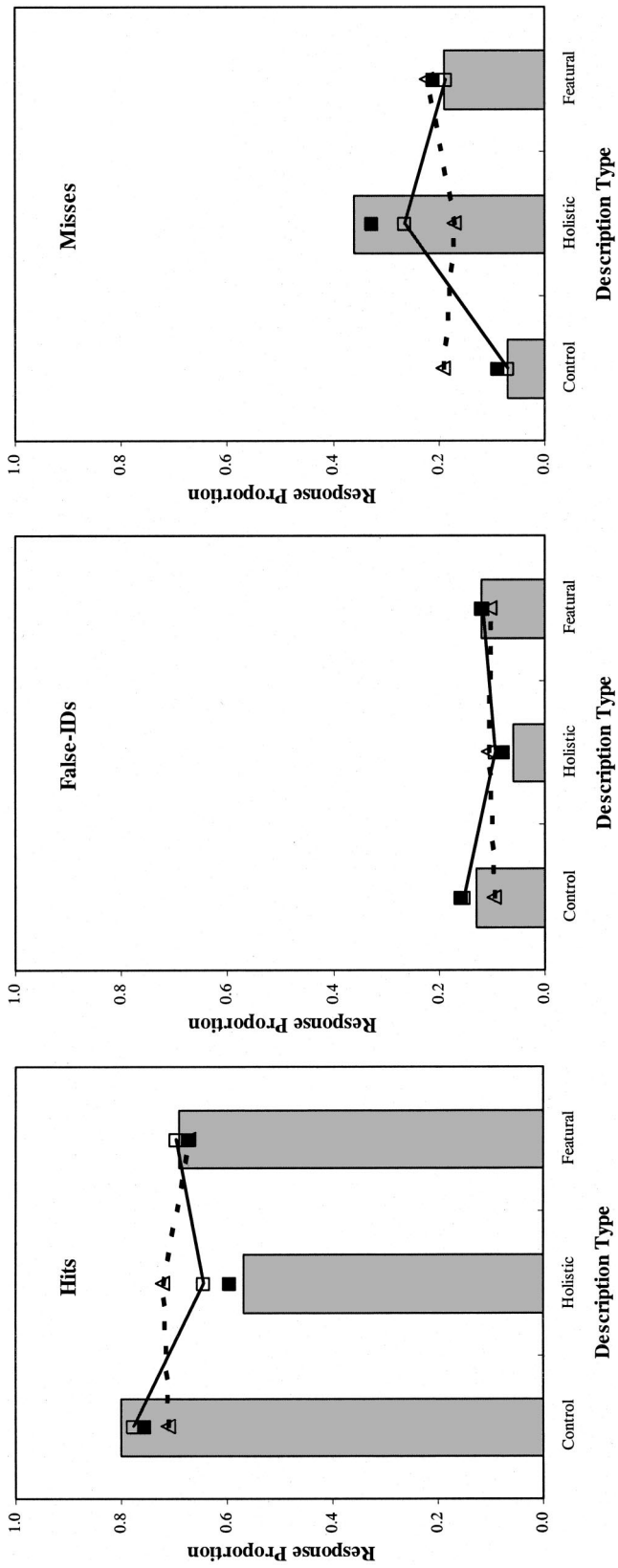


*Figure 1.* Experiment 1 perpetrator-present (PP) lineup data (bars) and three fits of WITNESS: (a) recognition criterion ($c_{rec}$) manipulated and fit to the total data set (indicated by solid lines and open squares), (b) *s* manipulated and fit to the PP optional-choice data only (indicated by dashed lines and open triangles), and (c) $c_{rec}$ manipulated and fit to the PP optional-choice data only (indicated by solid squares). ID = identification; *s* = proportion of the 100 encoded features veridically copied into the memory vector.

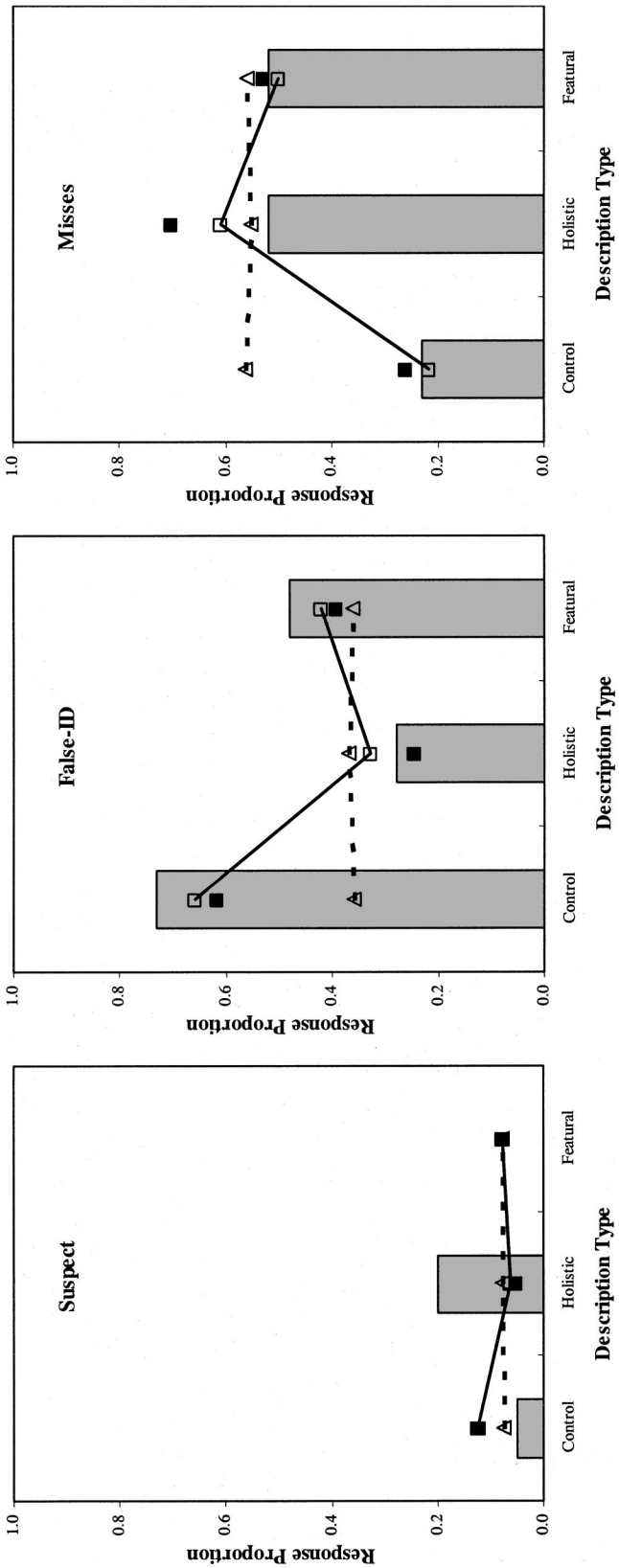


*Figure 2.* Experiment 1 perpetrator-absent lineup data (bars) and three fits of WITNESS: (a) recognition criterion ($c_{rec}$) manipulated and fit to the total data set (indicated by solid lines and open squares), (b) *s* manipulated and fit to the perpetrator-present (PP) optional-choice data only (indicated by dashed lines and open triangles), and (c) $c_{rec}$ manipulated and fit to the PP optional-choice data only (indicated by solid squares). ID = identification; *s* = proportion of the 100 encoded features veridically copied into the memory vector.

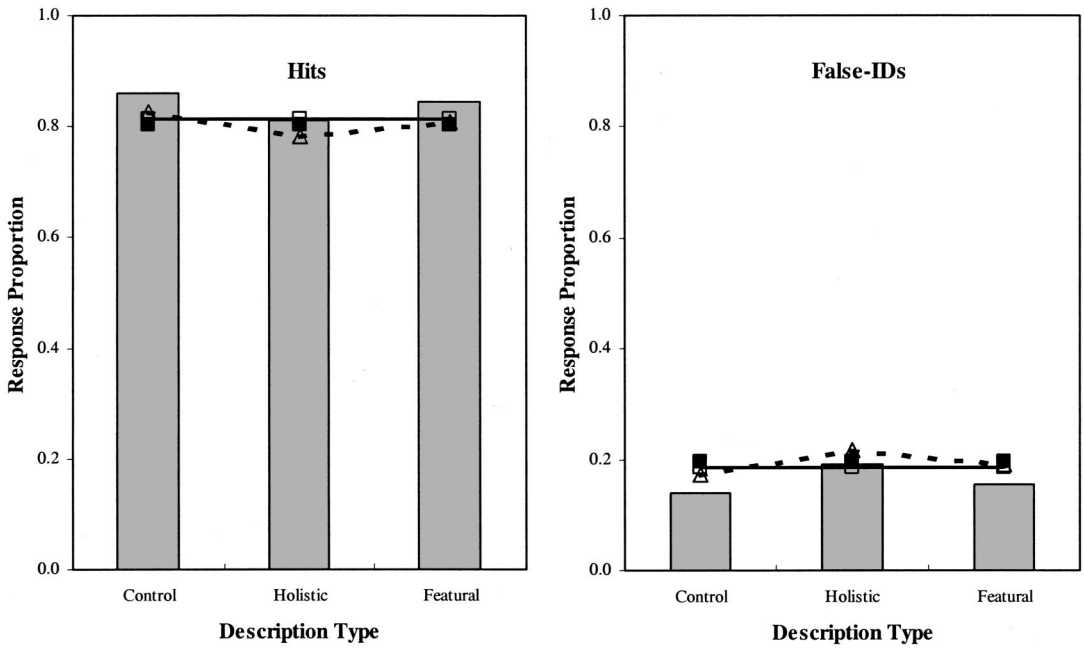


*Figure 3.* Experiment 2 lineup data (bars) and three fits of WITNESS: (a) recognition criterion ($c_{rec}$) manipulated and fit to the total data set (indicated by solid lines and open squares), (b) *s* manipulated and fit to the perpetrator-present (PP) optional-choice data only (indicated by dashed lines and open triangles), and (c) $c_{rec}$ manipulated and fit to the PP optional-choice data only (indicated by solid squares). ID = identification; *s* = proportion of the 100 encoded features veridically copied into the memory vector.

cations—which would be expected if memory deteriorates—but by additional erroneous lineup rejections (misses).

The model's failure is further confirmed by examining the predictions for PA lineups, which were obtained with the same best-fitting parameter values. The model again shows no evidence of verbal overshadowing for these lineups, quite contrary to the data. (The fact that the forced-choice results are captured, at least at a qualitative level, is of little consolation because it only reflects another instance of the model's pervasive inability to differentiate between verbalization conditions rather than a specific prediction). The failure of the memory-decrement model underscores the explanatory uniqueness of the criterion-shift explanation.

## General Discussion

### *Summary of Results*

Experiment 1 used standard description instructions and produced a strong effect of verbal overshadowing with a PP lineup. Participants in the two verbalization conditions were less likely than control participants to identify the perpetrator. Because false identification rates did not differ between conditions, this effect was due to people's increased reluctance to choose someone from the lineup following verbalization. The same pattern occurred with a PA lineup, with verbalization participants once again being more reluctant to choose someone. Because with a PA lineup any choice necessarily represents an error, verbalization improved performance in this case. This novel result escaped detection in most previous research, which did not consider PA lineups.

Next, Experiment 2 used forced-choice identification, which renders people's recognition criterion irrelevant. In line with the criterion explanation and contrary to the expectation of competing explanations, verbalization had no effect on identification in that study. The combined pattern of the first two experiments was replicated in Experiment 3, which under otherwise identical circumstances manipulated only the type of identification. As in the first two studies, optional-choice participants were found to be more conservative following verbalization, but no effect of verbalization was obtained when identification was forced.

Finally, the results from all three experiments were accommodated by the WITNESS model. With few free parameters, and by only varying the recognition criterion between optional-choice conditions, all of the data could be accounted for with considerable quantitative precision. Furthermore, the criterion version of WITNESS was found to be superior to an alternative model that held criterion constant while varying integrity of the memory trace. We next examine the implications of our results for the competing explanations of verbal overshadowing before considering the limitations of the criterion view.

### *Implications for the TIR Theory*

The TIR theory suggests that verbalization directs cognitive processing toward an inappropriate style, which inhibits retrieval of the nonverbal information required for facial recognition. This central tenet of the TIR theory was challenged by our results in several ways. First and foremost, Experiments 2 and 3 failed to find a VOE with forced-choice identification. Second, with

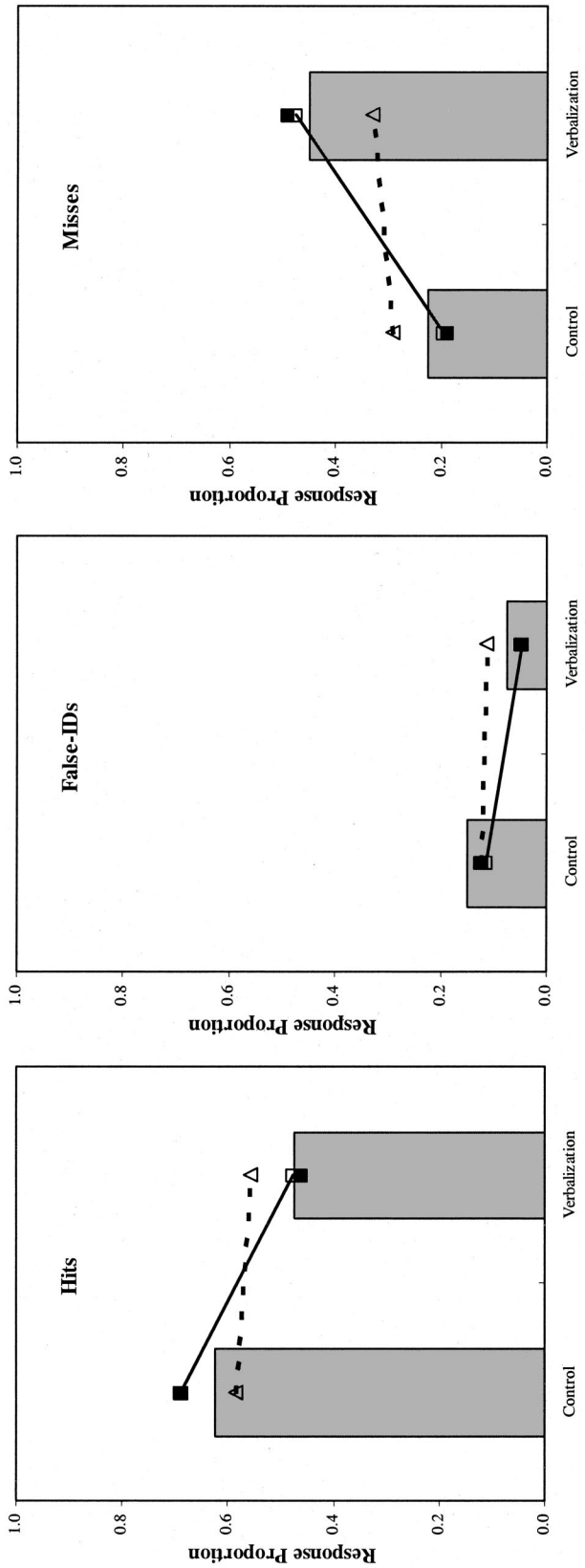


*Figure 4.* Experiment 3 lineup data from optional-choice condition (bars) and three fits of WITNESS: (a) recognition criterion ($c_{rec}$) manipulated and fit to the total data set (indicated by solid lines and open squares), (b) *s* manipulated and fit to the perpetrator-present (PP) optional-choice data only (indicated by dashed lines and open triangles), and (c) $c_{rec}$ manipulated and fit to the PP optional-choice data only (indicated by solid squares). ID = identification; *s* = proportion of the 100 encoded features veridically copied into the memory vector.

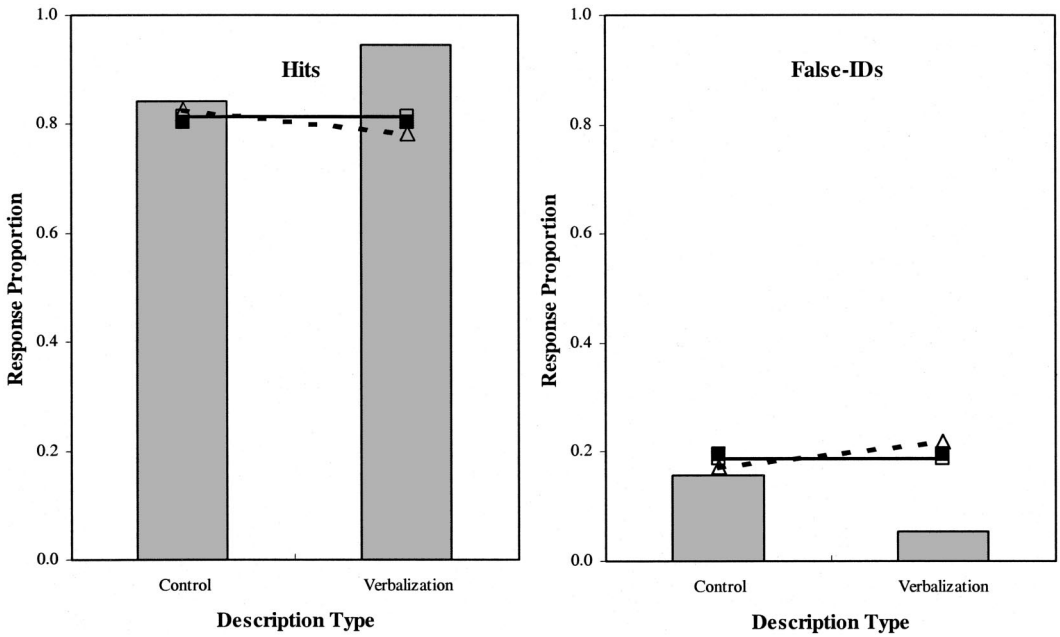


*Figure 5.* Experiment 3 lineup data from forced-choice condition (bars) and three fits of WITNESS: (a) recognition criterion ($c_{rec}$) manipulated and fit to the total data set (indicated by solid lines and open squares), (b) *s* manipulated and fit to the perpetrator-present (PP) optional-choice data only (indicated by dashed lines and open triangles), and (c) $c_{rec}$ manipulated and fit to the PP optional-choice data only (indicated by solid squares). ID = identification; *s* = proportion of the 100 encoded features veridically copied into the memory vector.

optional-choice identification, verbalization increased accuracy with PA lineups, notwithstanding the fact that the mismatch between the processes required for description and identification is unaffected by the type of lineup. It follows that the TIR view can only handle a small subset of our data, namely the optional-choice PP results from Experiments 1 and 3.

Schooler (2002) recently acknowledged and addressed other known issues for the TIR view, for example, the fact that identification accuracy can be affected even if no retrieval operations are engaged (Macrae & Lewis, 2002) and that unrelated nonverbal processes can alleviate the effect of verbalization on identification (Finger, 2002). In response, Schooler proposed a modification to the TIR view, which no longer considers retrieval to be essential for the proposed processing shift. Instead, under this more general framework, these other types of task may have similar effects. Schooler claimed that this generalized TIR view is consistent with all existing findings, including cases in which a VOE was absent. Schooler suggested that when a VOE is observed, sufficient verbal processing has taken place to induce a shift to an inappropriate processing style. By contrast, a failure to observe the VOE simply represents a failure to induce sufficiently extensive verbal processing. Although attractive at first glance, this explanation runs the risk of being circular and nonfalsifiable because no independent criteria exist for identifying the extent of verbal processing.

Table 6
*Best-Fitting Parameter Values for the Application of WITNESS to the PP Optional-Choice Data From Experiments 1–3 When Only s Was Manipulated Between Conditions*

| Parameter | Best-fitting estimate |
|---|---|
| Similarity (*sim*) | 0.29 |
| Recognition criterion ($c_{rec}$) | 1.78 |
| Encoding strength for Exp. 1 & 2 control [*s*(C12)] | 0.29 |
| Encoding strength for Exp. 3 control [*s*(C3)] | 0.25 |
| Reduction to holistic encoding strength [*decs*(H)] | 0.01 |
| Reduction to featural encoding strength [*decs*(F)] | 0.01 |

*Note.* PP = perpetrator present. *s* = proportion of the 100 encoded features veridically copied into the memory vector; Exp. = Experiment.

The criterion explanation also accommodates the presence (e.g., our Experiments 1 and 3) as well as the absence (e.g., Experiments 2 and 3) of verbal overshadowing. However, unlike Schooler's (2002) modified TIR, the criterion explanation postulates very clear and independently identifiable circumstances under which the VOE should or should not arise. Thus, the criterion explanation may provide a principled and integrated account of the cases in which previous research has failed to produce a VOE, which have combined to brand the VOE an unreliable phenomenon (Meissner & Brigham, 2001; Schooler, 2002).

### *Implications for the RBI Account*

The basic premise of RBI is that the content of verbalization influences the outcome of identification through retroactively interfering with the original memory trace. In the meta-analysis by Meissner and Brigham (2001), the applicability of RBI was reflected in the fact that the size of the VOE increases considerably after elaborative descriptions, that is, those description instructions

that force people to provide a description despite being unsure even if this will result in them having to guess.

Indeed, we offered evidence at the outset that elaborative descriptions may give rise to a different type of verbal overshadowing phenomenon than that observed with standard instructions. Although the evidence is mixed, with a number of studies failing to find a relationship between description accuracy and identification performance (e.g., Brown & Lloyd-Jones, 2002; Kitagami et al., 2002; Fallshore & Schooler, 1995; Finger, 2002; Schooler & Engstler-Schooler, 1990), when this correlation is present, it can typically be linked to an increase in the number of description errors (e.g., Finger & Pezdek, 1999; Meissner, 2002; Meissner et al., 2001). There is thus little doubt that retrieval-based interference occurs, albeit primarily after elaborative rather than standard descriptions. It turns out that the occurrence of RBI with elaborative descriptions is relevant to one of the limitations associated with the criterion view.

### *Limitations of the Criterion Account and Outlook*

Although the results reported here clearly support the criterion-shift account, at least two limitations deserve consideration. First, the criterion explanation insists that verbal overshadowing should not occur with forced-choice lineups. Although this expectation was confirmed in our Experiments 2 and 3, other researchers have obtained a VOE with forced-choice lineups (e.g., Fallshore & Schooler, 1995; Ryan & Schooler, 1998). At first glance, these reports seem to present a major challenge to the criterion view. However, closer inspection reveals that in most cases in which a VOE occurred with forced-choice lineups, participants were given elaborative description instructions. The selective effect of elaborative descriptions amplifies the preceding argument, that there may be two qualitatively different manifestations of verbal overshadowing, each tied to a specific type of description. With standard instructions, the data are best explained by a criterion shift and verbal overshadowing is limited to optional-choice lineups. With elaborative instructions, the likely errors in the description cause interference with existing memories, and verbal overshadowing may generalize to forced-choice lineups.

Further investigations are underway to extend this conclusion. Clare and Lewandowsky (2003) conducted a forced-choice experiment that compared elaborative and standard descriptions and found a VOE for the former, but not the latter, type of description. Clare and Lewandowsky additionally reevaluated the existing literature in another meta-analysis and found that the available data support the criterion view with standard description, but require the presence of another interference-based process, with elaborative descriptions.

Turning to the second limitation of the criterion view, our research to date does not explain why the recognition criterion is raised following verbalization. We examined two potential variables, identification confidence and perceived difficulty of verbalization, and neither was found to be related to identification accuracy (or by implication, to the criterion placement). However, both measures were taken after the identification had been attempted and thus may not be sensitive to transient shifts in criterion that arose as a consequence of verbalization. Thus, although this initial investigation cannot pinpoint why the criterion shifts after standard verbalization, the findings of Winkielman and Schwarz (2001) discussed at the outset support our suggestion that the subjective difficulty of the description combined with its presumed inadequacy induce reluctance during the subsequent identification task. Future research needs to develop measures that examine this presumed relationship between subjective experience and criterion setting, perhaps explicitly manipulating the recognition criterion prior to identification.

### *Applied Implications*

Perhaps the most important applied implication of our studies is that verbalization can have a positive effect on identification. Specifically, although verbalization resulted in fewer correct identifications from PP lineups, it also reduced the number of false identifications with PA lineups. Because in the real world it is always uncertain if the perpetrator is actually contained within a lineup, verbalization may thus protect innocent suspects from being falsely identified. This conclusion differs considerably from the existing broad consensus that any effects of verbalization on eyewitness identification are always deleterious.

## Conclusions

This article presented evidence that with standard description instructions, verbal overshadowing occurs because people become more reluctant to identify someone from a lineup after they provide a description of the perpetrator. Unlike existing theories, the criterion explanation is the only one that can simultaneously account for (a) a large verbal overshadowing effect in optional-choice PP lineups, (b) the absence of a VOE with forced-choice identification, and (c) the beneficial effect of verbalization with optional-choice PA lineups.

## References

Brandimonte, M. A., Schooler, J. W., & Gabbino, P. (1997). Attenuating verbal overshadowing through color retrieval cues. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 915–931.

Brigham, J. C., Meissner, C. A., & Wasserman, A. W. (1999). Applied issues in the construction and expert assessment of photo lineups. *Applied Cognitive Psychology, 13*, S73–S92.

Brigham, J. C., & Pressley, M. (1988). Cognitive monitoring and strategy choice in younger and older adults. *Psychology and Aging, 3*, 249–257.

Brown, C., & Lloyd-Jones, T. J. (2002). Verbal overshadowing in a multiple face presentation paradigm: Effects of description instruction. *Applied Cognitive Psychology, 16*, 873–885.

Clare, J., & Lewandowsky, S. (2003). *Towards a two-factor theory of verbal overshadowing: The integration of elaborative and standard verbalization.* Unpublished manuscript, University of Western Australia. Perth, Western Australia.

Clark, S. E. (2003). A memory and decision model for eyewitness identification. *Applied Cognitive Psychology, 17*, 629–654.

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review, 3*, 37–60.

Dodson, C. S., Johnson, M. K., & Schooler, J. W. (1997). The verbal overshadowing effect: Why descriptions impair face recognition. *Memory and Cognition, 25*, 129–139.

Fallshore, M., & Schooler, J. W. (1995). The verbal vulnerability of perceptual expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1608–1623.

Farrell, S., & Lewandowsky, S. (2002). An endogenous distributed model of ordering in serial recall. *Psychonomic Bulletin & Review, 9*, 59–79.

Finger, K. (2002). Mazes and music: Using perceptual processing to release verbal overshadowing. *Applied Cognitive Psychology, 16*, 887–896.

Finger, K., & Pezdek, K. (1999). The effect of the cognitive interview on face identification accuracy: Release from verbal overshadowing. *Journal of Applied Psychology, 84*, 340–348.

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

Gwyer, P., & Clifford, B. R. (1997). The effects of the cognitive interview on recall, identification, confidence and the confidence/accuracy relationship. *Applied Cognitive Psychology, 11*, 121–145.

Hintzman, D. L. (1991). Why are formal models useful in psychology. In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (pp. 39–56). Hillsdale, NJ: Erlbaum.

Kebbell, M. R., Milne, R., & Wagstaff, G. F. (1999). The cognitive interview: A survey of its forensic effectiveness. *Psychology, Crime, & Law, 5*, 101–115.

Kitagami, S., Sato, W., & Yoshikawa, S. (2002). The influence of test–set similarity in verbal overshadowing. *Applied Cognitive Psychology, 16*, 963–972.

Köehnken, G., Malpass, R. S., & Wogalter, M. S. (1996). Forensic applications of line-up research. In S. L. Sporer, R. S. Malpass, & G. Köehnken (Eds.), *Psychological issues in eyewitness identification* (pp. 205–231). Mahwah, NJ: Erlbaum.

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.

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

Koriat, A., Goldsmith, M., & Pansky, A. (2000). Toward a psychology of memory accuracy. *Annual Review of Psychology, 51*, 481–537.

Krafka, C., & Penrod, S. (1985). Reinstatement of context in a field experiment on eyewitness identification. *Journal of Personality and Social Psychology, 49*, 58–69.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99*, 22–44.

Lewandowsky, S. (1993). The rewards and hazards of computer simulations. *Psychological Science, 4*, 236–243.

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.

Lüüs, C. A. E., & Wells, G. L. (1991). Eyewitness identification and the selection of distracters for lineups. *Law and Human Behavior, 15*, 43–57.

Macrae, C. N., & Lewis, H. L. (2002). Do I know you? Processing orientation and face recognition. *Psychological Science, 13*, 194–196.

Meissner, C. A. (2002). Applied aspects of the instructional bias effect in verbal overshadowing. *Applied Cognitive Psychology, 16*, 911–928.

Meissner, C. A., & Brigham, J. C. (2001). A meta-analysis of the verbal overshadowing effect in face identification. *Applied Cognitive Psychology, 15*, 603–616.

Meissner, C. A., Brigham, J. C., & Kelley, C. M. (2001). The influence of retrieval processes in verbal overshadowing. *Memory and Cognition, 29*, 176–186.

Melcher, J. M., & Schooler, J. W. (1996). The misrememberance of wines past: Verbal and perceptual expertise differentially mediate verbal overshadowing of taste memory. *Journal of Memory and Language, 35*, 231–245.

Roediger, H. L. (1996). Memory illusions. *Journal of Memory and Language, 35*, 76–100.

Roediger, H. L., Jacoby, J. D., & McDermott, K. B. (1996). Misinformation effects in recall: Creating false memories through repeated retrieval. *Journal of Memory and Language, 35*, 300–318.

Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis*. New York: McGraw-Hill.

Ryan, R. S., & Schooler, J. W. (1998). Whom do words hurt? Individual differences in susceptibility to verbal overshadowing. *Applied Cognitive Psychology, 12*, S105–S125.

Schooler, J. W. (2002). Verbalization produces a transfer inappropriate processing shift. *Applied Cognitive Psychology, 16*, 989–997.

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology, 22*, 36–71.

Schooler, J. W., Fiore, S. M., & Brandimonte, M. A. (1997). At a loss *from* words: Verbal overshadowing of perceptual memories. *The Psychology of Learning and Motivation, 37*, 291–340.

Schooler, J. W., Ohlsson, S., & Brooks, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of Experimental Psychology: General, 122*, 166–183.

Schooler, J. W., Ryan, R. S., & Reder, L. (1996). The costs and benefits of verbally rehearsing memory for faces. In D. Herrmann, M. Johnson, C. McEvoy, C. Hertzog, & P. Hertel (Eds.), *Basic and applied memory: New findings* (pp. 51–65). Hillsdale, NJ: Erlbaum.

Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional Science, 26*, 113–125.

Wells, G. L., & Hryciw, B. (1984). Memory for faces: Encoding and retrieval operations. *Memory and Cognition, 12*, 338–344.

Wells, G. L., & Turtle, J. W. (1986). Eyewitness identification: The importance of lineup models. *Psychological Bulletin, 99*, 320–329.

Winkielman, P., & Schwarz, N. (2001). How pleasant was your childhood? Beliefs about memory shape inferences from experienced difficulty of recall. *Psychological Science, 12*, 176–178.

Yu, C. J., & Geiselman, R. E. (1993). Effects of constructing Identi-kit composites on photospread identification performance. *Criminal Justice and Behavior, 20*, 280–292.

Received May 15, 2003
Revision received December 11, 2003
Accepted December 22, 2003 ■