EUROPEAN JOURNAL OF COGNITIVE PSYCHOLOGY
2008, 20 (3), 414–455

A theoretical review and meta-analysis of the description-identification relationship in memory for faces

Christian A. Meissner
University of Texas at El Paso, TX, USA

Siegfried L. Sporer
University of Giessen, Giessen, Germany

Kyle J. Susa
University of Texas at El Paso, TX, USA

Verbal descriptions can sometimes impair (or "overshadow") and other times facilitate subsequent attempts at perceptual identification of faces; however, understanding the relationship between these two tasks and the theoretical mechanisms that bridge this relationship has often proven difficult. Furthermore, studies that have attempted to assess the description-identification relationship have varied considerably in demonstrating significant and null results, often across a variety of paradigms and design parameters. In the present paper we review the relevant literatures and theoretical positions proposed to explain this relationship, and we present the first meta-analysis of this effect across 33 research papers and a total of 4278 participants. Our results suggest that there does appear to be a small, but significant, relationship between the description measures of accuracy, number of incorrect descriptors, and congruence with that of subsequent identification accuracy. Furthermore, certain conditions were found to strengthen the magnitude of this relationship, including the use of face recognition versus eyewitness identification paradigms and the length of delays between relevant tasks. We discuss both the theoretical and practical implications of this relationship for understanding memory for faces.

In 1972, the United States Supreme Court addressed the admissibility of eyewitness identification obtained under suggestive circumstances in *Neil v. Biggers*. Biggers, the defendant, was convicted of rape based primarily upon his identification by the victim who testified that she had "no doubt" that Biggers was the assailant. Shortly after the crime, the victim had provided

---

Correspondence should be addressed to Christian A. Meissner, Department of Psychology, University of Texas at El Paso, El Paso, TX 79968,USA. E-mail: cmeissner@utep.edu

The authors would like to thank Jasmine Koestler for her assistance with coding the data.

http://www.psypress.com/ecp DOI: 10.1080/09541440701728581

police with a description of the assailant and was administered several showups and lineups of individuals who matched her description. She was not able to identify anyone as the assailant. Seven months later, the police conducted another showup where the officers asked Biggers to say "shut up or I will kill you" to the victim. At this point the victim immediately identified Biggers as the assailant.

Lower Courts suppressed the victim's identification of the defendant, ruling that the identification process used by police was overly suggestive. Thereafter, the US Supreme Court was presented with the issue of determining whether or not the victim's identification of Biggers was reliable. In reversing the decisions of the lower Courts, the Court listed five factors that should be taken into account when evaluating the reliability of an identification, including: (1) the witness's opportunity to view the criminal during the crime; (2) the length of time between the crime and the subsequent identification; (3) the level of certainty demonstrated by the witness at the identification; (4) the witness's degree of attention directed toward the event/perpetrator; and (5) the (apparent) accuracy of the witness's prior description of the criminal. In the *Neil v. Biggers* case, the Court's emphasis appeared to shift from a concern with suggestivity to an overriding concern with the reliability of the identification, even if it was obtained under suggestive circumstances. At the time of this decision, little published scientific research on eyewitness memory existed. The Court could, therefore, make only "educated guesses" about the factors that might influence eyewitness accuracy; however, scientific research conducted over the past three decades has permitted a systematic evaluation of the validity of the five criteria enumerated by the Court.

In short, the Court's assumptions regarding the reliability of eyewitness testimony appear to have been overly simplistic. While research findings indicate that the first two of the five *Neil v. Biggers* factors are clearly related to accuracy in the way that the Court assumed (namely, the opportunity to view and the retention interval between the event and identification; for a review, see Tredoux, Meissner, Malpass, & Zimmerman, 2004; Wells & Olson, 2003), research on the remaining factors present a much more complex interpretation of the relationships assumed by the Court. For example, despite the fact that jurors rely quite heavily on the degree of confidence expressed by a witness (Brewer & Burke, 2002), meta-analyses conducted on the confidence-accuracy relationship suggest only a weak to medium correlation (Bothwell, Deffenbacher, & Brigham, 1987; Sporer, Penrod, Read, & Cutler, 1995), though this relationship is generally higher for choosers than for nonchoosers (Sporer et al., 1995), and has occasionally been found to be quite high under certain conditions (e.g., Lindsay, Read, & Sharma, 1998). We should emphasise, however, that the confidence-accuracy relationship mentioned here only holds for the original statement at the time of the identification, not for any later expression of confidence (e.g., in the courtroom) which is likely to be contaminated by feedback and other factors (see Wells & Olson, 2003). With regard to the attentiveness of the witness (the Court's fourth factor), research has found that witnesses who exude a moderate degree of attention to a situation are likely to be more accurate when compared to those who did not pay attention, or to those who were distracted because they were in a stressful crime situation (Deffenbacher, 1983). However, even if a witness were trying to be attentive, high fear or stress (induced, possibly, by the presence of a weapon) is likely to interfere with memory and impair the accuracy of subsequent identifications (Deffenbacher, Bornstein, Penrod, & McGorty, 2004).

The current paper provides the first empirical review of the US Supreme Court's fifth factor—namely, the relationship between the accuracy of the description provided by a witness and their subsequent identification of the perpetrator. Was the Court correct in assuming a relationship between verbal description and perceptual identification processes in memory for faces? Unfortunately, the available research presents a rather murky picture of this relationship, with correlational studies suggesting a wide variety of moderately strong to nonsignificant findings. Furthermore, studies investigating the influence of generating a verbal description on subsequent identification have shown that the former process can *impair* later attempts at identification (Schooler & Engstler-Schooler, 1990), while in contrast a handful of studies have shown that the act of describing can *enhance* later identification (Brown & Lloyd-Jones, 2005, 2006; Davids, Sporer, & McQuiston-Surrett, 2006; Meissner, Brigham, & Kelley, 2001).

In this paper we examine the variability of the description-identification relationship across studies in several ways. First, we provide a theoretical literature review of the relationship, including studies on the verbal overshadowing and verbal facilitation effects in face identification. Second, we present the first meta-analysis of studies that have estimated the relationship between description performance and identification accuracy on memory for faces. In this meta-analysis we consider a number of different performance measures associated with verbal descriptions, and examine the potential for moderator variables that might explain some of the variability across studies. In closing, we return to a theoretical framework that might assist in understanding the description-identification relationship based upon the findings of the meta-analysis, and discuss the practical implications of this relationship for eyewitness evidence.

VERBAL OVERSHADOWING VERSUS VERBAL FACILITATION EFFECTS IN FACE IDENTIFICATION

The ability of an eyewitness to verbally translate a perceptual memory is an important component of eyewitness evidence in our legal system. In fact, at numerous stages throughout the justice process an eyewitness may be asked to describe what s/he witnessed at the crime scene, and most importantly his/her perception of the culprit. It seems intuitive that an eyewitness who is capable of giving a verbal description of a perpetrator's face would also have an accurate recollection of their facial features, and subsequently be capable of identifying the perpetrator from a photo array. Research has shown, however, that verbal descriptions can sometimes interfere with subsequent attempts at perceptual identification (verbal overshadowing) and at other times enhance attempts at perceptual identification (verbal facilitation).

Referred to as *verbal overshadowing*, Schooler and Engstler-Schooler (1990) first demonstrated in a series of studies that providing a verbal description of another person's face can significantly impair our ability to recognise that face in a subsequent lineup identification task. In the years since Schooler and Engstler-Schooler's original demonstration of the verbal overshadowing phenomenon, a number of studies have replicated these results within the facial memory paradigm (Dodson, Johnson, & Schooler, 1997; Fallshore & Schooler, 1995; Ryan & Schooler, 1998; Schooler, Ryan & Reder, 1996; Sporer, 1989). Furthermore, researchers have shown the overshadowing phenomenon to occur in other domains involving "difficult to describe" perceptual experiences, including wine tasting (Melcher & Schooler, 1996), visual forms (Brandimonte, Hitch, & Bishop, 1992), and Euclidean distance estimations (Fiore & Schooler, 2002). While the phenomenon has been replicated many times within the facial memory domain, notable failures to replicate have also occurred (Davids et al., 2006; Lovett, Small, & Engstrom, 1992; Meissner et al., 2001; Sauerland, Holub, & Sporer, 2008 this issue; Yu & Geiselman, 1993).

As a manner in which to examine this variability across studies, Meissner and Brigham (2001a) conducted a meta-analysis of the verbal overshadowing effect in face identification studies. The authors located 15 research papers comprising 29 tests of the overshadowing hypothesis and more than 2000 participants. Across studies, Meissner and Brigham observed a small, yet significant, negative effect of verbalisation on subsequent identification. Taken together, participants who were asked to describe the target face were 1.27 times more likely to *misidentify* the target than participants who did not generate a verbal description. A handful of studies were excluded from Meissner and Brigham's meta-analysis due to failures in following the constraints of the typical overshadowing paradigm, including the use of multiple target faces or alternative identification procedures. The authors analysed these studies separately and found something rather interesting—namely, a *verbal facilitation* effect in which participants who generated a description (or series of descriptions across multiple faces) were 1.38 times more likely to *correctly identify* the target faces when compared with no-description control participants.

The facilitating effects of verbalisation are not novel; in fact, a number of published studies have demonstrated improvements in face identification following generation of a verbal description (Brown & Lloyd-Jones, 2005, 2006; Chance & Goldstein, 1976; Davids et al., 2006; Itoh, 2005; Mauldin & Laughery, 1981; McKelvie, 1976; Meissner, 2002; Meissner et al., 2001; Read, 1979; Wogalter, 1991, 1996). In a recent series of studies, Brown and Lloyd-Jones (2005) replicated this verbal facilitation using a face recognition paradigm in which participants viewed and, in some cases described, a series of faces. Across four experiments, Brown and Lloyd-Jones demonstrated that participants who described the faces performed significantly better when later attempting to perceptually discriminate between previously viewed and novel faces at test. These facilitating effects occurred regardless of the type of description task participants were provided (e.g., similarities vs. differences; holistic vs. featural), and could be localised to faces that had been described previously.

How might we understand the cognitive mechanisms that can lead to verbal overshadowing versus facilitation effects in memory for faces? A review of the literature suggests that several moderator variables may distinguish these studies. First, the majority of verbal facilitation studies have typically employed face recognition or multiple face paradigms, whereas studies in the classic overshadowing literature have involved perception and verbalisation of a single target face. Interestingly, studies within the overshadowing paradigm that have employed multiple study–test trials have found that the interfering effects of description occur only on the first trial, while subsequent trials generally demonstrate null effects of verbalisation (e.g., Fallshore & Schooler, 1995; see Schooler et al., 1996, for a discussion of this "trial effect"). In addition to the multiple face distinction, studies of verbal facilitation versus overshadowing also vary in the degree of delay between encoding and description of the target face, with studies using multiple face paradigms generally requesting a verbal description immediately following presentation of the target face (in the context of the encoding task) and those in the single face paradigms generally providing a delay prior to the description task.

A second moderator variable that appears to distinguish these studies involves the extent to which participants are provided an opportunity to generate a verbal description of each target face. For example, Schooler and Engstler-Schooler (1990) had participants generate a detailed verbal description of the target face for 5 min prior to attempting identification, and subsequent studies employing this paradigm have used similar procedures. In contrast, studies that have demonstrated verbal facilitation effects often vary in the type of description task they employed and the length of the description interval. In the studies conducted by Brown and Lloyd-Jones (2005, 2006), for example, participants were provided 15 s to generate a description of each face prior to the presentation of a subsequent face. A study by Meissner et al. (2001) also examined the influence of the description task by directly manipulating participants' criterion of responding on the description task via an instructional manipulation. The authors' found that participants encouraged to provide lengthy and detailed descriptions (loose criterion) performed more poorly on the identification task when compared with those given a free recall instruction and those that were instructed to provide very brief, but precise, descriptions (strict criterion). Interestingly, participants in the strict criterion condition actually demonstrated a verbal facilitation effect when contrasted with the no-description control condition.

Finally, it is noteworthy that a handful of studies have observed verbal facilitation effects following lengthy delays (2 or more days) between the description and identification tasks (e.g., Davids et al., 2006; Itoh, 2005), while other studies have noted attenuation of the verbal overshadowing effect following significant delays (e.g., Finger & Pezdek, 1999; see Meissner & Brigham, 2001a). We can only speculate why such a reversal of the verbal overshadowing effect may occur after such a long delay. Perhaps, when the original memory trace has faded remembering some aspect of a face may serve as an effective retrieval cue, even though this cue is only available in verbal form (see Sporer, 2007). This reasoning is akin to the outshining hypothesis, according to which context reinstatement cues at testing are more likely to be effective when the memory trace is degraded (but not when the memory is still strong), thus "outshining" the effect of other retrieval cues (Cutler, Penrod, & Martens, 1987; Smith, 1988). Alternatively, the description task could serve to preserve the memory trace over the extended delay when compared with a no-description control condition—a phenomenon consistent with the notion of "output encoding" (Humphreys & Bowyer, 1980; see also Meissner & Brigham, 2001a).

Based upon studies largely in the verbal overshadowing domain, several theoretical accounts have been proposed to explain the cognitive mechanisms leading to verbal overshadowing versus facilitation. First, in their original demonstration of the overshadowing effect, Schooler and Engstler-Schooler (1990) proposed that the product of verbal description may inappropriately "recode" participants' representation of the target face and thereby interfere with subsequent attempts at identification. This *retrieval-based interference* explanation is quite consistent with the influence of retrieval processes on memory across a range of studies in the cognitive literature, including the role of *output encoding* in basic memory studies of the recognition-recall relationship (Humphreys & Bowyer, 1980). As discussed in a review by Roediger and Guynn (1996), variation in individuals' initial retrieval processes can significantly influence subsequent attempts at recollection, including both positive effects that aid subsequent recollection and negative effects that result in forgetting, interference, or even false recollections. In their research on the instructional bias effect, Meissner and colleagues (Meissner, 2002; Meissner et al., 2001) explored whether the product of participants' verbal description might mediate the overshadowing versus facilitation effects observed. Prior research has provided mixed support for the relationship between the contents of verbal descriptions and participants' identification performance; however, the instructional manipulation employed by Meissner and colleagues provided a unique opportunity to examine this relationship in the absence of *range restrictions* that were often observed in prior studies. Across their studies, the authors found a consistent relationship between identification performance and both description accuracy and the frequency of inaccurate details (see also Finger & Pezdek, 1999, for similar results). However, other studies have not found such a description-identification relationship (Brown & Lloyd-Jones, 2002; Fallshore & Schooler 1995; Kitagami, Sato, & Yoshikawa, 2002; Schooler & Engstler-Schooler, 1990), and this has served as a major source of scepticism for a retrieval-based account.

Although a retrieval-based account provides a viable explanation for verbal overshadowing when a relationship exists between description accuracy and identification performance, it does little to explain why verbal overshadowing can occur when such a relationship does not exist. Furthermore, a retrieval-based processing account has difficulty explaining overshadowing effects that result from verbalisation of a different stimulus from that of the target (Brown & Lloyd-Jones, 2003; Dodson et al., 1997; Westerman & Larsen, 1997). As a result of these and other findings, Schooler and colleagues (Schooler, 2002; Schooler, Fiore, & Brandimonte, 1997) have suggested that verbal overshadowing effects may be the result of separable cognitive processes mediating attempts at verbal description and perceptual identification. Referred to as *transfer inappropriate processing*, this theory posits that verbal descriptions instantiate a *featural* process orientation that carries over to the perceptual identification task and thereby conflicts with the *configural* or *holistic* process orientation that was likely employed at encoding. As a result, the processing orientation at retrieval fails to match that used at encoding and thereby disrupts attempts at identification (or more likely, undermines any potential facilitation that might be gained by matching processes; e.g., encoding specificity or transfer appropriate processing). This notion of processing differences is quite consistent with a variety of models that have proposed separable memory systems responsible for verbal vs. visual processing (e.g., Tulving, 1985; Tulving & Schacter, 1990) and those suggesting independent coding of verbal versus visual information in the cognitive system (Paivio, 1971; Woodhead & Baddeley, 1981).

Debate regarding the mechanisms underlying the effects of verbal description on face recognition has also recently seen the addition of a third perspective suggesting that verbalisation rather simply induces a *criterion shift*—that is, individuals who provide a description (regardless of its accuracy) are subsequently less likely to make a positive identification (regardless of accuracy) when compared with no-description control participants. In testing this hypothesis, Clare and Lewandowsky (2004) found that verbal description of a previously presented face impaired performance on target-present lineups only when participants were provided a "not present" option. Moreover, on target-absent lineups, verbalisation actually improved performance as the conservative shift led to fewer false identifications. Although a study by Sauerland and colleagues (2008 this issue) has recently confirmed this result, other studies have been less successful in replicating these findings. For example, researchers have found verbal overshadowing effects with paradigms that did not include a "not present" option (e.g., Fallshore & Schooler, 1995) and with the use of target-absent lineups (e.g., Meissner, 2002). In addition, use of a recognition paradigm introduced by Brown and Lloyd-Jones (2002, 2003) demonstrated verbal overshadowing (and verbal facilitation effects, see Brown & Lloyd-Jones, 2005, 2006) on signal detection measures of discrimination, but not on measures of response criterion.

Finally, researchers that have focused on verbal facilitation effects have often employed a *levels of processing* framework (Craik & Lockhart, 1972) to suggest that deeper encoding strategies (e.g., trait judgements regarding a target face) should facilitate later memory performance when compared with more shallow encoding strategies (e.g., categorical judgements of target race or gender) (however, see Sporer, 1991). Generally speaking, researchers have distinguished between those deeper processing strategies involving greater quantity/quality of visual processing (e.g., Wells & Hryciw, 1984; Winograd, 1981) and those involving greater semantic encoding (e.g., Anderson & Reder, 1979). While a recent study by Brown and Lloyd-Jones (2006) favours the semantic encoding alternative, further studies examining this approach appear warranted. Importantly, however, one must consider that studies invoking this approach have generally involved the use of multiple faces at encoding and have generally requested descriptions immediately following presentation of a target face (and preceding the presentation of a subsequent target face). Thus, descriptions are elicited as part of the encoding task in such studies rather than as a distinct memory phase of the experiment (as in most verbal overshadowing studies). One way in which to reconcile this distinction across studies may be to consider the extent to which the product of the description task might influence subsequent identification performance. For example, consistent with both the notion of retrieval-based processing (Meissner et al., 2001) and output encoding (Humphreys & Bowyer, 1980), descriptions with greater richness and accuracy may be more likely to facilitate subsequent memory performance when compared with those that lack detail or include inaccurate aspects.

Taken together, it appears that retrieval-based processing, transfer inappropriate processing, and levels of processing frameworks can account for a variety of conditions that lead to verbal overshadowing versus verbal facilitation. Given recent suggestions that the encoding and recognition of faces involve both verbal/featural and visual/holistic processing elements, it is possible that multiple mechanisms may work together to produce the variety of negative and positive effects observed in the literature—this possibility will be discussed later. One of the greatest difficulties for the retrieval-based processing account, however, has involved the perceived lack of a relationship between the contents of verbal descriptions and subsequent identification performance. In the next section we seek to provide the first quantitative review of this relationship by examining studies that have estimated this correlation. In our analysis we also consider various methodological variables (relating to those described earlier) that might mediate the variance across studies.

THE DESCRIPTION-IDENTIFICATION RELATIONSHIP IN MEMORY FOR FACES

It seems quite intuitive that witnesses who are better at describing a perpetrator should also be better at identifying him/her. The nature of this relationship is inherent in the arguments posed in many eyewitness cases where inconsistencies between a witness's initial description of a perpetrator and the appearance of the suspect are highlighted to undermine the credibility of the identification. Both the US Supreme Court (*Neil v. Biggers, 1972*) and the German Supreme Court (for reviews, see Meurer, Sporer, & Rennig, 1990; Odenthal, 1999) have used the quality of person descriptions as indicators of the accuracy of person identifications in criminal trials (see Sporer & Cutler, 2003). Unfortunately, at the time of these rulings little empirical research had been conducted from which the Courts might have based their decisions.

Despite the appeal of the belief that a strong relationship should exist between face description quality and identification accuracy, estimates of this relationship appear to vary considerably across studies. A few circumstances have been identified under which a significant relationship has been observed. For example, studies that have compared the relative ease with which different faces can be described versus recognised (e.g., multiple face, recognition paradigms) have noted significant description-identification correlations. Wells (1985) showed participants multiple faces and then examined both their ability to describe and recognise each face. He found that distinctive faces tended to be both easier to describe and easier to recognise than less distinctive faces, thereby leading to a modest relationship between recognition accuracy and description quality ($r = .27$) across faces. Along similar lines, Wickham and Swift (2006) have demonstrated that typical faces tend to produce verbal overshadowing effects, whereas easier-to-describe distinctive faces were unaffected by a verbalisation task.

A second condition under which a relationship between description quality and recognition performance has been observed involves studies in which participants were forced to generate rather elaborate descriptions of faces and were later asked to identify these individuals in a lineup identification task (Finger & Pezdek, 1999; Meissner, 2002; Meissner et al., 2001). In these studies, it appears that the elicitation of elaborate verbal descriptions may lead participants to generate inaccurate details, producing a self-generated misinformation effect that subsequently impairs recognition performance (Meissner, 2002; Meissner et al., 2001; Sauerland et al., 2008 this issue). As noted previously, studies that have varied response output on the description task also reduce the likelihood of range restriction on estimates of description quality, and thereby maximise their opportunity to estimate the description-identification relationship.

Interestingly, reviews of description and identification performance in memory for faces have noted many conditions that influence both recall and recognition measures in the same manner (either positively or negatively; see Meissner, Sporer, & Schooler, 2007; Sporer, 1996; Tredoux et al., 2004). These include a variety of encoding manipulations (e.g., opportunity to view the target face, stress or anxiety, alcohol intoxication at the time of encoding, and weapon focus), testing conditions (e.g., context reinstatement instructions), and the retention interval between encoding and retrieval (e.g., length of the retention interval, misinformation effects, and cowitness effects). Furthermore, studies have suggested that factors such as the age of the witness can influence both description quality and identification accuracy (e.g., Haas & Sporer, 1989). In fact, finding conditions under which the two measures become dissociated may prove quite valuable to understanding the constraints of the relationship between the two processes. One notable example has involved the cross-race effect in memory for faces (for reviews, see Meissner & Brigham, 2001b; Sporer, 2001). While studies have consistently demonstrated the cross-race effect in face identification, only a handful of studies have examined the quality of descriptions generated for same- and other-race target faces, producing rather mixed results (Dore, Brigham, Moussallie, Bennett, & Butz, 2005; Ellis, Deregowski, & Shepherd, 1975; McQuiston-Surrett & Topp, 2004; Mitchell, Meissner, & MacLin, in press; Shepherd & Deregowski, 1981).

In short, despite the intuition that witnesses who are better at describing a target should also be better at recognising him/her, this relationship has often proven quite difficult to demonstrate empirically. It may be that while verbal and perceptual tasks overlap to a certain degree in the processes demanded at encoding and retrieval (cf. Cabeza & Kato, 2000; Wickham & Swift, 2006), description tasks may encourage somewhat more featural processing (e.g., Sporer, 1989; Wells & Turtle, 1988), and make difficult the expression of less verbalisable, configural information regarding a face (e.g., Diamond & Carey, 1986). This minimal degree of covariance for featural information across description and identification tasks may well explain the modest correlations between description and identification performance frequently observed across studies (Meissner et al., 2007; Sporer, 1996), as well as the stimulus-based effects noted by other researchers (Wells, 1985; Wickham & Swift, 2006).

Given the variability of findings across studies and the frequency with which Courts often rely upon witness descriptions as indicators of identification accuracy, we sought to conduct the first meta-analysis of the description-identification relationship in memory for faces. As will be described later, we assessed this relationship across a variety of description quality measures (including accuracy, quantity, correct details, and incorrect details) and examined a number of methodological variables that might moderate the relationship. After presenting the results of this meta-analysis, we return to a brief discussion of its impact on the theoretical underpinnings of the relationship between description quality and identification accuracy and address the practical implications this relationship for eyewitness memory.

META-ANALYSIS OF THE DESCRIPTION-IDENTIFICATION RELATIONSHIP IN MEMORY FOR FACES

Method

Studies

Research papers were obtained via several methods, including: (a) searches of the *PsycINFO*, *Dissertation Abstracts*, *Social Sciences Citation Index*, and *First Search* databases using the key words "face identification", "face description", "verbal overshadowing", "eyewitness memory/recall/recognition", and "facial memory/recall/recognition"; (b) a search of selected conference programs (e.g., American Psychological Association, American Psychology–Law Society, Association for Psychological Science, Psychonomic Society, and Society for Applied Research in Memory and Cognition) over the past 5 years; and (c) contact with colleagues in the field who may have had knowledge of research studies that had neither been published nor presented at a conference.

Inclusion versus exclusion criteria. A total of 33 research papers were identified for inclusion in the meta-analysis, representing the responses of 4278 participants. Six of these manuscripts represented unpublished data, while three additional manuscripts had recently been accepted for publication. To be included in the primary analysis, studies must have required participants to both verbally describe and perceptually identify one or more faces, and must have provided an estimate of the relationship between some measure of description quality (or quantity) and identification accuracy. Studies may have employed either a single face or a multiple face paradigm, and could utilise either a lineup identification task or a recognition paradigm at test. Studies were generally excluded from the analyses if they failed to provide an estimate of the relationship between description quality and identification accuracy.

When there was reason to assume that an experimental condition (e.g., presence of a weapon in studies on weapon focus, or rereading one's prior description before identification) would lead to different correlations that would not be considered representative of the associations normally found between the description variables and identification accuracy, we used only those correlations for our analyses that were not likely to have changed from experimental manipulations along with their respective sample sizes. Thus, we used only the no-weapon control group condition for a study on weapon focus (Bothwell, Trahan, & Newsome, 1991), and the description-only condition in studies that also had a description rereading group (Davids et al., 2006; Sauerland et al., 2008 this issue; Sporer, 2007). These instances are documented in the notes to Appendix A.

Estimates of effect size. Most studies included in our meta-analyses reported point-biserial correlations between identification accuracy and the description variables. As correlations reported in the original studies should not be used for meta-analytic purposes (Lipsey & Wilson, 2001; Rosenthal, 1991), we first converted the respective *r*s using Fisher's *Zr* transformation. All analyses were conducted with these *Zr* values; however, for better understanding, all reports of effect sizes and confidence intervals were back-transformed using the inverse of Fisher's *Zr* transformation.

Some authors simply reported that a specific correlation was "not significant". In cases where we could not obtain the exact values by writing to the authors, these correlations were set to $r = .00$. This was only necessary for a few correlations (overall 8 of the 119 correlations analysed), and such instances are documented in the results (Table 1) as well as in Appendix A. Some studies reported separate correlations for different experimental conditions (e.g., target-present vs. target-absent lineups), or for subgroups of participants (e.g., children vs. adults, choosers vs. nonchoosers). In most of these cases we used the values of each condition as independent estimates of the associations of interest, coding the respective conditions as potential moderator variables.

Measures of description quality. Five measures of description quality were examined in the current meta-analysis. First, the relationship between identification accuracy and *description accuracy* (generally calculated as the proportion of correct details divided by the total number of details provided) was calculated and presented in 21 of the research papers and resulting in $k = 32$ hypothesis tests of this relationship across a total of 2973 participants. Second, we identified 18 research papers that estimated the relationship between identification accuracy and *description quantity* (generally calculated as the total number of facial descriptors, not including subjective or personality aspects, generated by a participant), representing $k = 33$ hypothesis tests of the relationship across 2578 participants. Third, 13 research papers assessed the relationship between identification accuracy and the *number of correct descriptors* generated for a given face, representing

TABLE 1
Meta-analytic summary of correlations between measures of description quality and identification accuracy

Summary statistics	*Accuracy*	*Quantity*	*Correct descriptors*	*Incorrect descriptors*	*Congruence*
k	32	33	22	16	5
k (est. = .00)	4	1	1	2	0
N	2973	2578	1932	1640	279
Min r	−.22	−.61	−.33	−.46	−.25
Max r	.39	.69	.31	.11	.25
Unweighted mean r	.12	.02	.01	−.16	.07
Weighted mean r	.14	−.04	−.02	−.18	.12
p	.000	.065	.335	.000	.046
95% CI—low	.11	−.08	−.07	−.23	.00
95% CI—high	.18	.00	.02	−.13	.24
N_{FS}	440	−32	−22	217	1
Homogeneity Q	66.52	102.16	43.05	41.89	6.51
$df(Q)$	31	32	21	15	4
$p(Q)$	.000	.000	.003	.000	.164

k = number of significance tests; k (est. = .00) = number of "*ns*" results set to $r = .00$; N_{FS} = failsafe N.

$k = 22$ tests of the hypothesis across 1932 participants. Fourth, the relationship between identification accuracy and the *number of incorrect descriptors* generated for a given face was estimated in nine research papers, resulting in $k = 16$ tests of the hypothesis across 1640 participants. Finally, four research papers assessed the relationship between description accuracy and the *congruence* (or degree of similarity) between the description provided and the face that was identified from the photo array, representing $k = 5$ tests of the hypothesis across 279 participants.

Coding of study characteristics as moderator variables. Two independent raters coded a host of study characteristics that served not only to document the type of studies reviewed but also to allow us (and future researchers) to analyse for systematic differences across studies. The following study characteristics were recorded for each study: (a) degree of realism at encoding (i.e., slide or photo vs. video vs. live event), (b) type of description task (i.e., rating scales vs. checklist vs. cued recall vs. free recall), (c) time permitted to encode each target face, (d) number of faces encoded at study, (e) delay between encoding and the description task, (f) delay between generating the description and subsequent identification, and (g) publication status (i.e., unpublished vs. published). We also coded a host of other study characteristics. Although moderator analyses with these variables were also conducted, they are somewhat redundant with those reported here, but are available from the authors upon request. Few disagreements occurred in the coding, and these differences were resolved by the first and second authors. Appendix B provides the codings of study characteristics across the sample of studies.

Results and discussion

Effect size analyses

In this section, we present the results of separate meta-analyses involving various relationships between description quality and identification accuracy in memory for faces. For each of these meta-analytic integrations we (1) calculate the mean relationship across studies, along with 95% confidence levels (CI); (2) assess whether the results can be considered homogeneous; and (3) in case of heterogeneity search for relevant moderator variables that may account for differences across studies. The meta-analytic procedures used followed recommendations by Lipsey and Wilson (2001), as well as Rosenthal (1991), and the respective chapters in Cooper and Hedges (1994). To assess computational accuracy, we used different software algorithms developed by Lipsey and Wilson, as well as adaptations of the examples provided by Becker (1994) and Hedges and Olkin (1985) that were programmed by the second author.

Primary analyses are reported as weighted means calculated by using the inverse variance weights for each coefficient (which is based on the sample size of the respective studies); however, unweighted means are also provided as an additional measure of central tendency. Appendix A provides the point-biserial correlations of the various description-identification relationships that were used as the basis for our meta-analyses. In most cases, the units of analyses are identical with individual experiments; however, in some experiments, different conditions were likely to lead to different correlations between description quality and identification accuracy. Whenever available, we recorded separate correlations for subgroups of experiments. Therefore, the unit of analysis for any meta-analysis conducted is the number of hypothesis tests, *k*, available for the particular association. Appendix B provides a description of all the variables coded from the respective method sections of the reports and/or from consulting with the authors. The primary results of our meta-analyses are summarised in Table 1. As we discuss the significance and magnitude of each relationship, we base our interpretations within the context of effect size conventions proposed by Cohen (1988). Specifically, a point-biserial correlation of .10 is considered a small effect size, .24 a medium effect size, and .37 a large effect size. Table 1 also contains fail-safe numbers (N_{FS}) based on unweighted integrations involving *p*-values using the Stouffer method (see Rosenthal, 1991). Figure 1 displays the distributions of relationships investigated (median, quartiles, and deciles of the *Zr* values).

Practically all of the relationships examined were highly heterogeneous (see Table 1). One way to address this problem is to identify potential outliers that may have obscured meaningful patterns in the relationships studied. We utilised both graphical methods (stem-and-leaf plots and line plots that involved rank ordered individual study effect sizes with 95% confidence intervals; see Begg, 1994), as well as more formal meta-analytic techniques such as those suggested by Hedges and Olkin (1985).[1] As such, we provide an assessment of the impact of such outliers for each effect size

[1] Hedges and Olkin (1985) have suggested the use of standardised residuals as well as homogeneity statistics to search for outliers. Unfortunately, they only provide formulae for the effect size *d*. We have adjusted these procedures to *Zr* as effect size, calculating the residuals and homogeneity statistics for *Zr* analogous to the procedure adopted by Hedges and Olkin for *d*. The basic rationale of this procedure is to calculate adjusted mean effect sizes after removing the effect size in question and then examine the standardised residual of this particular effect size, as well as the homogeneity *Q* after removal of this study. Residuals larger than 2 are considered as potential outliers. Removal of one, or several, outliers should reduce the observed heterogeneity indicated by a failure to reject the null hypothesis of homogeneity for the reduced number of studies. However, removal of these studies may inadvertently lead to a drop of important "exceptions" of the general observed pattern which (through the conduct of moderator analyses) could be particularly interesting in understanding the underlying theoretical mechanisms.

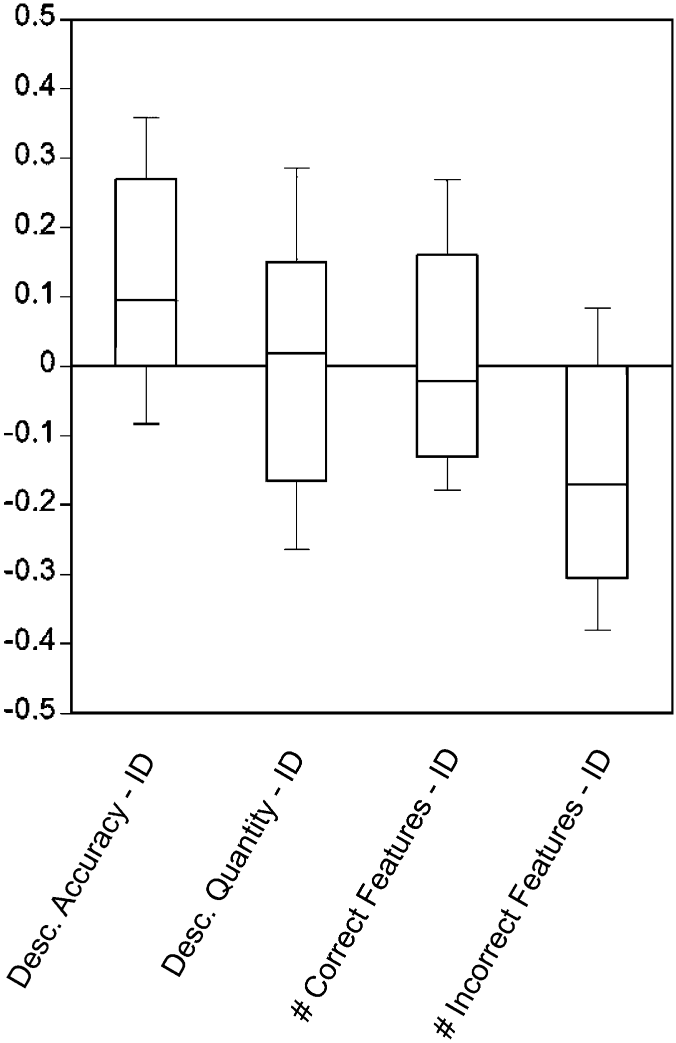

Figure 1. Box plots of correlations between measures of description quality and identification accuracy.

calculated. Box plots of all relationships were also examined here (Figure 1; see also Light, Singer, & Willett, 1994).[2]

Relationship between description accuracy and identification accuracy. Most of the studies reviewed reported point-biserial correlations

[2] Additional figures displaying these relationships are available from the second author.

between description accuracy and identification accuracy, leading to a $k = 32$ hypothesis tests of this relationship with $N = 2973$. The weighted mean effects size was $r = .14$, a small yet significant effect, $p < .001$, with CIs of .11 and .18, $N_{FS} = 440$. Results were heterogeneous, $Q(31) = 66.52$, $p < .001$. Analysis of the data for outliers suggested that studies by Bothwell et al. (1991, Exp. 2) and Pigott, Brigham, and Bothwell (1990) had shown the two most negative relationships, while studies by Meissner (2002, Exp. 1) and Meissner et al. (2001, Exp. 1) had observed significant positive relationships with fairly large samples. These four studies all had standardised residuals larger than 2.0 when compared with the weighted mean effect size after removal of the respective studies. Although dropping none of the studies individually would make the results homogeneous, removing all four as outliers led to a nonsignificant $Q(27) = 36.53$, $p > .10$. The removal of these studies, however, did not significantly alter the mean weighted effect size, $r = .12$, $p < .001$, CIs of .08 and .16, now based upon $k = 28$ hypothesis tests with $N = 2244$. Nonetheless, the variability across studies is still large enough to take a closer look at potential moderators.

Relationship between the description quantity and identification accuracy. In actual criminal cases, the accuracy of person descriptions cannot be ascertained as this requires knowledge of the true perpetrator; however, description quantity can be assessed. A total of $k = 33$ tests of the hypothesis involving $N = 2578$ resulted in a nonsignificant weighted mean of $r = -.04$, $p = .07$, with CIs of $-.08$ and .00. The results were highly heterogeneous, $Q(32) = 102.16$, $p < .001$, with *r*s ranging from $-.61$ to .69. Outlier analyses revealed 11 studies with standardised residuals above 2.0 (including Studies 34, 39, 23, 29, 38 with negative, and 51, 45, 28, 19, 32, and 31 with positive values; see Appendix A). Although the heterogeneity is extreme, we decided not to remove these outliers, but rather to search for moderators that might explain these differences.

Relationship between the number of correct descriptors and identification accuracy. Besides description accuracy, some studies have also reported the number of correct and/or the number of incorrect descriptors and their relationship to identification accuracy. For $k = 22$ tests of the hypothesis involving $N = 1932$, there was no significant relationship between the number of correct descriptors and identification accuracy, weighted mean $r = -.02$, with CIs of $-.07$ and .02. Results were also heterogeneous, $Q(21) = 43.05$, $p < .01$. Analysis of outliers suggested that studies by Memon, Rose, Searcy, and Bartlett (1999, Exp. 2), Finger (2002, Exp. 1), Gwyer and Clifford (1997, target-present lineups), Geiselman, Schroppel, Tubridy, Konisi, and Rodriguez (2000, Exp. 2, target-absent lineups), and Hosch and Bothwell (1990) had standardised residuals over 2.0. The study by Memon et al. (1999), with an effect size of $r = -.33$, appears particularly out of the range of most other values. Removal of three studies (Studies 26, 22, and 28) with more extreme standardised residuals (one with a negative, two with positive effect sizes) resulted in a homogeneous distribution, $Q(18) = 25.42$, $p > .10$; however, the overall conclusion that there is no relationship between number of correct descriptors and identification accuracy does not change when these more extreme values are removed, weighted mean $r = -.03$, CIs $-.08$ and .01 based upon $k = 19$ hypothesis tests with $N = 1741$.

Relationship between the number of incorrect features and identification accuracy. For the $k = 16$ hypothesis tests with $N = 1640$ investigating incorrect details, a significant effect was observed, with mean weighted $r = -.18$, $p < .001$, CIs of $-.23$ and $-.13$, $Nfs = 217$. Results of this relationship were also heterogeneous, $Q(15) = 41.89$, $p < .001$. Analysis of outliers suggested that the study by Geiselman et al. (2000, Exp. 2, target-present lineups) had a standardised residual close to 3.0, while studies by Sauerland et al. (2008 this issue), Meissner (2002, Exp. 2), Kitagami et al. (2002), and Geiselman et al. (2000, Exp. 2, target-absent lineup) showed standardised residuals slightly above 2.0. Removal of these studies rendered the distribution homogeneous, $Q(10) = 15.13$, $p > .10$, though the relationship between incorrect descriptors and identification accuracy remained significant, with a mean weighted $r = -.21$, $p < .001$, CIs of $-.26$ and $-.15$, based upon $k = 11$ tests of the hypothesis with $N = 1214$ participants.

Relationship between congruence and identification accuracy. Finally, only five studies with a total of 279 participants investigated the relationship between congruence and identification accuracy. The weighted mean $r = .12$, $p < .05$, with CIs of .00 and .24, $N_{FS} = 1$. The results appear to be homogeneous, $Q(4) = 6.51$, $p = .16$, with *r*s ranging from $-.25$ to .25. It should be noted, however, that the Wells (1985) study used the number of faces as the unit of analysis, which led to an $r = .19$.

Relationships among the description quality measures. Though ancillary to our primary analyses, some studies provided estimates of the correlation between various measures of description quality. We focus briefly on two of these relationships. First, $k = 9$ studies ($N = 1098$) examined the relationship between the number of descriptors and description accuracy. A significant correlation was observed across studies, with a mean weighted $r = -.26$, $p < .001$, and CIs of $-.32$ and $-.21$, $N_{FS} = 79$. The results were heterogeneous, $Q(8) = 63.22$, $p < .001$, with *r*s ranging from $-.48$ to .34. The two extreme values were observed in studies by Meissner (2002), $N = 432$) and Davids et al. (2006), $N = 48$), both of which could be considered outliers.

When removed from the analysis, the results become homogeneous, $Q(6) = 7.22$, $p = .301$, though the significance of the relationship remains with a mean weighted $r = -.13$, $p < .001$, with CIs of $-.21$ and $-.05$, $N_{FS} = 12$. Second, as only two studies investigated the relationship between description accuracy and congruence, no detailed meta-analytic analyses were performed; however, we include these studies here to note the rather high correlations observed by both Wells (1985), $r(125) = .87$, and Pigott et al. (1990), $r(24) = .67$. The weighted mean effect size across the two studies was $r = .85$, with CIs of .80 and .89, $N_{FS} = 103$.

Moderator analyses

As noted previously, practically all of the effect size relationships examined were highly heterogeneous, suggesting that a search for study characteristics may help us to understand the differences across studies. In the current analyses we examine the association between effect sizes across studies and the study characteristics recorded. Unfortunately, the number of studies is too small to conduct more complex blocking or multiple regression analyses due to mutual dependencies (collinearity) among predictors. Hence, we focus only on a few moderators that were likely to have affected these relationships. Tables 2, 3, 4, and 5 display the results of moderator analyses of effect sizes as a function of study characteristics. For each analysis, the Q_{BET} indicates whether the subgroups differed as a function of this moderator. Q_{WIT} indicates whether the remaining variance is homogeneous. When no studies or only a single study was available for a certain subgroup, subgroups were collapsed. Next we discuss the significant patterns observed in these analyses and present the mean weighted *r*s for each condition where applicable. Given the small and nonsignificant nature of the relationships observed in the earlier effect size analyses, we further restrict our discussion

TABLE 2
Moderator analysis of the relationship between description accuracy and identification accuracy

Predictor variable	Q_{BET}	*df*	*p*	Q_{WIT}	*df*	*p*
Encoding realism	32.07	2	.00	34.45	29	.22
Description task	3.08	3	.38	63.43	28	.00
Encoding time	3.93	2	.14	62.59	29	.00
Encoding load	3.96	1	.05	62.56	30	.00
Encoding-description delay	6.40	2	.04	60.12	29	.00
Description-identification delay	22.20	2	.00	44.32	29	.03
Publication status	0.00	1	.95	66.51	30	.00

TABLE 3
Moderator analysis of the relationship between description quantity and identification accuracy

Predictor variable	Q_{BET}	*df*	*p*	Q_{WIT}	*df*	*p*
Encoding realism	14.82	2	.00	87.34	30	.00
Description task	2.86	2	.24	99.29	30	.00
Encoding time	15.02	2	.00	87.14	30	.00
Encoding load	2.77	2	.25	99.38	30	.00
Encoding-description delay	24.77	2	.00	77.39	30	.00
Description-identification delay	9.38	2	.01	92.77	30	.00
Publication status	0.00	1	.98	102.16	31	.00

to those moderators that produced patterns yielding significant relationships between description measures and identification accuracy.

Relationship between description accuracy and identification accuracy. Table 2 displays the results of moderator analyses conducted on the relationship between description accuracy and identification accuracy. A few notable findings merit discussion. First, this correlation appears to decrease when realism increases at encoding such that the strongest correlation is observed in studies that employed photographic stimuli, $r = .26$, $p < .001$, $k = 9$, followed by studies that utilised video stimuli, $r = .12$, $p < .001$, $k = 12$. In contrast, studies that employed a live event showed no significant relationship between description accuracy and identification accuracy, $r = .04$, *ns*, $k = 11$. Second, face recognition paradigms that employed more than one target face generally showed larger relationships, $r = .27$, $p < .001$, $k = 5$, when compared with lineup identification studies that utilised only a single target face, $r = .13$, $p < .001$, $k = 27$.

Finally, the delay between encoding, description, and identification phases across experiments appears to have moderated the relationship

TABLE 4
Moderator analysis of the relationship between number of correct details and identification accuracy

Predictor variable	Q_{BET}	*df*	*p*	Q_{WIT}	*df*	*p*
Encoding realism	7.10	2	.03	35.95	19	.01
Description task	3.61	1	.06	39.43	20	.01
Encoding time	0.41	2	.82	42.64	19	.00
Encoding load	3.42	1	.06	39.63	20	.01
Encoding-description delay	6.37	1	.01	36.68	20	.01
Description-identification delay	2.31	2	.31	40.73	19	.00
Publication status	4.89	1	.03	38.16	20	.01

TABLE 5
Moderator analysis of the relationship between number of incorrect details and identification accuracy

Predictor variable	Q_{BET}	*df*	*p*	Q_{WIT}	*df*	*p*
Encoding realism	11.67	2	.00	30.22	13	.00
Description task	0.77	1	.38	41.12	14	.00
Encoding time	0.96	2	.62	40.93	13	.00
Encoding load	0.22	1	.64	41.67	14	.00
Encoding-description delay	5.03	1	.02	36.86	14	.00
Description-identification delay	20.17	2	.00	21.72	13	.06
Publication status	0.77	1	.38	41.12	14	.00

between description accuracy and identification accuracy. In particular, shorter delays between encoding and generation of a description produced stronger effects when compared with delays longer than 1 day: no delay, $r = .11$, $p < .001$, $k = 14$; 5 min to 1 day delay, $r = .19$, $p < .001$, $k = 16$; more than 1 day delay, $r = .09$, $p < .05$, $k = 2$. In contrast, the presence of a delay between the description and identification phases yielded stronger correlations between description accuracy and identification accuracy: no delay, $r = .06$, $p < .05$, $k = 17$; 5 min to 1 day delay, $r = .23$, $p < .001$, $k = 10$; more than 1 day delay, $r = .23$, $p < .001$, $k = 5$.

Relationship between the description quantity and identification accuracy. The moderator analyses for the relationship between description quantity and identification accuracy are shown in Table 3. Consistent with the prior analyses, studies that employed photographic stimuli showed a small, but significant, negative association, $r = -.11, p < .001, k = 11$, while studies utilising video stimuli, $r = .03$, *ns*, $k = 11$, and live interactions, $r = .07$, *ns*, $k = 11$, produced nonsignificant, though positive, effects. Studies that employed shorter encoding times for each target face also produced significant negative effects: 1 to 5 s, $r = -.10, p < .001, k = 9$; 5 to 60 s, $r = -.07, p \leq .06$, $k = 12$, while those that permitted longer encoding times produced significant positive effects: greater than 60 s, $r = .09, p < .05, k = 12$.

The delays between encoding, description, and identification phases across studies appears to moderate the relationship between description quantity and identification accuracy as well. With regard to the encoding to description phase delay, it appears that a modest amount of delay leads to the significant negative association between quantity and identification accuracy: 5 min to 1 day delay, $r = -.12, p < .001, k = 13$. In contrast, studies that employed either no delay, $r = .06$, $p \leq .06$, $k = 14$, or lengthy delays, $r = .12$, $p \leq .08$, $k = 6$, produced only marginally significant, but positive, relationships. Delays between the description and identification phases also yielded stronger negative correlations between description quantity and identification accuracy: no delay, $r = .04$, *ns*, $k = 18$; 5 min to 1 day delay, $r = -.08$, $p < .001$, $k = 10$; more than 1 day delay, $r = -.11$, $p \leq .07$, $k = 5$.

Relationship between the number of correct descriptors and identification accuracy. Moderator analyses for the relationship between the number of correct descriptors and identification accuracy were much less successful in shedding light on the diverse findings (see Table 4). In fact, only two of the moderator analyses produced conditions under which a significant relationship appears to exist. In particular, and consistent with prior analyses, studies utilising photographic stimuli showed a small, but significant, negative association, $r = -.07$, $p < .05$, $k = 9$, while studies utilising video stimuli, $r = .02$, *ns*, $k = 9$, and live interactions, $r = .11$, $p \leq .09$, $k = 4$, produced nonsignificant positive effects. Furthermore, the encoding-description delay also affected this relationship. When giving descriptions immediately, the relationship tended to be positive, $r = .11$, $p \leq .06$, $k = 6$, while studies with delays of 5 min up to several days showed a small but significantly negative relationship, $r = -.05$, $p \leq .06$, $k = 16$.

Relationship between the number of incorrect descriptors and identification accuracy. A small to medium relationship appears to exist between the number of incorrect descriptors mentioned and identification accuracy. Although there were only 16 hypothesis tests of this relationship, some of the moderator variables further illuminate the observed relationship (see Table 5). Once again, studies utilising photographic stimuli showed the strongest negative association, $r = -.23$, $p < .001$, $k = 9$, while studies utilising video stimuli, $r = -.06$, *ns*, $k = 5$, and live interactions, $r = .00$, *ns*, $k = 2$, produced nonsignificant effects. The effects of delay were also evident here. Longer delays between encoding and generation of a description produced stronger effects: 5 min to 1 day delay, $r = -.21$, $p < .001$, $k = 11$, while studies employing no delay showed nonsignificant effects, $r = -.06$, *ns*, $k = 5$. (It should be noted that no studies using this dependent measure included delays longer than 1 day.) In contrast, the presence of a delay between the description and identification phases yielded a stronger negative correlation when compared with the absence of a delay: no delay, $r = -.01$, *ns*, $k = 6$; 5 min to 1 day delay, $r = -.23$, $p < .001$, $k = 6$; more than 1 day delay, $r = -.30$, $p < .001$, $k = 4$.

Multiple regression analyses with encoding time, encoding-description delay, and description-identification delay

The results regarding encoding time, encoding-description delay, and description-identification delay were puzzling. One of the reasons may have

TABLE 6
Regression-based moderator analysis of the relationship between description accuracy and identification accuracy

Predictor variable	*B*	*SE*	*CI low*	*CI high*	*Z*	*p*	*Beta*
Encoding time	−.06	.04	−.14	.01	−1.58	.11	−.19
Encoding-description delay	.00	.02	−.05	.05	0.09	.93	.01
Description-identification delay	.06	.02	.02	.10	3.13	.00	.41

been that the categorical coding we used may not adequately reflect the large variations in encoding and delay times across studies. There is also the possibility that the respective encoding and delay times may have been confounded in the studies reviewed, which cannot be detected when only one variable is investigated at a time. Therefore, we conducted (multiple) regression analyses to postdict the description-identification associations from the actual times in seconds (as opposed to the categorisations reported earlier) used for encoding and for delay intervals. As the distributions of these times were heavily skewed, we employed a logarithmic transformation of these values, adding the constant 1 as many studies used no delay intervals (i.e., encoding time = log10(s + 1); delay = log10(min + 1)). These analyses were carried out for measures of description accuracy and description quantity. For the other description measures, there was not a sufficient number of studies as a prerequisite for these types of analyses. The multiple regression model involving the description accuracy measure yielded a significant $Q(3) = 13.39$, $p < .01$. Table 6 displays the results of this meta-regression analysis following the procedures by Lipsey and Wilson (2001). Only the description-identification delay reliably predicted the size of the association; however, the residual variance remained significant, $Q(28) = 53.12$, $p < .001$. An analogous multiple regression analysis was carried out for the measure of description quantity (see Table 7). The model was significant, $Q(3) = 23.30$, $p < .001$. Encoding time showed a significant positive association such that the longer participants had time to encode the target face, the stronger the relationship between description quantity and identification accuracy. In addition, description-identification delay showed

TABLE 7
Regression-based moderator analysis of the relationship between description quantity and identification accuracy

Predictor variable	*B*	*SE*	*CI low*	*CI high*	*Z*	*p*	*Beta*
Encoding time	.11	.03	.06	.16	4.17	.00	.42
Encoding-description delay	−.02	.02	−.06	.01	−1.15	.25	−.12
Description-identification delay	−.05	.02	−.09	−.02	−2.75	.01	−.28

a significant negative association such that increasing the description-identification delay produced a smaller relationship between description quantity and identification accuracy. However, as the residual variance remained highly significant, $Q(29) = 78.86$, $p < .001$, other factors may also be responsible for these associations.

Summary of findings

Across 33 research papers and a total of 4278 participants, the current meta-analysis found some support for the relationship between description quality and identification accuracy in memory for faces. More specifically, three of the five relationships examined demonstrated significant small-to-moderate effect sizes (see Table 1) such that: (a) More accurate descriptions were significantly associated with greater accuracy in identification; (b) descriptions that contained more incorrect details were associated with greater inaccuracy in identification; and (c) greater congruence between the description and the person identified was associated with greater accuracy in identification.

As is typically the situation, however, it is important to consider the boundary conditions that moderate these effects. Our analyses on this front suggest that the strongest relationships were produced in studies that employed photographic stimuli (e.g., face recognition paradigms), whereas studies that utilised more realistic stimuli such as video or live events (e.g., eyewitness identification paradigms) produced smaller and often nonsignificant effects. On a similar basis, face recognition paradigms that employed more than one target face generally showed larger relationships between description accuracy and identification accuracy when compared with lineup identification studies that utilised only a single target face.

Another important factor appeared to involve delays between the encoding, description, and identification phases of the experiment. With regard to the delay between encoding and the description phase, shorter delays produced stronger effects for the measure of description accuracy, while longer delays produced stronger effects for the number of incorrect descriptors generated. Given the effect of delay on the likely accuracy of a description, the variability of this moderator across the description measures appears reasonable and further supports the role of retrieval-based processes. Finally, across the description measures of accuracy, quantity, and number of incorrect details, a delay between the description and identification tasks produced stronger effects when compared with studies that presented the identification task immediately thereafter.

GENERAL DISCUSSION

In the present paper we have reviewed the empirical and theoretical literatures that have addressed the description-identification relationship in memory for faces. As noted previously, descriptions can sometimes impair (or "overshadow") and other times facilitate subsequent attempts at perceptual identification; however, understanding the relationship between these two tasks and the theoretical mechanisms that bridge this relationship has proven difficult. Furthermore, given the Court's assumptions regarding the usefulness of examining a witness's description when assessing the likely accuracy of his/her identification (*Neil v. Biggers*, 1972), we felt it was important to further examine this relationship across studies spanning four decades of empirical analysis. Taken together, the results of our meta-analysis suggest that a significant relationship does exist between the description measures of description accuracy, number of incorrect descriptors, and congruence with subsequent identification accuracy. Furthermore, certain conditions appear to exacerbate the magnitude of this relationship, including the use of face recognition versus eyewitness identification paradigms, and the length of delays between relevant tasks. Next we discuss both the theoretical, methodological, and practical implications of our findings, and suggest future directions for research in this area.

Theoretical implications

As reviewed previously, descriptions can render both positive and negative effects on subsequent identification performance, leading to either verbal facilitation or verbal overshadowing, respectively. Several theoretical accounts have been proposed to explain these effects, including retrieval-based processing, transfer inappropriate processing, and levels of processing theories. Each of these theories can account for a variety of findings in this literature. For example, retrieval processes can explain the effects of instructional bias and facial typicality, and are generally consistent with the effects of self-generated misinformation and repeated testing (cf. Meissner et al., 2001; Roediger, Wheeler, & Rajaram, 1993). Similarly, a levels of processing framework can account for the influence of description generation on subsequent identification performance, including verbal facilitation (Brown & Lloyd-Jones, 2005, 2006), and the role of output encoding on secondary task performance (Hintzman & Hatry, 1990). In contrast, transfer inappropriate processing can justify the effects of perceptual tasks that "release" the overshadowing of verbal descriptions and the effects of featural tasks (such as describing a different stimulus from that of the target face) that can lead to overshadowing on subsequent identification (for a review see

Schooler, 2002). Transfer inappropriate processing is also consistent with a variety of models that have proposed separable memory systems responsible for verbal versus visual processing (e.g., Tulving, 1985; Tulving & Schacter, 1990) and those suggesting independent coding of verbal vs. visual information in the cognitive system (Paivio, 1971; Woodhead & Baddeley, 1981).

One difficulty for a retrieval-based processing account has involved the often variable association between the contents of verbal descriptions and later attempts at identification. To address this issue, we sought to conduct the first meta-analysis of the description-identification relationship in memory for faces. As summarised previously, the results of this analysis provide a basis for supporting a significant weak-to-moderate relationship between the description measures of accuracy and number of incorrect descriptors on subsequent identification that is consistent with a retrieval-based (or output encoding) mechanism. However, the modest nature of this correlation may also be used to support the relative independence of verbal and visual processing that is consistent with a transfer inappropriate processing account (Schooler, 2002; see also Flexser & Tulving, 1978; Kahana, Hirsuto, & Schneider, 2005). Furthermore, the vast majority of studies exclude subjective descriptors for which it is difficult to determine accuracy, including such aspects as personality or comparative judgements (e.g., his chin is similar to that of Dick Cheney) that might be considered more "holistic" aspects. A transfer inappropriate processing account would propose that a focus on such holistic information would lead to less accurate identifications. Though we were unable to assess this prediction using the current database of studies, it is noteworthy that several recent studies conducted by Brown and Lloyd-Jones (2005, 2006) failed to find any significant correlations between the production of holistic features in a face description and subsequent recognition performance—though the authors were able to confirm the transfer inappropriate processing prediction when participants were explicitly instructed to generate holistic versus featural descriptions (but only when the analysis was conducted across items; Brown & Lloyd-Jones, 2005, Exp. 4).

While the current analysis appears to provide some support for the role of retrieval-based processing in the description-identification relationship, should this be interpreted to preclude any possibility that transfer inappropriate processing might account for some of the effects of verbal description on perceptual identification? That is, should these two accounts be considered mutually exclusive? The simple answer, we believe, is "no". Just as the effects of retrieval processes and repeated testing have been shown throughout the study of human memory, so too have the effects of encoding specificity and transfer appropriate processing. As such, we believe that most models of cognitive processing would readily accommodate the two perspectives as situationally determined effects—that is, conditions arise whereby one or both of the processes may come to influence subsequent attempts at identification. It is this line of future research, namely to determine the parameters under which the two theories might independently versus conjointly account for the effects of verbal overshadowing versus verbal facilitation, that we believe will prove fruitful to our understanding of the relationship between verbal and perceptual processes in memory for faces.

For example, one manner in which to account for a modest description-identification relationship may involve the extent to which certain processes are invoked by both recall and recognition tasks. For example, dual-process theories of memory propose that while *familiarity-based processes* support performance on recognition and implicit memory tasks, *recollection-based processes* can be applied to both recognition and recall tasks (see Yonelinas, 2002). Thus, to the extent that recollection may be employed at the time of identification, overlap may be seen in performance on description and identification tasks. For example, Brown and Lloyd-Jones (2006) found evidence of an increase in recollection-based judgements for previously verbalised faces (though an increase in familiarity-based judgements was also seen), and this recollection increase was associated with the verbal facilitation effect observed in their study. To the extent that verbalisation might invoke recollection-based processes that can subsequently be applied in an identification task, we would expect a greater relationship between performance across the two tasks—including both verbal facilitation and verbal overshadowing depending upon the veracity of the description output. This role of recollection in facilitating performance across tasks is also consistent with levels of processing manipulations within the dual-process memory literature (e.g., Gardiner, 1988).

Several moderator variables were found to exacerbate the magnitude of the description-identification relationship, including the use of facial recognition paradigms and various delays between the encoding, description, and identification phases of an experiment. We believe that theoretical accounts of memory for faces and the effect of verbal overshadowing versus verbal facilitation must take into consideration these conditions. First, studies employing facial recognition paradigms generally demonstrated stronger correlations between description and identification performance based upon study characteristics such as the use of more than one target face (as opposed to a single target face) and a shorter delay between encoding of the face and generation of a description. Although we did not code the specific nature of the identification task (i.e., recognition vs. lineup identification), this may also be an important variable to consider. Our sense is that the vast majority of such studies incorporated the description task within the context of encoding and thereby created a stronger (and more precise) representation that could be applied on the subsequent identification task. This is consistent with the predictions of both a levels of processing framework and retrieval-based mechanisms. In addition, Wells (1985) has noted that the description-identification relationship may be driven not by the ability of the participant to describe and identify faces, but rather by the variation across faces that renders some faces more easily described and identified than others. As such, facial recognition paradigms that employ a diversity of target faces provide a stronger basis upon which to estimate the description-identification relationship across target faces.

Delays between the description and identification phases also produced stronger effects. Longer delays (of several days) may facilitate memories for described faces in yet little understood ways. Some authors have recently argued that verbal facilitation effects will be more likely when the memory trace is poor, be it for a lack of encoding or an increased retention interval (Itoh, 2005). Distinctive targets may be encoded by labelling these distinctive aspects in ways that serve as retrieval cues (Sporer, 1989, 2007) or as a means of rehearsal (Read, 1979; Sporer, 1988), consistent with the notion of output encoding. Alternatively, the vagueness of descriptions may make participants aware that their memory is apparently rather poor, leading to a cautious shift in target-absent lineups (Brown & Lloyd-Jones, 2005; Sauerland et al., 2008 this issue).

Methodological implications

Our attempts to provide a meta-analytic synthesis of previous findings was hampered by some difficulties we encountered when comparing and coding the studies. While we are very grateful to many authors who have provided us with additional information on request, we believe that to be able to draw firm conclusions about the role of verbal processes and their relationship to identification several standards of reporting should be followed. As both encoding times and the delays between encoding, description, and identification seem to be quite important for the relationships observed, exact times for presentation of the stimulus and the two delays should be routinely reported. To the extent that forensic implications are sought, there is also a clear need for longer retention intervals (the longest ones we found in this literature were 2 days and 1 week, respectively, while the majority of studies used recognition tests in the same session). Researchers should also be mindful of the types of descriptions asked from their participants (see Appendix B). Estimating somebody's height and weight is unlikely to be related to a recognition task where only faces are shown, and defining accuracy simply by agreement may be an unreliable estimate of measuring description accuracy. While counting the number of features mentioned is unlikely to pose a problem when measuring description quantity, pilot testing and training of coders to assess description accuracy should lead to a better estimate of the construct of description accuracy. Although not included in Tables 2, 3, 4, and 5, additional analyses showed that effect sizes were stronger for studies that reported interrater reliabilities than for those that had either reported no reliability data or had simply indicated that accuracy was merely coded "by agreement". As is evident from the studies by Meissner et al. (2001), exact instructions regarding the amount of information requested from witnesses may also be crucial as they are likely to change the decision criterion for reporting (incorrect) details. Given the lack of methodological detail provided in the vast majority of studies, this variable could not be appropriately coded across studies. Indirectly, this problem may also be inherent in the use of checklists, which may foster featural processing by isolating individual aspects of a face to be described. Requesting many details or ratings may also induce perceivers to mark responses they may not really remember at all (Sporer, 2007; Wogalter, 1991, 1996).

We were stunned by the large variations across studies in the number of features provided and the accuracy rates obtained. Obviously, these differences are not only likely to reflect differences in description tasks (e.g., instructions to participants about the reporting criterion), but also differences in scoring and operationalisations of "accuracy". Some studies involved only physical descriptions of faces (for which our vocabulary appears to be rather limited), while others contained descriptions of body and clothing characteristics, as well as estimates of height, weight, and age. While these aspects of descriptions may add to the quantity of the numbers of features mentioned, they may also obscure the correlations between description accuracy and a face identification task (see Davids et al., 2006; Sauerland et al., 2008 this issue; Sporer, 1996). Furthermore, person descriptions frequently contain subjective impressions of the targets described (e.g., "attractive", "aggressive") that refer to inferred (personality) characteristics for which accuracy cannot be scored (see Sporer, 1996), and most studies appear to have omitted such aspects from their feature calculations. Researchers should be careful to report what proportions of the descriptions provided refer to such characteristics and how these were considered/eliminated in the description measures used for reporting description-identification correlations.

In general, we cannot expect to find substantive correlations between two variables if one, or both, of the variables is ill-defined and not measured with objectivity. Similarly, such correlations are likely to be attenuated if one of the variables refers to various aspects of a person (e.g., body and character descriptions) while identification is only measured via a photospread or a facial recognition task. Future studies should clearly specify the instructions given to participants and carefully operationalise the various aspects of person descriptions measured. For each characteristic of a description, evidence of high interrater reliability needs to be established. To the extent that previous studies have not followed these guidelines they may have underestimated the true relationships between these variables.

Practical implications

In *Neil v. Biggers* (1972), the US Supreme Court rendered the assertion that a witness's description could be used as a basis from which to evaluate the veracity of his/her memory and subsequent identification of the suspect. The current paper has provided the first opportunity to evaluate the Court's ruling regarding the description-identification relationship via a meta-analysis of the available literature. Across 33 research papers and a total of 4278 participants, our analysis found support for a significant relationship between description accuracy and identification accuracy, as well as a relationship between the number of incorrect descriptors recalled and identification (in)accuracy. While this would seem to provide support for the Court's assertions, these relationships were somewhat small and it is important to consider the conditions under which the strongest effects were identified.

In particular, results of a moderator analysis suggested that the relationship between description accuracy and identification accuracy was *strongest* in studies that employed facial recognition paradigms involving the presentation of multiple target faces in the context of photographic materials and involving a short delay between encoding and description, and *weakest* in studies that utilised eyewitness identification paradigms that focused encoding and identification on a single target face presented in a more realistic manner (via a videotaped or live event). As noted earlier, one interpretation of this effect is that descriptions generated in multiple-face paradigms help to preserve memory against interference effects and provide individuals an opportunity to generate elaborate, individuated encodings regarding each stimulus face. In fact, this is quite consistent with studies of verbal facilitation reviewed earlier (e.g., Brown & Lloyd-Jones, 2005, 2006) and the "trial effect" observed in studies of verbal overshadowing (see Schooler et al., 1996). Alternatively, it may be that multiple faces and/or trials are required to vary description quality and thereby supersede the effects of range restriction inherent in description performance or, as Wells (1985) has suggested previously, that stimulus variability in the distinctiveness of faces serves to increase this variance in description quality and is associated with both ease/difficulty of description and identification (see also Sporer, 1989). Regardless of one's interpretation of this effect, however, it is clear that the courts may have considerable difficulty attempting to utilise a witness's single description of the perpetrator as a basis to judge the likely veracity of his/her identification.

Of course, the central issue of concern in transferring these research findings from the laboratory to the field involves that of determining *ground truth* with respect to accuracy. That is, how do we determine in practice that a witness has accurately described the perpetrator of the crime without assuming (sometimes falsely) that the suspect identified by the witness is, in fact, the perpetrator? As Wells (1985) has noted, what the courts mistakenly refer to as accuracy is rather the *congruence* between a witness's description and the individual they identified from the lineup. While only a handful of studies have examined this measure of description quality, the results suggest that it demonstrates a small, but significant, relationship with identification accuracy; and, more importantly, two of these studies suggest that measures of congruence show a rather strong correlation with that of description accuracy (of the actual target face). This relationship between description accuracy, congruence, and identification accuracy merits further research in our estimation—particularly given its relevance to the court's desire to evaluate the description quality of witnesses at trial. Ideally, these studies should examine targets differing in distinctiveness (and race), as well as vary encoding times and ecologically valid retention intervals (weeks or months).

One measure of a description that the courts could directly rely upon involves that of *description quantity*. Unfortunately, the current meta-analysis suggested that only a small, marginally significant, relationship existed between description quantity and identification accuracy, and this effect is likely counter to the court's assumptions regarding what constitutes description quality—as more complete descriptions were associated with a greater likelihood of *inaccurate* identification. Furthermore, certain conditions only exacerbated this negative relationship, including the use of more contrived face recognition paradigms, shorter encoding times, and lengthier delays between description and identification tasks. Finally, quantity of the description was significantly associated with description accuracy across studies, but again this relationship is likely counter to the court's assumption regarding description quality as more complete descriptions tended to *less accurately* describe the target face. Taken together, the court should be careful of how it uses description quantity as an estimate of the quality of a witness's memory for the perpetrator.

CONCLUSIONS

Verbal descriptions can lead to both negative (verbal overshadowing) and positive (verbal facilitation) effects on subsequent identification accuracy. Understanding the relationship between verbal description and perceptual identification processes is important to appreciating the factors that lead to these two effects. The current meta-analysis of the description-identification relationship in memory for faces suggests that performance on these two tasks is related to a certain degree, particularly in measures of description accuracy and number of incorrect details recalled. Furthermore, moderator analyses suggested that this effect is strongest in facial recognition studies that employ larger stimulus sets and provide more opportunities for assessing the relationship within and across both participants and faces. Delays between the encoding, description, and identification phases also appeared to moderate this relationship. Taken together, the findings provide some support for a retrieval-based account of the effects of verbal description on subsequent identification performance, though they do not preclude the role of featural versus configural processing (transfer inappropriate processing). The courts have suggested that a witness's description may be used to assess the veracity of his/her identification of the suspect (*Neil v. Biggers*, 1972); however, the current review questions the application of this assumption and suggests future research that might lead to the use of measures of *congruence* in applied settings.

(*Note*: Studies denoted with an asterisk were included in the meta-analysis.)

REFERENCES

Anderson, J. R., & Reder, L. M. (1979). An elaborative processing explanation of depth of processing. In L. Cermak & F. Craik (Eds.), *Levels of processing in human memory* (pp. 385–403). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Becker, B. J. (1994). Combining significance levels. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 215–230). New York: Russell Sage Foundation.

Begg, C. B. (1994). Publication bias. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 399–409). New York: Russell Sage Foundation.

*Bothwell, R. K. (1985). *The effects of gender and arousal on facial recognition and recall.* Unpublished doctoral dissertation, Florida State University, Tallahassee, FL.

Bothwell, R. K., Deffenbacher, K. A., & Brigham, J. C. (1987). Correlation of eyewitness accuracy and confidence: Optimality hypothesis revisited. *Journal of Applied Psychology*, *72*, 691–695.

*Bothwell, R. K., Trahan, M. A., & Newsome, S. R. (1991). *Eyewitness description accuracy as a function of weapon presence and delay.* Paper presented at the annual meeting of the Southeastern Psychological Association, New Orleans, LA.

Brandimonte, M. A., Hitch, G. J., & Bishop, D. V. M. (1992). Influence of short-term memory codes on visual image processing: Evidence from image transformation tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 157–165.

Brewer, N., & Burke, A. (2002). Effects of testimonial inconsistencies and eyewitness confidence on mock-juror judgments. *Law and Human Behavior*, *26*, 353–364.

*Brigham, J. C., & Pigott, M. (1983). *The relationship between accuracy of prior description and facial recognition*. Paper presented at the American Psychological Association, Anaheim, CA.

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.

Brown, C., & Lloyd-Jones, T. J. (2003). Verbal overshadowing of multiple face and car recognition: Effects of within- versus across-category verbal descriptions. *Applied Cognitive Psychology*, *17*, 183–201.

*Brown, C., & Lloyd-Jones, T. J. (2005). Verbal facilitation of face recognition. *Memory and Cognition*, *33*, 1442–1456.

*Brown, C., & Lloyd-Jones, T. J. (2006). Beneficial effects of verbalization and visual distinctiveness on remembering and knowing faces. *Memory and Cognition*, *34*, 277–286.

Cabeza, R., & Kato, T. (2000). Features are also important: Contributions of featural and configural processing to face recognition. *Psychological Science*, *11*, 429–433.

Chance, J., & Goldstein, A. G. (1976). Recognition of faces and verbal labels. *Bulletin of the Psychonomic Society*, *7*, 384–386.

Clare, J., & Lewandowsky, S. (2004). Verbalizing facial memory: Criterion effects in verbal overshadowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 739–755.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cooper, H., & Hedges, L. V. (1994). *The handbook of research synthesis.* New York: Russell Sage Foundation.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671–684.

Cutler, B. L., Penrod, S. D., & Martens, T. K. (1987). The reliability of eyewitness identification: The role of system and estimator variables. *Law and Human Behavior*, *11*, 233–258.

*Davids, M. S. C., Sporer, S. L., & McQuiston-Surrett, D. (2006). *The verbal facilitation effect: Person descriptions as a means to improve identification performance*. Manuscript submitted for publication.

Deffenbacher, K. A. (1983). The influence of arousal on reliability of testimony. In S. M. A. Lloyd-Bostock & B. R. Clifford (Eds.), *Evaluating witness evidence* (pp. 235–251). Chichester, UK: Wiley.

Deffenbacher, K. A., Bornstein, B. H., Penrod, S. D., & McGorty, E. K. (2004). A meta-analytic review of the effects of high stress on eyewitness memory. *Law and Human Behavior*, *28*, 687–706.

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

Dore, H. S., Brigham, J. C., Moussallie, N. M., Bennett, L. B., & Butz, D. A. (2005). *What did he look like? Is there an "other-race effect" in description accuracy?* Unpublished manuscript.

Ellis, H. D., Deregowski, J. B., & Shepherd, J. W. (1975). Descriptions of white and black faces by white and black subjects. *International Journal of Psychology*, *10*, 119–123.

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

*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 verbal description on face identification accuracy: "Release from verbal overshadowing". *Journal of Applied Psychology*, *84*, 340–348.

Fiore, S. M., & Schooler, J. W. (2002). How did you get here from there? Verbal overshadowing of spatial mental models. *Applied Cognitive Psychology*, *16*, 897–909.

Flexser, A., & Tulving, E. (1978). Retrieval independence in recognition and recall. *Psychological Review*, *85*(3), 153–171.

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory and Cognition*, *16*, 309–313.

*Geiselman, R. E., Schroppel, T., Tubridy, A., Konisi, T., & Rodriguez, V. (2000). Objectivity bias in eyewitness performance. *Applied Cognitive Psychology*, *14*, 323–332.

*Goldstein, A. G., Johnson, K. S., & Chance, J. E. (1979). Does fluency of face description imply superior face recognition? *Bulletin of the Psychonomic Society*, *13*, 15–18.

*Grass, E., & Sporer, S. L. (1991, March). *Richtig oder falsch? Zur Vorhersage von Identifizierungsleistungen durch weitere Aussagen von Zeugen* [Correct or false? Post-dicting eyewitness identification accuracy from verbal statements]. Paper presented at the 33rd Tagung experimentell arbeitender Psychologen, Giessen, Germany.

*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.

Haas, S., & Sporer, S. L. (1989, April). *Zur Entwicklung von Enkodierungsstrategien beim Wiedererkennen von Gesichtern* [On the development of encoding strategies in the recognition of human faces]. Paper presented at the 31st Tagung experimentell arbeitender Psychologen, Bamberg, Germany.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hintzman, D., & Hatry, A. L. (1990). Item effects in recognition and fragment completion: Contingency relations vary for different subsets of words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 965–969.

*Hosch, H. M., & Bothwell, R. K. (1990). Arousal, description and identification accuracy of victims and bystanders. *Journal of Social Behavior and Personality*, *5*, 481–488.

Humphreys, M. S., & Bowyer, P. A. (1980). Sequential testing effects and the relation between recognition and recognition failure. *Memory and Cognition*, *8*, 271–277.

Itoh, Y. (2005). The facilitating effect of verbalization on the recognition memory of incidentally learned faces. *Applied Cognitive Psychology*, *19*, 421–433.

*Jenkins, F., & Davies, G. (1985). Contamination of facial memory through exposure to misleading composite pictures. *Journal of Applied Psychology*, *70*, 164–176.

Kahana, M. J., Hirsuto, D. S., & Schneider, A. R. (2005). Theoretical correlations and measured correlations: Relating recognition and recall in four distributed memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 933–953.

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

Light, R. J., Singer, J. D., & Willett, J. B. (1994). The visual presentation and interpretation of meta-analyses. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 439–453). New York: Russell Sage Foundation.

Lindsay, D., Read, J., & Sharma, K. (1998). Accuracy and confidence in person identification: The relationship is strong when witnessing conditions vary widely. *Psychological Science*, *9*(3), 215–218.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

Lovett, S. B., Small, M. Y., & Engstrom, S. A. (1992, November). *The verbal overshadowing effect: Now you see it, now you don't*. Paper presented at the annual meeting of the Psychonomic Society, St Louis, MO.

Mauldin, M. A., & Laughery, K. R. (1981). Composite production effects on subsequent facial recognition. *Journal of Applied Psychology*, *66*, 351–357.

McKelvie, S. J. (1976). The effects of verbal labeling on recognition memory for schematic faces. *Quarterly Journal of Experimental Psychology*, *28*, 459–474.

McQuiston-Surrett, D., & Topp, L. D. (2004, March). *Cross-race face recall: Assessing perceptual expertise in a description and facial composite task*. Paper presented at the annual conference of the American Psychology–Law Society, Scottsdale, AZ.

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

Meissner, C. A., & Brigham, J. C. (2001b). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, *7*, 3–35.

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

Meissner, C. A., Sporer, S. L., & Schooler, J. W. (2007). Person descriptions as eyewitness evidence. In R. Lindsay, D. Ross, J. Read, & M. Toglia (Eds.), *Handbook of eyewitness psychology: Memory for people* (pp. 3–34). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Melcher, J. M., & Schooler, J. W. (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.

*Memon, A., & Bartlett, J. (2002). The effects of verbalization on face recognition in young and older adults. *Applied Cognitive Psychology*, *16*, 635–650.

*Memon, A., & Rose, R. (2002). Identification abilities of children: Does a verbal description hurt face recognition? *Psychology, Crime, and Law*, *8*, 229–242.

*Memon, A., Rose, R., Searcy, J., & Bartlett, J. C. (1999). *The verbal overshadowing effect and eyewitness identification: A developmental and aging perspective.* Unpublished manuscript.

Meurer, D., Sporer, S. L., & Rennig, C. (1990). Der Beweiswert von Personenidentifizierungen: Auf dem Weg von alltagspsychologischen Erfahrungssaetzen zu empirisch ueberpruefbaren Fragestellungen [The probative value of eyewitness identifications: From common sense psychology to empirically testable hypotheses]. In D. Meurer & S. L. Sporer (Eds.), *Zum Beweiswert von Personenidentifizierungen* (pp. 1–18). Marburg, Germany: N. G. Elwert.

Mitchell, T. L., Meissner, C. A., & MacLin, O. H. (2006). *The influence of the cross-race effect on lineup construction and fairness.* Manuscript submitted for publication.

Neil v. Biggers, 409 US 188 (1972).

Odenthal, H. J. (1999). *Die Gegenüberstellung im Strafverfahren* [The lineup in criminal procedure] (3rd ed.). Stuttgart, Germany: Boorberg.

Paivio, A. (1971). *Imagery and verbal processes.* New York: Holt.

*Pigott, M. A., & Brigham, J. C. (1985). Relationship between accuracy of prior description and facial recognition. *Journal of Applied Psychology*, *70*, 547–555.

*Pigott, M. A., Brigham, J. C., & Bothwell, R. K. (1990). A field study on the relationship between quality of eyewitnesses' descriptions and identification accuracy. *Journal of Police Science and Administration*, *17*, 84–88.

*Pozzulo, J. D., & Warren, K. L. (2003). Descriptions and identification of strangers by youth and adult eyewitnesses. *Journal of Applied Psychology*, *88*, 315–323.

Read, J. D. (1979). Rehearsal and recognition of human faces. *American Journal of Psychology*, *92*, 71–85.

Roediger, H. L., & Guynn, M. J. (1996). Retrieval processes. In E. L. Bjork & R. A. Bjork (Eds.), *Memory* (pp. 197–231). San Diego, CA: Academic Press.

Roediger, H. L., Wheeler, M. A., & Rajaram, S. (1993). Remembering, knowing, and reconstructing the past. In D. L. Medin (Ed.), *Advances in research and theory: Vol. 37. The psychology of learning and motivation* (pp. 97–134). San Diego, CA: Academic Press.

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Newbury Park, CA: Sage.

Ryan, R. S., & Schooler, J. W. (1998). Whom do words hurt? Individual differences in susceptibility to verbal overshadowing. *Applied Cognitive Psychology*, *12*, 105–126.

*Sauerland, M., Holub, F. E., & Sporer, S. L. (2008). Person descriptions and person identifications: Verbal overshadowing or recognition criterion shift? *European Journal of Cognitive Psychology*, *20*, 496–528.

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. In D. Medin (Ed.), *Handbook of learning and motivation* (Vol. 37, pp. 291–340). Orlando, FL: Academic Press.

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

*Searcy, J. H., Bartlett, J. C., Memon, A., & Swanson, K. (2001). Aging and lineup performance at long retention intervals: Effects of metamemory and context reinstatement. *Journal of Applied Psychology*, *86*, 207–214.

Shepherd, J. W., & Deregowski, J. B. (1981). Races and faces: A comparison of the responses of Africans and Europeans to faces of the same and different races. *British Journal of Social Psychology*, *20*, 125–133.

Smith, S. M. (1988). Environmental context-dependent memory. In G. Davies & D. Thomson (Eds.), *Memory in context: Context in memory* (pp. 13–34). Chichester, UK: Wiley.

Sporer, S. L. (1988). Long-term improvement of facial recognition through visual rehearsal. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 182–188). London: Wiley.

Sporer, S. L. (1989). Verbal and visual processes in person identification. In H. Wegener, F. Loesel, & J. Haisch (Eds.), *Criminal behavior and the justice system: Psychological perspectives* (pp. 303–324). New York: Springer.

Sporer, S. L. (1991). Deep—deeper—deepest? Encoding strategies and the recognition of human faces. *Journal of Experimental Psychology, Learning, Memory, and Cognition*, *17*, 323–333.

*Sporer, S. L. (1992). Postdicting eyewitness identification accuracy: Confidence, decision-times and person descriptions among choosers and non-choosers. *European Journal of Social Psychology*, *22*, 157–180.

Sporer, S. L. (1996). Describing others: Psychological issues. In S. L. Sporer, R. S. Malpass, & G. Koehnken (Eds.), *Psychological issues in eyewitness identification* (pp. 53–86). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Sporer, S. (2001). Recognizing faces of other ethnic groups: An integration of theories. *Psychology, Public Policy and Law*, *7*, 36–97.

*Sporer, S. L. (2007). Person descriptions as retrieval cues: Do they really help? *Psychology, Crime, and Law*, *13*, 591–609.

Sporer, S. L., & Cutler, B. L. (2003). Identification evidence in Germany: Common sense assumptions, empirical evidence, guidelines, and judicial practices. In P. J. van Koppen & S. D. Penrod (Eds.), *Adversarial vs. inquisitorial justice: Psychological perspectives on criminal justice systems* (pp. 191–208). New York: Plenum.

Sporer, S. L., Penrod, S., Read, D., & Cutler, B. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin*, *118*, 315–327.

Tredoux, C. G., Meissner, C. A., Malpass, R. S., & Zimmerman, L. A. (2004). Eyewitness identification. In C. Spielberger (Ed.), *Encyclopedia of applied psychology* (pp. 875–887). San Diego, CA: Academic Press.

Tulving, E. (1985). How many memory systems are there? *The American Psychologist*, *40*, 385–398.

Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science*, *247*, 301–306.

*Wegener, H. (1966). *Experimentelle Untersuchungen ueber die Personenbeschreibung und identifizierung 10- bis 12-jaehriger Kinder* [Experimental investigations on person descriptions and identifications of 10- to 12-year-old children]. In J. Gerchow (Ed.), *An den Grenzen von Medizin und Recht* (pp. 115–125). Stuttgart, Germany: Enke.

Wells, G., & Hryciw, B. (1984). Memory for faces: Encoding and retrieval operations. *Memory and Cognition*, *12*(4), 338–344.

*Wells, G. L. (1985). Verbal descriptions of faces from memory: Are they diagnostic of identification accuracy. *Journal of Applied Psychology*, *70*, 619–626.

Wells, G. L., & Olson, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology*, *54*, 277–295.

Wells, G. L., & Turtle, J. W. (1988). What is the best way to encode faces? In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues* (pp. 163–168). New York: Wiley.

Westerman, D. L., & Larson, J. D. (1997). Verbal overshadowing effect: Evidence for a general shift in processing. *American Journal of Psychology*, *110*, 417–428.

Wickham, L. H. V., & Swift, H. (2006). Articulatory suppression attenuates the verbal overshadowing effect: A role for verbal encoding in face identification. *Applied Cognitive Psychology*, *20*, 157–169.

Winograd, E. (1981). Elaboration and distinctiveness in memory for faces. *Journal of Experimental Psychology: Human Learning and Memory*, *7*, 181–190.

Wogalter, M. S. (1991). Effects of post-exposure description and imaging on subsequent face recognition performance. *Proceedings of the Human Factors Society*, *35*, 575–579.

*Wogalter, M. S. (1996). Describing faces from memory: Accuracy and effects on subsequent recognition performance. *Proceedings of the Human Factors and Ergonomics Society*, *40*, 536–540.

Woodhead, M., & Baddeley, A. (1981). Individual differences and memory for faces, pictures, and words. *Memory and Cognition*, *9*(4), 368–370.

*Yarmey, A. D. (1986). Verbal, visual, and voice identification of a rape suspect under different levels of illumination. *Journal of Applied Psychology*, *71*, 363–370.

*Yarmey, A. D. (2004). Eyewitness recall and photo identification: A field experiment. *Psychology, Crime and Law*, *10*, 53–68.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*, 441–517.

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

APPENDICES

TABLE A1
Correlations (effect size *r*) between measures of description quality and identification accuracy across studies

Study ID	*Authors*	*Year*	*Experiment*	*N*	*Accuracy*	*Quantity*	*Correct descriptors*	*Incorrect descriptors*	*Congruence*
1	Wegener	1966	Boys	63	.03	na	na	na	na
2	Wegener	1966	Girls	32	.12	na	na	na	na
3	Goldstein et al.	1979		22	.15	na	na	na	na
4	Brigham & Pigott	1983	Lineup A	32	na	na	na	na	.01
5	Brigham & Pigott	1983	Lineup B	36	na	na	na	na	−.25
6	Bothwell	1985		128	.08	na	na	na	na
7	Pigott & Brigham	1985		120	.00[a]	na	na	na	.16[b]
8	Jenkins & Davies	1985	2—Control	29	−.16	na	na	na	na
9	Wells	1985	Choosers	127	.27	.05	na	na	.19
10	Yarmey	1986		128	.00[c]	na	na	na	na
11	Hosch & Bothwell	1990	2	42	na	na	.31	na	na
12	Schooler & Engstler-Schooler	1990	1	35	−.01	.11	−.01	na	na
13	Schooler & Engstler-Schooler	1990	2	40	.23	−.14	−.15	na	na
14	Schooler & Engstler-Schooler	1990	4	36	.23	−.26	.04	na	na
15	Pigott et al.	1990		47	−.16	.09	na	na	.25[d]
16	Bothwell et al	1991	1	31	.02[f]	na	na	na	na
17	Bothwell et al	1991	2	70	−.22[f]	na	na	na	na
18	Grass & Sporer	1991		79	na	−.06	−.05	na	na
19	Sporer	1992		49	na	.28	na	na	na

Table A1 (*Continued*)

Study ID	*Authors*	*Year*	*Experiment*	*N*	*Accuracy*	*Quantity*	*Correct descriptors*	*Incorrect descriptors*	*Congruence*
20	Wogalter	1996		48	.38	na	na	na	na
21	Wogalter	1996		48	.35	na	na	na	na
22	Gwyer & Clifford	1997	TP	70	.00[c]	.00[c]	.26	.00[c]	na
23	Gwyer & Clifford	1997	TA	70	.00[c]	−.28	.00[c]	.00[c]	na
24	Finger & Pezdek	1999	1	75	na	na	−.22	−.32	na
25	Finger & Pezdek	1999	2	69	na	na	.16	.00	na
26	Memon et al.	1999	2	60	.19	na	−.33	na	na
27	Geiselman et al.	2000	2—TP	99	−.05	.02	−.03	.11	na
28	Geiselman et al.	2000	2—TA	61	.06	.27	.29	.09	na
29	Meissner et al.	2001	1	180	.34	−.23	−.15	−.29	na
30	Meissner et al.	2001	2	60	.38	−.08	−.03	−.38	na
31	Searcy et al.	2001	Young—TP	23	na	.69	na	na	na
32	Searcy et al.	2001	Old—TP	25	na	.48	na	na	na
33	Searcy et al.	2001	Young—TA	22	na	.15	na	na	na
34	Searcy et al.	2001	Old—TA	24	na	−.61	na	na	na
35	Finger	2002	1	89	na	.07	.19	−.18	na
36	Finger	2002	2	73	na	−.23	−.13	−.16	na
37	Kitagami et al.	2002		110	na	.04	−.03	.02	na
38	Meissner	2002	1	432	.27	−.18	−.07	−.25	na
39	Meissner	2002	2	108	.27	−.30	−.16	−.37	na
40	Memon & Bartlett	2002	Young	34	.00	na	na	na	na
41	Memon & Bartlett	2002	Senior	36	.14	na	na	na	na
42	Memon & Rose	2002		25	−.02	na	na	na	na
43	Pozzulo & Warren	2003	1—TP	76	na	−.07	na	na	na
44	Pozzulo & Warren	2003	1—TA	74	na	.15	na	na	na
45	Pozzulo & Warren	2003	2—TP	85	na	.24	na	na	na
46	Pozzulo & Warren	2003	2—TA	86	na	.08	na	na	na

(*Continued overleaf*)

Table A1 (*Continued*)

Study ID	*Authors*	*Year*	*Experiment*	*N*	*Accuracy*	*Quantity*	*Correct descriptors*	*Incorrect descriptors*	*Congruence*
47	Yarmey	2004		590	.09	na	na	na	na
48	Sauerland et al.	2008 this issue	Description	48	.39	−.18	.07	−.46	na
49	Brown & Lloyd-Jones	in press-a		84	na	.02	na	na	na
50	Brown & Lloyd-Jones	in press-b	1	56	na	−.16	na	na	na
51	Brown & Lloyd-Jones	in press-b	4	89	na	.21	na	na	na
52	Sporer	in press	Description	25	na	.31	na	na	na
53	Davids et al.	2006	Description, Target 1	48	.10	−.02	.08	−.09	na
54	Davids et al.	2006	Description, Target 2	48	.30	.12	.25	−.18	na

Correlations with identification accuracy are point-biserial correlations. All other correlations are Pearson product-moment correlations.

[a]Averaged across choosers and nonchoosers as well as other conditions. [b]Based on $N = 37$ only. [c]Estimated as .00 from report of being "*ns*". [d]Based on $N = 25$ only. [e]Based on $N = 24$ only. [f]No weapon condition only. [g]Rereading condition was not considered here.

TABLE A2
Coding of study characteristics

Study ID	*Realism of encoding event*	*Description task*	*Encoding time (s)*	*No. of faces encoded*	*Encoding-description delay (min)*	*Description-identification delay (min)*	*Publication status*
1	Live event	Recall & checklist	30	1	0	5	Journal article
2	Live event	Recall & checklist	30	1	0	5	Journal article
3	Photograph	Checklist	2.5	10	0	0	Journal article
4	Live event	Checklist	15	1	0	5	Unpublished
5	Live event	Checklist	15	1	0	5	Unpublished
6	Live event	Cued recall	15	1	0	20	Journal article

Table A2 (*Continued*)

Study ID	*Realism of encoding event*	*Description task*	*Encoding time (s)*	*No. of faces encoded*	*Encoding-description delay (min)*	*Description-identification delay (min)*	*Publication status*
7	Live event	Checklist	15	1	0	5	Journal article
8	Video	Checklist	50	1	15	20, 2880, & 10,080	Journal article
9	Photograph	Recall	240	1	10	10	Journal article
10	Photograph	Recall	120	1	0	0	Journal article
11	Live event	Recall	20	1	10	0	Conference Presentation
12	Video	Recall	30	1	20	0	Journal article
13	Video	Recall	30	1	20	0	Journal article
14	Video	Recall	30	1	0	10	Journal article
15	Live event	Recall	90	1	270	0	Journal article
16	Live event	Recall	15	1	20	0	Journal article
17	Live event	Recall	15	1	20	0	Journal article
18	Live event	Recall	70	1	10,080	0	Conference Presentation
19	Live event	Recall	20	1	10,080	0	Journal article
20	Photograph	Recall & checklist	5	6	0	5	Journal article
21	Photograph	Recall	5	6	0	5	Journal article
22	Live event	Recall	3	1	30	0	Journal article
23	Live event	Recall	3	1	30	0	Journal article
24	Photograph	Recall	240	1	10	10	Journal article
25	Photograph	Recall	240	1	10	60	Journal article
26	Video	Recall	90	1	30	0	Journal article
27	Video	Recall	5	1	0	0	Journal article
28	Video	Recall	5	1	0	0	Journal article
29	Photograph	Recall	10	1	5	0 & 30	Journal article
30	Photograph	Recall	10	1	5	0	Journal article
31	Live event	Recall	1200	1	43,200	0	Journal article
32	Live event	Recall	1200	1	43,200	0	Journal article
33	Live event	Recall	1200	1	43,200	0	Journal article

(*Continued overleaf*)

Table A2 (*Continued*)

Study ID	*Realism of encoding event*	*Description task*	*Encoding time (s)*	*No. of faces encoded*	*Encoding-description delay (min)*	*Description-identification delay (min)*	*Publication status*
34	Live event	Recall	1200	1	43,200	0	Journal article
35	Photograph	Recall	30	1	5	5	Journal article
36	Photograph	Recall	30	1	5	5	Journal article
37	Photograph	Recall	10	1	5	0	Journal article
38	Photograph	Recall	5	1	15	5	Journal article
39	Photograph	Recall	5	1	15	5 & 10,080	Journal article
40	Video	Recall	60	1	45	2	Journal article
41	Video	Recall	60	1	45	2	Journal article
42	Live event	Recall	480	1	1440	0	Journal article
43	Video	Recall	500	1	0	5	Journal article
44	Video	Recall	500	1	0	5	Journal article
45	Live event	Recall	500	1	0	5	Journal article
46	Live event	Recall	500	1	0	5	Journal article
47	Live event	Cued recall	15	1	2240	0	Journal article
48	Video	Recall & cued recall	18	1	0	2880	Conference Presentation
49	Photograph	Recall	5	24	0	0	Journal article
50	Photograph	Recall	2	23	0	0	Journal article
51	Photograph	Recall	2	23	0	0	Journal article
52	Video	Recall & checklist	18	1	0	10,080	Journal article
53	Video	Recall & cued recall	42	2	0	2880	Unpublished
54	Video	Recall & cued recall	70	2	0	2880	Unpublished