

A Meta-analysis of the Verbal Overshadowing Effect in Face Identification

CHRISTIAN A. MEISSNER* and JOHN C. BRIGHAM

Florida State University, USA

SUMMARY

Recent studies have demonstrated that requesting individuals to produce a verbal description of a previously seen face can hinder subsequent attempts at identification. This phenomenon, termed 'verbal overshadowing', has been studied rather extensively in the face-identification paradigm; however, studies have not always replicated the general effect. Based upon both practical and theoretical interests in the phenomenon, a meta-analysis of 29 effect size comparisons ($N = 2018$) was conducted. Across the sample of studies there was a small, yet significant, negative effect (Fisher's $Z_r = -0.12$), indicating some degree of verbal impairment or overshadowing. A fixed-effects analysis of several moderating variables demonstrated a significant effect of post-description delay and type of description instruction. The pattern of means indicated that overshadowing effects were more likely to occur when the identification task immediately followed the description task, and when participants were given an elaborative, as opposed to a standard (free recall), instruction during the description task. Inconsistencies in the literature are discussed, as well as various theoretical and applied issues regarding the verbal overshadowing effect. Copyright © 2001 John Wiley & Sons, Ltd.

Recent research on the susceptibility of eyewitness memory suggests that individuals may be influenced by a variety of factors involving encoding and retrieval of event-related information (for reviews, see Ross *et al.*, 1994; Sporer *et al.*, 1996). While this line of research has had some effect on the legal community's notion of eyewitness memory for objects and events at the crime scene (Wells, 1993), until recently it was thought that the procedure of obtaining a verbal description of the suspect had relatively minor consequences for the memory of the witness. For centuries, law enforcement officials have operated on this premise in their daily protocol of investigative procedures. Furthermore, several studies have previously demonstrated the positive effects of verbal description, rehearsal, and elaboration on later recognition of faces (Chance and Goldstein, 1976; McKelvie, 1976; Read, 1979; Wogalter, 1991, 1996).

However, a series of six experiments by Schooler and Engstler-Schooler (1990) demonstrated apparent deleterious effects of verbal description on subsequent identification of a target face, an effect they termed *verbal overshadowing*. In these experiments, participants were initially presented with a 30 s video presentation of a bank robbery, including a salient perpetrator, after which they were assigned to either a *description condition* or a *no-description control condition*. Those in the description condition were instructed to use the next five minutes to describe the facial features of the robber, while

*Correspondence to: Christian A. Meissner, Department of Psychology, Florida State University, Tallahassee, FL 32306-1270, USA

Received 5 October 1999

Accepted 25 April 2000

control participants were given an innocuous filler task. Identification was assessed for each condition using an 8-photo target-present lineup. Schooler and Engstler-Schooler's results consistently demonstrated that when participants were asked to provide a post-exposure description of the target, they were significantly less able to identify him or her when compared with a control group.

Since their initial set of experiments, a number of studies have further replicated the overshadowing phenomenon within the facial memory paradigm (Dodson *et al.*, 1997; Fallshore and Schooler, 1995; Finger and Pezdek, 1999; Read and Schooler, 1994, presentation at the Third Practical Aspects of Memory Conference, College Park, MD; Ryan, 1992, unpublished thesis; Ryan and Schooler, 1998; Schooler *et al.*, 1996). However, several attempts at replicating the effect in other labs have proven unsuccessful (Lindsay, 1990, personal communication cited in Schooler *et al.*, 1996; Lovett *et al.*, 1992, unpublished manuscript; Meissner *et al.*, 2001; Memon *et al.*, 1999, unpublished manuscript; Tunnicliff and Clark, 1999, unpublished raw data; Yu and Geiselman, 1993). In addition, several shortcomings in the effect have been demonstrated. For example, it has been shown that the effect may be *attenuated* on subsequent trials within-subject (Fallshore and Schooler, 1995; Houser *et al.*, 1997; Melcher and Schooler, 1995; Schooler *et al.*, 1996), or following re-presentation of the original stimulus prior to identification (Schooler *et al.*, 1996), or when a significant delay between the description and identification phases exists (Finger and Pezdek, 1999).

Regarding this latter post-description delay effect, Finger and Pezdek (1999) have recently demonstrated that when participants were provided a 24-minute delay prior to identification of the target face, a 'release of verbal overshadowing' occurred. In particular, the delayed description condition outperformed not only the no-delay description condition, but also the no-delay control condition. This increase in accuracy was said to result from a release of interference created by the description task. Although some researchers have observed similar effects (cf. Yu and Geiselman, 1993), others have been unable to replicate these results (Boelter and Reisberg, 1999, presentation at the Society for Applied Research in Memory and Cognition Conference, Boulder, CO; Meissner *et al.*, 2001; Schooler and Engstler-Schooler, 1990). We will attempt to further examine this potential moderating effect in the present meta-analysis.

As a result of various empirical inconsistencies in the literature, theoretical accounts of the overshadowing phenomenon have often been difficult to formulate. In a recent review, Schooler *et al.* (1997) concluded that the majority of research appeared to provide support for both the *modality mismatch* hypothesis (i.e. the notion of competing representations in memory: verbal versus non-verbal information) and the *availability* assumption (i.e. the notion that the visual representation remains available in memory despite the temporary verbal impairment). However, Schooler and colleagues noted little empirical support for a *recoding interference* hypothesis (i.e. the belief that overshadowing effects are due to a non-veridical verbal description of the target stimulus), as studies within the facial memory domain have generally failed to find a relationship between quality of description and subsequent identification accuracy (Fallshore and Schooler, 1995; Schooler and Engstler-Schooler, 1990).

In contrast to this conclusion, though, several studies have found evidence of retrieval-based factors influencing whether or not the overshadowing effect occurs (Finger and Pezdek, 1999; Meissner *et al.*, 2001). For example, in a recent study, we attempted to test the possibility that impairment on the identification task depended upon the manner in which individuals were initially instructed to describe the target from memory. This

hypothesis is rather consistent with the findings of Finger and Pezdek (1999), who also observed that variation in the type of description task (i.e. standard interview versus cognitive interview) influenced the likelihood of a verbal overshadowing effect. In testing this retrieval-based effect, we found that inducing participants to generate many descriptors by lowering their response criterion resulted in impairment or overshadowing on a photo lineup task both immediately and 30 minutes following the description task. However, conditions in which participants were encouraged to adopt a moderate or strong response criterion failed to yield subsequent impairment. Based upon this evidence, we concluded that failures to replicate the verbal overshadowing effect may have involved differences in the response criterion (set perhaps via variations in instructions or degree of exhortation used by experimenters) that participants used when attempting to describe the target stimulus. This effect of instruction type will also be assessed in the present meta-analysis.

Given the current degree of variability in results across studies, we felt that a statistical synthesis of the literature would be useful and appropriate in order to generalize current findings, clarify theoretical inconsistencies, and identify productive directions for future research. An additional practical concern is that verbal overshadowing could have an important influence on the manner in which law-enforcement officials obtain information from the eyewitness. If the impairment due to generating a description were found to be substantial across studies, it would seem important to inform law-enforcement officials of the potential harm in such a procedure.

The current meta-analysis took the approach advocated by Hedges and Olkin (1985) in which a mean weighted effect size for the sample of studies was initially calculated, followed by prediction of effect size based upon several moderating variables (see Johnson *et al.*, 1995, for a discussion of various approaches). We had three goals in synthesizing the verbal overshadowing literature. First, we sought to estimate the true effect size for the verbal overshadowing effect with regard to facial stimuli, including as many published and unpublished studies as could be obtained. Second, we wanted to examine whether several moderating variables inherent to the design of most verbal overshadowing studies might predict when the effect would more likely occur. Finally, we examined several sets of studies which failed to meet the methodological criteria for inclusion in the overshadowing paradigm, but which involved post-encoding verbalization or recall of facial details. Our interest here was to investigate whether the overshadowing effect might be found in other paradigms (i.e. face-recognition paradigm), in other identification tasks (i.e. mug shot sorting task), and in other forensically relevant recall tasks (i.e. facial composite reconstruction).

METHOD

Studies

For the primary analysis of verbal overshadowing on face identification, a total of 29 effect size comparisons described in 15 research articles were located, representing the responses of 2018 participants. Studies were obtained by several methods, including: (a) searches of the *PsycINFO*, *Sociofile*, *Dissertation Abstracts*, and *First Search* databases using the key words 'verbal overshadowing', 'face identification', 'face description', 'eyewitness memory', and 'facial memory'; (b) a search of the *Social Sciences Citation Index* using the previous key words and the Schooler and Engstler-Schooler (1990) principal citation;

(c) cross-referencing of the bibliographies of review papers on the topic (Schooler *et al.*, 1996, 1997); (d) a search of several selected conference programmes (i.e. Psychonomic Society, APA, APS, and SARMAC) over the past 5 years; and (e) contact with colleagues in the field who may have had knowledge of fugitive literature which had neither been published nor presented at a conference. Of the 29 effect size comparisons located, 11 (38%) appeared in unpublished manuscripts or theses/dissertations.

Inclusion/exclusion criteria

To be included in the primary analysis of verbal overshadowing in face identification, studies must have involved a contrast between participants who were asked to both describe and identify a target individual (*description group*), and those asked only to identify the target individual (*no-description control group*). The statistical difference between these two groups is known as the verbal overshadowing effect. Second, given the potential influence of confusion of multiple target faces on later verbalization, and the evidence of attenuation of the effect when using multiple trials within-subject (see Schooler *et al.*, 1996), we included only studies that used a single stimulus face and identification task for each participant. Finally, participants must have been presented with a target-present photo lineup identification task, a predominant characteristic of studies in the overshadowing paradigm.¹ Studies which were excluded due to one or more criteria ($N=9$) are included in a separate analysis. Several other studies ($N=8$) which involved the use of an intervening facial composite recall task are also subsequently evaluated.

Coded variables²

Four moderator variables, described below, were coded for each study. Moderators were derived from aspects of the standard verbal overshadowing paradigm and included: (a) the manner of stimulus presentation (i.e. live/video versus photograph); (b) the presence of a post-encoding delay; (c) the presence of a post-description delay; and (d) the type of description instruction given to participants (i.e. standard-free recall versus elaborative-forced recall).³ While both stimulus type and description instruction were represented as dichotomous variables, post-encoding and post-description delay variables were log

¹Two experiments by Fallshore and Schooler (1995) utilized a paradigm in which participants were exposed to a series of trials in which they were asked to view, describe, and identify a given target face. Because attenuation of the effect occurred on later trials, only data from the first trial were used in the present analysis. Other studies were excluded due to various reasons. In particular, Chance and Goldstein (1976, Experiment 3) reported in Dodson *et al.* (1997), McClure (1998, unpublished dissertation), and Woglater (1991, 1996) were excluded from the overall analysis due to the use of multiple target faces at presentation. Mauldin and Laughery (1981) and Thompson (1979, unpublished dissertation) were excluded due to the use of an alternative identification procedure (mug shot sorting task). Experiment 1 reported in Westerman and Larsen (1997) was also excluded from the analysis due to the lack of a no-description control condition resulting from the within-subjects design of the study.

²Two independent raters coded each of the studies, after which percent agreement and Kappa (κ) values were calculated for each moderator variable. All Kappa's were in the excellent range (0.75 +), as suggested by Fleiss (1981). A third rater was used to resolve any discrepancies in the codings.

³A fourth moderator variable that might seem appropriate involved whether the target and test stimuli were identical ('photo recognition'), or whether they varied in pose and context ('face recognition'). However, only two early studies (Chance and Goldstein, 1976; Mauldin and Laughery, 1981) used identical stimuli at encoding and recognition. Consequently, this moderator was not included in the analysis.

transformed and used in their continuous form, as performed when simulating decay/forgetting in many cognitive models of memory (see Rubin and Wenzel, 1996).

First, the *manner of stimulus presentation* was recorded as either involving a short video presentation of a mock witness event or the display of a single photograph of the target face. A previous meta-analysis on facial identification by Shapiro and Penrod (1986) also examined the role of stimulus presentation. They hypothesized that performance would be best when more retrieval cues were available at encoding; thus, a video presentation should result in a higher rate of correct identifications when compared with a photograph of the target. Their review, however, showed only mixed support for this hypothesis. Within the verbal overshadowing paradigm, we might expect a similar process to be working, such that variations in perspective and context might provide more cues for verbalization, and thus increase identification accuracy when compared with a stimulus photograph.

Second, the presence of a *post-encoding delay* was coded. A majority of studies have used a task to distract participants from rehearsing the target face, with the delay between encoding and the description task varying between 5 and 20 minutes. However, a few studies have immediately requested participants to provide a description of the target face. If the mechanism of verbal overshadowing involves the veridicality of the description, then one might predict that the presence of a delay prior to verbalization would decrease the accuracy of the description due to a degradation of the original stimulus face in memory. Consequently, the likelihood of impairment on later identification could also increase.

Third, the presence of a *post-description delay* was coded. While most studies have assessed participants' memory immediately after the description, a few researchers have examined the practical implications of inserting a delay between the description and identification tasks. Such post-description delays have ranged from as little as 3 minutes to as long as 2 days. As noted previously, Finger and Pezdek (1999) demonstrated that participants who described the target and then experienced a 24-minute delay prior to identification outperformed participants in the no-delay control condition. This effect was assessed across the current sample of studies.

Finally, studies were coded for the *type of instruction* that participants were presented during the description task. Given the paucity of information provided in most manuscripts regarding the description instructions, authors were contacted and asked whether they used any methods that may have 'encouraged participants to verbalize for the entire description period'. We were particularly interested in whether the authors explicitly encouraged their participants to go beyond their normal criterion of free recall and to provide more elaborative descriptions. We contrasted this with standard (free recall) instructions in which participants were allowed to establish their own recall criterion. Consistent with our previous findings (Meissner *et al.*, 2001), we hypothesized that studies using an elaborative instruction would be more likely to demonstrate overshadowing effects when compared with studies using a standard instruction.

Measure of effect size

Fisher's Z_r was used as the measure of effect size across the studies. In most cases, a single degree of freedom (*df*) analysis (i.e. χ^2) between the description and control conditions was transformed into the effect size estimate. When studies presented only the proportion of correct identifications (hit rates), a χ^2 statistic was computed using the number of

correct versus incorrect decisions across the control and description conditions. χ^2 values were then converted to r (correlation coefficient) by way of

$$r = \text{SQRT}(\chi^2/N) \quad (1)$$

Finally, to control for skewness in estimating the true population parameter, r was transformed to Fisher's Z_r by way of

$$Z_r = 0.5 * \log_e[(1 + r)/(1 - r)] \quad (2)$$

Both the r and Z_r transformation formulae were obtained from Rosenthal (1994). Because verbalization is thought to impair subsequent identification (Description–Control < 0), effect sizes demonstrating the overshadowing effect would be in the *negative* direction. Conversely, a *positive* effect size would demonstrate some degree of verbal facilitation.

RESULTS

Effect size analysis

Effect sizes (Z_r) computed for each study are presented in Table 1, in addition to sample size, and coded variables. The mean weighted effect size across studies was $Z_r = -0.12$, a significant effect size, $Z = -5.27$, $p < 0.001$, with 95% confidence intervals of $(-0.17, -0.08)$. No significant outliers were noted in the sample of studies. Overall, the verbal overshadowing effect accounted for 1.44% of the variability across studies, and indicated that participants who described the target were 1.27 times more likely to *misidentify* the target when compared with participants in the no-description control condition.

Moderator variables

A test of the homogeneity of variances across the sample of weighted effect sizes indicated a significant degree of variability, exceeding that expected on the basis of sampling error alone, $\chi^2(28) = 59.42$, $p < 0.01$. Thus, the design moderators discussed earlier were used to predict the variability across the sample of effect sizes. Due to the small sample of studies, we utilized a fixed-effects approach in which a weighted least squares regression analysis was run on the sample of effect sizes with the four design moderators (stimulus, post-encoding delay, post-description delay, and description instruction) as predictors (see Hedges, 1994).⁴ Studies were weighted in the analysis as a function of their sample sizes. Results indicated that both the post-description delay, $Z_j = 2.00$, $p < 0.05$, $r_s = 0.21$, and description instruction variables, $Z_j = 2.83$, $p < 0.01$, $r_s = 0.40$, significantly influenced the magnitude of observed effect sizes. However, neither the manner of stimulus presentation, $Z_j = 1.69$, *ns*, nor the presence of a post-encoding delay, $Z_j = 1.19$, *ns*, were significant predictors. The pattern of weighted means for the post-description delay variable indicated that verbal overshadowing effects were seen in studies using either an 'immediate'

⁴A random effects model would have been most appropriate given other considerations; however, it is not as sensitive an approach and would have provided only inconclusive results given the small number of studies (see Raudenbush, 1994). In fact, results of a random effects analysis indicated only marginal effects for both post-description delay, $t = 1.73$, $p < 0.10$, and type of description instruction, $t = 1.78$, $p < 0.10$.

Table 1. Sample size (N) effect size (Z_r), and moderator variables across studies

Study	Exp./condition	N	Fisher's Z_r	Stimulus	Post-encoding delay	Post-description delay	Description instruction
Boelter and Reisberg (1999 unpublished raw data)		36	-0.21	Video	48 h	Immediate	Elaborative
		36	-0.17	Video	Immediate	48 h	Elaborative
Dodson <i>et al.</i> (1997)	Exp. #1	40	-0.17	Video	20 min	Immediate	Standard
	Exp. #2	80	-0.13	Photo	5 min	Immediate	Standard
Fallshore and Schooler (1995)	Exp. #1	120	-0.34	Photo	5 min	Immediate	Elaborative
	Exp. #3	160	-0.05	Photo	Immediate	Immediate	Elaborative
Finger and Pezdek (1999)	Exp. #3	54	-0.38	Photo	5 min	Immediate	Elaborative
Lovett <i>et al.</i> (1992, unpublished manuscript)	Exp. #1	48	-0.05	Video	30 min	Immediate	Elaborative
	Exp. #2	48	0.14	Photo	30 min	Immediate	Elaborative
	Exp. #1	59	0.13	Photo	30 min	Immediate	Elaborative
Meissner <i>et al.</i> (2001)	Exp. #1	60	0.20	Photo	5 min	30 min	Standard
		60	0.00	Photo	5 min	Immediate	Standard
	Exp. #2	60	-0.34	Photo	5 min	Immediate	Elaborative
Memon <i>et al.</i> (1999, unpublished manuscript)	Exp. #3	39	-0.16	Video	20 min	Immediate	Elaborative
Miner and Reisberg (1999 unpublished raw data)	Exp. #1	30	-0.42	Video	15 min	3 min	Elaborative
	Exp. #2	24	0.09	Video	15 min	3 min	Elaborative
	Exp. #3	20	0.00	Video	15 min	3 min	Elaborative
Ryan (1992, unpublished thesis)	Incidental	229	-0.08	Video	2 min	Immediate	Elaborative
Ryan and Schooler (1998)		165	-0.11	Photo	2 min	Immediate	Elaborative
Schooler and Engstler-Schooler (1990)	Exp. #1	88	-0.27	Video	20 min	Immediate	Elaborative
	Exp. #2	70	-0.33	Video	20 min	Immediate	Elaborative
	Exp. #4	78	-0.22	Video	Immediate	10 min	Elaborative
	Exp. #5	67	-0.25	Video	Immediate	48 h	Elaborative
	Exp. #6	56	-0.31	Photo	5 min	Immediate	Elaborative
Schooler <i>et al.</i> (1996)	Exp. #1	60	-0.17	Photo	Immediate	Immediate	Elaborative
	Exp. #2	60	-0.07	Photo	Immediate	Immediate	Elaborative
Tunnichiff and Clark (1999, unpublished raw data)		62	0.15	Video	Immediate	1 week	Standard
Westerman and Larsen (1997)	Exp. #2	59	-0.11	Video	20 min	Immediate	Standard
Yu and Geiselman (1993)		50	0.33	Video	Immediate	48 h	Standard

assessment of identification accuracy ($Z_r = -0.16$), or a 'short' (≤ 10 minutes) delay between description and identification phases ($Z_r = -0.13$). However, in studies that employed a 'long' delay (≥ 30 minutes), differences between the description and control conditions were marginal ($Z_r = 0.07$), but in the direction of verbal facilitation. The pattern of weighted means for the description instruction variable indicated that studies using an elaborative instruction ($Z_r = -0.16$) were more likely to elicit verbal overshadowing effects when compared with studies using a standard instruction ($Z_r = 0.04$). Finally, goodness of fit of the regression model was significant, $\chi^2(24) = 42.43$, $p < 0.01$, indicating the presence of additional residual variance unexplained by the present moderators.

From the above results it is unclear what might be responsible for the differences in performance across the post-description delay variable. One possibility is that a 'release of verbal overshadowing' may occur such that performance in the description condition significantly increases across the delay (Finger and Pezdek, 1999). Alternatively, differences in performance could result from memory decay across the delay for the control condition. Finally, a combination of the above processes may be responsible for the interaction (i.e. increased accuracy for the description condition and decreased accuracy for the control condition across the delay). To examine more closely this effect, we analysed the pattern of identification accuracy for the control and description conditions across the three post-description delay conditions (immediate versus short delay versus long delay). Although variation in relative levels of identification accuracy may occur as a function of the particular stimuli used, we felt that such an analysis might nevertheless provide some preliminary estimate of the influence of a post-description delay. Unfortunately, a mixed factorial ANOVA on the mean proportion of correct identifications demonstrated a nonsignificant verbal overshadowing \times post-description delay interaction, $F(2, 26) = 2.34$, $p = 0.12$, $MSE = 0.013$, $\eta^2 = 0.15$. Nevertheless, planned comparisons confirmed that performance did significantly differ in the control condition when the identification task followed either immediately or after a short delay (≤ 10 minutes) versus a long delay (≥ 30 minutes), $t_s > 2.55$, $p_s < 0.05$. Conversely, participants who had previously generated a description of the target stimulus showed no significant performance differences across the three delays, $t_s < 0.30$, *ns*. Figure 1 displays the pattern of means for the interaction.

Publication bias

As there has been some difficulty in replicating the verbal overshadowing effect within the research community, it is possible that, despite our efforts to accumulate both published and unpublished studies, some degree of publication bias exists in the sample. As a measure of publication bias, we calculated the number of null studies necessary for our effect to no longer be significant. Also known as the 'file drawer calculation' (see Begg, 1994) it was estimated that 155 studies with an effect size of zero would be necessary for the mean weighted effect size to no longer be statistically significant. Thus, it appears that we can maintain a good degree of confidence in the results of the meta-analysis indicating a verbal overshadowing effect.⁵

⁵In addition, a sensitivity analysis was conducted such that we could determine whether one or more studies exerted undue influence over the pattern of results (see Greenhouse and Iyengar, 1994). No study appeared to contribute a large degree of influence, as average effect sizes produced by the analysis fell within an acceptable range.

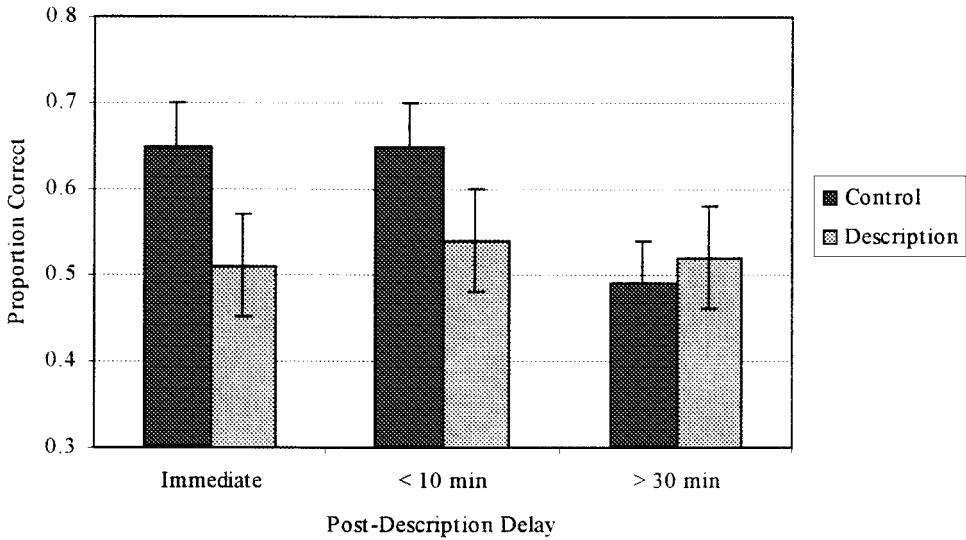


Figure 1. Proportion correct identification as a function of post-description delay

Excluded studies

Several other studies were obtained from the literature search that were generally consistent with the constraints of the verbal overshadowing paradigm, but were excluded from the meta-analysis due to their use of multiple target faces with each participant (Chance and Goldstein, 1976; McClure, 1998, unpublished dissertation; Wogalter, 1991, 1996), or their use of an alternative identification procedure (Mauldin and Laughery, 1981; Thompson, 1979, unpublished dissertation) (see Table 2).⁶ The mean weighted effect size across these studies was $Z_r = 0.15$, indicating a significant improvement in identification accuracy following generation of a verbal description (verbal facilitation), $Z = 3.28$,

Table 2. Sample size (N) and effect size (Z_r) estimates for excluded studies

Study	Exp./condition	N	Fisher's Z_r
Chance and Goldstein (1976)		40	0.40
Mauldin and Laughery (1981)	4 s	26	0.33
	15 s	22	0.50
McClure (1998, unpublished dissertation)	Exp. #1	134	0.13
	Exp. #2; 3-min delay	88	0.12
	Exp. #2; immediate	88	0.02
Thompson (1979, unpublished dissertation)		50	0.00
Wogalter (1991)		43	0.12
Wogalter (1996)		48	0.18

⁶Experiment 3 of Dodson *et al.* (1997) and Experiment 1 of Westerman and Larsen (1997) were not included in this set. Exclusion of the first was due to the use of a mixed design in the description condition (i.e. participants described only one of two faces but were later tested on identification for both). The latter was excluded due to the lack of a no-description control condition resulting from the within-subjects design of the study.

Table 3. Sample size (N) and effect size (Z_r) estimates for studies involving a facial composite recall composite recall task

Study	Exp./condition	N	Fisher' Z_r
Davies <i>et al.</i> (1978)	Exp. #2: < 48 h	20	-0.14
	Exp. #2: 3 weeks	20	-0.20
Mauldin and Laughery (1981)	4 s: immediate	20	0.79
	4 s: 48 h	20	0.17
	15 s: immediate	20	0.62
	15 s: 48 h	20	0.37
Thompson (1979, unpublished dissertation)		50	0.26
Yu and Geiselman (1993)		47	0.03

$p < 0.001$, with 95% confidence intervals of (0.06, 0.23). In sum, the current set of studies indicated that individuals who generated a description were 1.38 times more likely to *correctly identify* the target(s) when compared with individuals in the no-description control condition.

Facial composite recall task

Several studies ($N = 8$) have examined the influence of a facial composite recall task (i.e. Photofit or Identi-Kit) on later identification of a target face. It is possible that such a task could act to overshadow or interfere with an individual's ability to make a subsequent identification. However, because the task maintains a visual component as well, one might predict that the processes acting on visual recall and visual recognition would compliment, rather than contradict, one another (see Schooler *et al.*, 1997). To examine this effect, studies that followed the basic overshadowing methodology, but which involved a facial composite reconstruction rather than a verbal description task, were collected (see Table 3). Across studies, the mean weighted effect size was $Z_r = 0.21$, indicating a significant improvement in recognition accuracy following facial composite reconstruction, $Z = 2.92$, $p < 0.01$, with 95% confidence intervals of (0.07, 0.35). Overall, this effect accounted for 4.41% of the variability across studies, and indicated that participants who generated a facial composite were 1.56 times more likely to *correctly identify* the target when compared with a no-description control condition.

DISCUSSION

The results of our meta-analysis indicated a small, yet significant, verbal overshadowing effect. This overall effect accounted for only 1.44% of the variability across studies, a finding that may explain other researchers' failure in replicating the general overshadowing effect. Interestingly, a sample of studies which were excluded from the primary analysis, but which involved a comparison between description and control conditions, exhibited a significant verbal *facilitation* effect. Given that the overshadowing effect is known to attenuate across repeated trials within-subject (see Schooler *et al.*, 1996), this effect was somewhat expected. However, less is known regarding the overshadowing effect using alternative identification procedures, such as a mug shot sorting task (Mauldin

and Laughery, 1981; Thompson, 1979, unpublished dissertation). Further research assessing the reliability of these results would be beneficial.

Due to the variance in effect size estimates across studies, a fixed-effects analysis (weighted least-squares regression) was run in which four moderators were entered as predictors. Both post-description delay and type of description instruction were significant. Regarding the post-description delay effect, examination of the mean weighted effect sizes indicated that verbal overshadowing occurred when identification followed the description task either immediately or shortly thereafter (≤ 10 minutes). However, when a long delay (≥ 30 minutes) was inserted prior to the identification phase, no significant differences were noted between the control and description conditions.

A more in-depth analysis of the proportion correct across studies indicated that the long delay (≥ 30 minutes) influenced only participants in the no-description control condition who demonstrated a significant degree of *forgetting*. Conversely, participants who had previously described the target face showed no change in performance across the delay. Overall, it appears that the process of retrieving the stimulus face from memory during the description task allowed participants to maintain or preserve the memory trace across the extended post-description delay when compared with the control condition. Curiously, such a finding is inconsistent with that of a recent study by Finger and Pezdek (1999) in which the description condition demonstrated significant improvement in identification performance across a 24-minute delay. Unfortunately, the authors included only an immediate control condition; thus, we were unable to determine whether significant forgetting might also occur for participants not asked to describe the target face. The general effect of post-description delay was also inconsistent with an experiment conducted by Schooler and Engstler-Schooler (1990; Experiment 5: 2-day delay), and a study by Boelter and Reisberg (1999, presentation at the Society for Applied Research in Memory and Cognition, Boulder, CO; 2-day delay), in which significant verbal overshadowing effects were observed.

The fixed-effects analysis also indicated a significant influence of description instruction on the magnitude of observed effects, a result consistent with our previous studies (Meissner *et al.*, 2001). Specifically, studies that employed instructions intended to elicit an elaborative and detailed description of the target face were more likely to demonstrate verbal overshadowing when compared with studies that employed only a standard (free recall) instruction. First, we believe this effect to have certain implications for the methods that law-enforcement officials might use to elicit descriptions from eyewitnesses. Namely, it appears that officers should allow witnesses to establish their own criterion of responding when providing a description of the suspect, including the freedom to withhold information of which they may be unsure. Although such a procedure involves a cost of less information in the description, law enforcement must weigh this against the potential for later misidentification of the suspect.

Second, with regard to the theoretical mechanism of description instructions, our empirical results (Meissner *et al.*, 2001) have indicated that this effect is likely due to the influence of erroneous descriptors that are more often elicited under elaborative recall. This erroneous coding of the stimulus appears to confuse participants when they later attempt to distinguish that which was internally generated as correct versus incorrect, and thus to match their memory for the target to members in the photo lineup. This retrieval-based account is consistent with Schooler and Engstler-Schooler's (1990) original recoding interference hypothesis, and is also in agreement with the results of Finger and Pezdek (1999).

Several alternatives to this account also seem plausible. First, Schooler and colleagues (1997) have recently proposed that the interference caused by verbalization may result from the type of memory processes demanded by the paradigm (i.e. configural versus featural processing of faces (see Diamond and Carey, 1986)), a theory they termed *transfer inappropriate retrieval* (TIR). In general, the authors suggested that an individual's description of a target face would contain only those aspects of the stimulus that were readily verbalizable (i.e. featural information), a process that would reduce access to (or overshadow) those aspects which were not recalled or those which could not be verbalized (i.e. configural information). As a result of the suppression of crucial visual information, individuals would later demonstrate difficulty in matching their (description-based) memory for the target with that of the identification stimulus presented them.

With regard to the description instruction effect, TIR might predict that the criterion shift produced by the elaborative instructions would simply lead to more (continuous) verbalization such that overemphasis is placed upon the featural (verbalizable) information. This overemphasis would then further overshadow crucial non-verbalizable (configural) details that appear to be important for identification. While Schooler and colleagues (1997) discuss a number of other findings in support of TIR, including evidence on cross-racial effects (Fallshore and Schooler, 1995), re-presentation of the target face prior to the identification task (Schooler *et al.*, 1996), and description of an alternative non-target face (Dodson *et al.*, 1997), one particular finding from the present meta-analysis is also noteworthy. Namely, participants who engaged in a facial composite recall task (i.e. Identi-Kit or Photofit) demonstrated significant improvement in identification performance over that of a no-recall control condition. As the processes activated by a facial composite task are primarily *visual*, presumably preserving participants' ability to rely on *configural* information, confirmation of increased performance appears to provide yet further evidence as to the viability of competing memory processes.⁷

In addition to the TIR approach that Schooler and colleagues (1997) have advocated, a second theoretical line has involved examining characteristics of the observer (participant) him/herself. For example, a study by Ryan and Schooler (1998) found that *individual differences* in participants' perceptual and verbal abilities appeared to influence the overshadowing effect in face identification (although see Memon *et al.*, 1999, unpublished manuscript). In addition, several studies have indicated that increased *perceptual skill* in a given domain (e.g. own-race faces, wine tasting, music, etc.) may also mediate overshadowing in tasks involving stimuli taken from the specified domain (Fallshore and Schooler, 1995; Houser *et al.*, 1997, unpublished manuscript; Melcher and Schooler, 1996; see also Schooler *et al.*, in press). With regard to the description instruction effect, individual differences may play a role in the recall criterion that participants naturally select on such tasks. For example, certain individuals may demonstrate an ability to more effectively monitor the responses that they generate from memory (see Koriat and Goldsmith, 1996). Future research would be beneficial in clarifying the potential for an individual differences account of verbal overshadowing, and for generalizing to overshadowing effects beyond the facial identification domain.

⁷In only two effect size estimates was this finding not supported (Davies *et al.*, 1978). Unlike other studies in this sample, Davies and colleagues' participants were asked to return to the lab either less than 48 hours or 3 weeks after initially viewing the target stimulus. However, given such a small sample of studies, it is difficult to conclude whether this moderating variable might solely be responsible for the observed decline in performance.

Overall, it is apparent from the present analysis that verbal overshadowing is a genuine and reliable phenomenon. However, various inconsistencies have plagued the experimental literature, and conclusions as to the mechanisms responsible for the phenomenon appear complex and elusive. Although we have made progress in assessing a handful of these inconsistencies, it is our hope that future empirical and theoretical work will clarify the various conditions under which the effect is reliably observed. Once this is accomplished, the applied issues associated with the verbal overshadowing effect may be further addressed, and a more grounded account of the findings can be presented to law-enforcement officials for their use in designing everyday investigative procedures.

ACKNOWLEDGEMENTS

We are to indebted Anders Ericsson, Colleen Kelly, Kathy Pezdek, Jonathan Schooler, and Rick Wagner for their comments on earlier versions of this manuscript.

REFERENCES

References denoted by an asterisk (*) were included in some portion of the meta-analysis.

- Begg CB. 1994. Publication bias. In *The Handbook of Research Synthesis*, Cooper H, Hedges LV (eds). Russell Sage Foundation: New York; 399–409.
- *Chance J, Goldstein AG. 1976. Recognition of faces and verbal labels. *Bulletin of the Psychonomic Society* **7**: 384–386.
- *Davies G, Ellis H, Shepherd J. 1978. Face identification: the influence of delay upon accuracy of photofit construction. *Journal of Police Science and Administration* **6**: 35–42.
- Diamond R, Carey S. 1986. Why faces are and are not special: an effect of expertise. *Journal of Experimental Psychology: General* **115**: 107–117.
- *Dodson CS, Johnson MK, Schooler JW. 1997. The verbal overshadowing effect: why descriptions impair face recognition. *Memory and Cognition* **25**: 129–139.
- *Fallshore M, Schooler, JW. 1995. The verbal vulnerability of perceptual expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **21**: 1608–1623.
- *Finger K, Pezdek K. 1999. The effect of verbal description on face identification accuracy: 'release from verbal overshadowing'. *Journal of Applied Psychology* **84**: 340–348.
- Fleiss JL. 1981. *Statistical Methods for Rates and Proportions* (2nd edn). Wiley: New York.
- Greenhouse JB, Iyengar S. 1994. Sensitivity analysis and diagnostics. In *The Handbook of Research Synthesis*, Cooper H, Hedges LV (eds). Russell Sage Foundation: New York; 383–398.
- Hedges LV. 1994. Fixed effects models. In *The Handbook of Research Synthesis*, Cooper H, Hedges LV (eds). Russell Sage Foundation: New York; 285–300.
- Hedges LV, Olkin I. 1985. *Statistical Methods for Meta-analysis*. Academic Press: Orlando, FL.
- Johnson BT, Mullen B, Salas E. 1995. Comparison of three major meta-analytic approaches. *Journal of Applied Psychology* **80**: 94–106.
- Koriat A, Goldsmith M. 1996. Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review* **103**: 490–517.
- *Mauldin MA, Laughery KR. 1981. Composite production effects on subsequent facial recognition. *Journal of Applied Psychology* **66**: 351–357.
- McKelvie SJ. 1976. The effects of verbal labeling on recognition memory for schematic faces. *Quarterly Journal of Experimental Psychology* **28**: 459–474.
- *Meissner CA, Brigham JC, Kelly CM. 2001. The influence of retrieval processes in verbal overshadowing. *Memory and Cognition* **29**: 176–186.
- Melcher JM, Schooler JW. 1996. The misremembrance of wines past: verbal and perceptual expertise differentially mediate verbal overshadowing of taste memory. *Journal of Memory and Language* **35**: 231–245.

- Raudenbush SW. 1994. Random effects models. In *The Handbook of Research Synthesis*, Cooper H, Hedges V (eds). Russell sage foundation: New York; 301–321.
- Read JD. 1979. Rehearsal and recognition of human faces. *American Journal of Psychology* **92**: 71–85.
- Rosenthal R. 1994. Parametric measures of effect size. In *The Handbook of Research Synthesis*, Cooper H, Hedges LV (eds). Russell Sage Foundation: New York; 383–398.
- Ross DF, Read JD, Toglia MP. 1994. *Adult Eyewitness Testimony: Current Trends and Developments*. Cambridge University Press: New York.
- Rubin DC, Wenzel AE. 1996. One hundred years of forgetting: a quantitative description of retention. *Psychological Review* **103**: 734–760.
- *Ryan RS, Schooler JW. 1998. Whom do words hurt?: Individual differences in susceptibility to verbal overshadowing. *Applied Cognitive Psychology* **12**: 105–126.
- *Schooler JW, Engstler-Schooler TY. 1990. Verbal overshadowing of visual memories: some things are better left unsaid. *Cognitive Psychology* **22**: 36–71.
- Schooler JW, Fiore SM, Brandimonte MA. 1997. At a loss from words: verbal overshadowing of perceptual memories. In *Handbook of Learning and Motivation* (Vol. 37), Medin D (ed.). Academic Press: Orlando, FL.
- Schooler JW, Ryan RS, Fallshore M, Melcher JM. in press. Knowing more than you can tell: The relationship between language and expertise. In *The Psychology of Expertise*, Nisbett RE, Caverni J (eds). Elsevier: Amsterdam.
- *Schooler JW, Ryan RS, Reder LM. 1996. The costs and benefits of verbalization. In *Basic and Applied Memory: New findings*, Herrmann D, Johnson M, McEvoy C, Hertzog C, Hertel P. (eds). Erlbaum: Hillsdale, NJ; 51–65.
- Shapiro PN, Penrod S. 1986. Meta-analysis of facial identification studies. *Psychological Bulletin* **100**: 139–156.
- Sporer SL, Malpass RS, Koehnken G. 1996. *Psychological Issues in Eyewitness Identification*. Erlbaum: Mahwah, NJ.
- Wells GL. 1993. What do we know about eyewitness identification? *American Psychologist* **48**: 553–571.
- *Westerman DL, Larsen JD. 1997. Verbal overshadowing effect: evidence for a general shift in processing. *American Journal of Psychology* **110**: 417–428.
- *Wogalter MS. 1991. Effects of post-exposure description and imaging on subsequent face recognition performance. *Proceedings of the Human Factors Society* **35**: 575–579.
- *Wogalter MS. 1996. Describing faces from memory: accuracy and effects on subsequent recognition performance. *Proceedings of the Human Factors and Ergonomics Society* **40**: 536–540.
- *Yu CJ, Geiselman RE. 1993. Effects of constructing identi-kit composites on photospread identification performance. *Criminal Justice and Behavior* **20**: 280–292.