Externalizing Visual Images

Examining the Accuracy of Facial Descriptions vs. Composites as a Function of the Own-Race Bias

Dawn McQuiston-Surrett\textsuperscript{1} and Lisa D. Topp\textsuperscript{2}

\textsuperscript{1}Arizona State University, Glendale, AZ, \textsuperscript{2}University of Texas at El Paso, both USA

\textbf{Abstract.} Two studies investigated whether the recall accuracy of facial details would differ depending on recall method, and also explored the own-race/ethnic bias in face recall. In Experiment 1, Hispanic participants (\(N = 120\)) viewed either a Hispanic or a White face and then were asked to recall the face by either giving a physical description or constructing a facial composite. Independent judges then determined the accuracy of the recalled information in both a rating task and a matching task. Results revealed reliably higher accuracy scores for the descriptions over the facial composites but showed no evidence for an own-race bias. A second experiment (\(N = 120\)) was conducted to attempt to replicate the description-advantage effect and to further explore the own-race bias in a stronger test using Black faces as the cross-ethnic group. This experiment again showed a description advantage and provided some evidence for an own-race bias in recall similar to that found in the facial recognition literature. Directions for future research based on the current findings are discussed.

\textbf{Keywords:} facial descriptions, facial composites, own-race bias

Obtaining physical information about a culprit from an eyewitness is an initial step in an investigation, generally forming the basis for further police work by narrowing the search for a suspect. Based on this, police may inspect mugshots for individuals who match the description or have been arrested for a similar crime or in a similar area. If the eyewitness is unaware of the culprit’s identity and the police do not have a suspect in mind, witnesses are often asked to assist in creating a visual representation of the culprit. An important empirical question concerning these recall procedures is how effective current verbal and visual recall methods are in assisting police. Information should be obtained from witnesses using procedures that maximize the likelihood that it accurately conveys the culprit’s appearance. How well do people perform these tasks and how successful are current procedures at helping represent the intended individual?

Several description elicitation techniques have been developed, ranging from free recall to cued description methods (Meissner, Sporer, & Schooler, 2006). Descriptions can be quite accurate but are often incomplete or vague (Sporer, 1996) and methods that involve cued recall of information are generally advantageous (Meissner & Brigham, 2001a). Concerning facial composite production, many procedures are available to law enforcement, including sketch artistry and manual (e.g., IdentiKit, Photofit) and computerized systems (e.g., Faces, PROfit). Using the actual systems which are quite popular among police (e.g., McQuiston-Surrett et al., 2006), witnesses select individual features from sets of pictorial images and assemble them into a likeness of the culprit’s face.

Sporer (1996) has described the process by which physical information is transferred from a witness to an investigator as a two-pronged approach: The witness translates the visual image of the culprit into words understandable by the investigator, who attempts to form a visual impression of the face. There are many practical challenges, including lacking the necessary vocabulary to communicate information accurately (Ellis, 1984; Sporer, 1988; Wogalter, 1991), the witness and interviewer not sharing an understanding of particular descriptors (Davies, 1986), and limitations in composite system design or operator skill (Shepherd & Ellis, 1996). More conceptually, maintaining the visual image in memory while decomposing it to its constituent features for recall is said to be difficult and may contaminate the original memory (Ellis, 1984, 1986; Ellis, Shepherd, & Davies, 1975), perhaps because faces are said to be processed holistically which is in contrast to the featural nature of recall tasks (Sergent, 1988).

Given these limitations, one purpose of this research is to assess the accuracy of eyewitness recall in the form of descriptions versus composites. A handful of early studies examined the efficacy of these procedures. For example, Christie and Ellis’s (1981) participants generated a description and a Photofit construction of a face, after which judges matched the information to the correct face from a photoarray, and sorted the information to the correct face amongst a set of faces. Results indicated greater accuracy of the descriptions on both assessment tasks. Flin, Markham, and Davies’ (1989) results among a sample of children paralleled Christie and Ellis’s (1981). These data are not surprising given the data...
concerning the general inadequacy of composites, however, a theoretical position also supports these results. Constructing a composite while simultaneously holding a visual image in memory may produce modality specific interference since these two processes work through the visual modality, thereby interfering with the witness’s ability to adequately judge the accuracy of the likeness (Brooks, 1968; Christie & Ellis, 1981; Shepherd & Ellis, 1996). In contrast, generating a description, which is a verbal process, may not be subject to this interference and presumably not disrupt the memory, thus resulting in greater accuracy for information produced via the verbal channel than that portrayed by composite. The results of these studies suggest that police should place greater emphasis on the description of a culprit than a composite, but it could be argued that existing research does not wholly inform current practices since outdated composite systems were utilized and the results have not been replicated using either age group of participants. On a more conceptual note, little is known about how well people can reproduce facial details using verbal versus visual methods. Thus, one purpose of this research was to investigate the efficacy of face recall methods using current practices.

A related point understudied in the literature but theoretically relevant is whether people are susceptible to the own-race bias (ORB) in face recall. Research shows the ORB in recognition – that individuals are better at recognizing faces of their own ethnic group than less familiar groups (Meissner & Brigham, 2001b) – but whether there is an ORB in face recall is not clear. The process of recall differs from that of face recognition; indeed, people generally perform better on recognition than recall tasks (Baddeley, 1998; Davies, Shepherd, & Ellis, 1978; Goldstein, Johnson, & Chance, 1979; Wells & Hryciw, 1984) and the relationship between the two is weak (e.g., Pigott & Brigham, 1985). Few studies have examined the ORB in recall and some are limited in important ways. Deregowski, Ellis, and Shepherd (1975) found that Blacks provided more details of Black and White faces than did Whites, and Black faces elicited more descriptors from Whites and Blacks than did White faces, but description accuracy was not reported. In investigating Whites’ recall in describing Black and White faces, Fallshore and Schooeler (1995) found no cross-ethnic differences. In investigating facial composites, Ellis, Davies, and McMurrain (1979) showed that Whites and Blacks produced better quality composites of White faces than Black faces, however limitations inherent to the composite system used may have confounded their results; composites of any other ethnicity than White are difficult to construct. Taken together, we currently know very little about the accuracy or underlying mechanisms of cross-race recall, suggesting the need for additional research. This reflects the second purpose of the current studies.

### Experiment 1

Experiment 1 manipulated participants’ recall method (facial description or composite) and target ethnicity (own- or cross-ethnicity). Hispanics, the fastest-growing ethnic minority in the United States (United States Census Bureau, 2000), remain largely unstudied in experimental research in general (see Dovidio, Gaertner, Kawakami, & Hodson, 2002), and in psycho-legal research and facial identification studies in particular, thus Hispanics comprised the majority of participants. A main effect of recall was predicted such that descriptions would yield more accurate information than composites. Experiment 1 also examined whether an ORB would emerge in the recall of facial information, but since there is little foundational research on this topic no specific prediction was made.

### Method

This study consisted of two stages: (1) face recall by description or composite, and (2) evaluation of their respective accuracy. In the recall stage, participants were shown a face and then either gave a description or constructed a composite of it. In the evaluation stage, judges determined the accuracy of the recalled information in two tasks.

#### Face Recall Stage

**Participants and Design**

Hispanic participants (N = 120, 63% female, M = 20 years) were recruited to serve as witnesses. A 2 (Recall Method: Description vs. Facial Composite) × 2 (Target Ethnicity: Hispanic vs. White) between-groups design assessed the impact of recall method and target ethnicity on recall accuracy.

**Materials**

**Stimulus Faces.** Twenty head and shoulder color photographs of males (ten Hispanic and ten White) who appeared to be in their twenties were used as the targets (this group was chosen since young males are known to commit the most crimes). The faces represented a range of distinctiveness based on judges’ ratings. None of the targets wore glasses or had distinguishing facial characteristics as determined by independent raters.

**Recall Methods.** Participants in the description condition gave a description of the face they viewed using a cued

---

1 Distinctive faces are better recognized than nondistinctive faces, while it is possible that more typical faces will lead to better quality composites since they are more average (Bruce, 1990). Therefore, a range of distinctive faces was included in order to allow this factor to vary. To achieve this, the photographs in each of our databases were initially rated by judges (whose ethnicity corresponded with the ethnicity of the photos they were asked to rate) on a 7-point distinctiveness rating scale. Based on the ratings, researchers then randomly selected 10 photographs from each database that represented a range of values spanning the rating scale.
description method in which they were prompted to provide details about ten facial characteristics. Those in the composite conditions constructed a composite of the face they viewed using Faces 3.0 software (Cote, 1998), a system widely used by police (McQuiston-Surrett et al., 2006).

**Questionnaire.** There is no information in the literature regarding people’s subjective experience constructing a composite, therefore participants in the composite condition answered a three-item questionnaire following completion of their composite.

**Procedure**

Participants completed the experiment in groups of one to 20. Following informed consent, participants were instructed they would view a photograph of an individual for 60 sec. and that, later, they would be asked questions about what they viewed. Participants were randomly assigned to view either a White or Hispanic target face for 60 s. Following a 20 min filler task (crossword puzzle), participants either gave a written description or created a composite of the face they viewed, also counterbalanced across conditions. Participants constructing a composite did so following an 8-min tutorial, needing little assistance from the experimenter. These participants then filled out a questionnaire addressing aspects of the facial composite process. All participants were debriefed and dismissed.

**Evaluation Stage**

The accuracy of the descriptions and composites was assessed in a rating task and matching task. The rating task was chosen as a global measure of recall accuracy. The matching task was implemented as a measure of “investigatory usefulness”; since the goal of recalled information is to narrow the pool of suspects, how well this information can be correctly matched to the intended face by nonwitnesses represents an important measure of recall accuracy.

**Rating Task**

Forty students (80% Hispanic, 15% White, 3% Black, 2% Asian; 58% female; M = 20 years) served as rating judges. This task assessed how well the recalled information accurately represented the target face. Judges were run in groups from one to 20, each completing a rating task on two separate occasions, two days apart. In order to avoid fatigue, judges rated half the descriptions and half the composites in total, and this task was carried out on two separate occasions, two days apart. Ultimately, each description and composite was rated by 20 judges. During each session, judges received a set of either 30 descriptions or 30 composites (half of Hispanics and half of Whites), with the presentation of this information counterbalanced similar to that described above. The task was to match the descriptions/composites to the intended target from the photoarray. The score for each description and composite was the number of times it was matched correctly out of 20 judges.

**Matching Task**

Forty psychology students (65% Hispanic, 23% White, 10% Black, 2% Puerto Rican; 50% female; M = 19) were recruited to serve as matching judges. The purpose of this task was to determine the usefulness of descriptions and composites for finding the intended target. Photographs of 20 moderately similar Hispanic and 20 moderately similar White males (the ten targets randomly interspersed among ten new faces) were assembled in two separate photoarrays for inspection by judges, with each 20-person photoarray mounted on a board. Judges were run in groups from one to six. In order to avoid fatigue, judges rated half the descriptions and half the composites in total, and this task was carried out on two separate occasions, two days apart. Ultimately, each description and composite was rated by 20 judges. During each session, judges viewed either a White or Hispanic target face for 60 s. Following a 20 min filler task (crossword puzzle), participants either gave a written description or created a composite of the face they viewed, also counterbalanced across conditions. Participants constructing a composite did so following an 8-min tutorial, needing little assistance from the experimenter. These participants then filled out a questionnaire addressing aspects of the facial composite process. All participants were debriefed and dismissed.

**Results and Discussion**

An initial set of analyses was run to assess whether judges’ ethnicity was related to evaluations of the own-race and cross-race recalled information. For the rating task, when broken down by judges’ ethnicity, each group’s mean accuracy rating of the Hispanic faces and White faces ranged between M = 6.00–6.63 and M = 6.00–6.70, respectively, with no significant effect of judges’ ethnicity on rating accuracy, F(3, 36) < 1, p > .05. For the matching task, broken down by ethnic group, judges accuracy for the Hispanic and White faces ranged between 10–25% and 23–28%, respectively, with no effect of judges’ ethnicity on matching accuracy, F(3, 36) < 1, p > .05. Thus, the following analyses are collapsed across all judges.

A 2 (Type of Recall: Physical description vs. Facial composite) × 2 (Target Ethnicity: Hispanic vs. White) ANOVA on rating accuracy revealed a significant main effect for Type of Recall, F(1, 116) = 96.95, p < .001; η² = 0.46 (see Table 1). The accuracy of the descriptions (M = 7.04, SD = 0.84) was reliably higher than the accuracy of composites (M = 5.23, SD = 1.15). The main effect of Target Ethnicity was not statistically significant, nor was the interaction, both p values > .05. Next, a 2 (Type of Recall) × 2 (Target Ethnicity) ANOVA on matching accuracy was conducted. The main effect of Type of Recall was statistically signifi-
Table 1. Rating and matching scores for recall method by target ethnicity (Experiments 1 and 2)

<table>
<thead>
<tr>
<th></th>
<th>Experiment 1</th>
<th></th>
<th>Experiment 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Physical descriptions</td>
<td>Facial composites</td>
<td>Physical descriptions</td>
<td>Facial composites</td>
</tr>
<tr>
<td></td>
<td>Own-race (Hispanic)</td>
<td>Cross-race (White)</td>
<td>Own-race (Hispanic)</td>
<td>Cross-race (Black)</td>
</tr>
<tr>
<td>Mean accuracy rating (1–10 scale)</td>
<td>7.12 (0.80)</td>
<td>6.95 (0.87)</td>
<td>5.08 (1.12)</td>
<td>5.38 (1.17)</td>
</tr>
<tr>
<td>Proportion correctly matched</td>
<td>0.28</td>
<td>0.30</td>
<td>0.20</td>
<td>0.18</td>
</tr>
</tbody>
</table>

Note. Standard deviations are in parentheses.

cant, $F(1, 116) = 7.50, p < .01; \eta^2 = 0.06$. The proportion of descriptions correctly matched to the target face in the phototarget (29%) was reliably higher than that of the composites (19%). The main effect of Target Ethnicity was not statistically significant, nor was the interaction, $p$ values > .05. The correlation between the rating and matching tasks was significant, $r = .30, n = 120, p = .001$, indicating a moderate relationship between our measures of recall accuracy.

Participants who constructed a facial composite ($N = 60$) were asked about their experience constructing a composite. When asked To what extent do you think your facial composite is similar to the actual face you viewed?, participants' mean rating was 3.03 ($SD = 1.15$) on a scale where 1 was “not at all similar” and 7 was “extremely similar.” When asked How easy or difficult was it for you to create a facial composite that accurately represents the actual face you viewed?, participants' mean rating on a 7-point scale (1 = extremely easy, 7 = extremely difficult) was 5.28 ($SD = 1.33$), with 75% of participants' answers above the midpoint of the scale. None of these responses were related to the ethnicity of facial composite constructed, $p$ values > .05. Last, participants were asked If you thought the process of creating the composite was difficult, what made it so? with four alternatives (instructions indicated to circle all answers that applied). Sixty-three percent endorsed “searching through all the features interfered with my memory of the face,” 43% endorsed “the facial feature catalog was limited; I could not find the features I was looking for,” 23% endorsed “it was difficult to position individual features where I wanted in the facial space,” and 43% endorsed “I couldn’t remember the face well enough to reconstruct it.”

The results of Experiment 1 confirmed our hypothesis: The descriptions were rated as more accurate and were matched to the correct target face more often than were the composites. These findings support those of early studies reporting an advantage for the use of descriptive information over facial composites in the identification process (e.g., Christie & Ellis, 1981). The finding of a description-advantage is consistent with participants’ below-average rating of their satisfaction with the similarity between the facial composite they created and the original target face ($M = 3.03$ on a 7-point scale), as well as their overall sense of the difficulty of the composite construction task ($M = 5.28$ on a 7-point scale)$^2$.

The typical ORB found in recognition did not emerge in the recall of facial details: The information recalled about own-race faces in both the description and facial composite tasks was not more accurate than that recalled about cross-race faces in both the rating the matching tasks. One possibility is that Hispanic and White populations rely on similar characteristics in facial identification. That is, these groups might find certain facial features informative in distinguishing between faces in the identification of both Hispanic and White faces, but there is little evidence to support this claim (see Brigham, Bennett, Meissner, & Mitchell, 2006). Alternatively, this null finding raised an important question with respect to our conceptualization of race expertise: Did our selection of Whites appropriately qualify as a true cross-race group for Hispanics? Although Hispanics comprise the numerical majority (nearly 80%) in the population in which this research was conducted, it is probable that members of this broad ethnic group have greater exposure to Whites than to any other group. Thus, Hispanics in a Hispanic-majority population might be more susceptible to the ORB with other ethnic minority groups with which they have less experience (Meissner & Brigham, 2001b; see also Wright, Boyd, & Tredoux, 2003).

To investigate this assertion, we asked 81 Hispanic university students to fill out the Social Experience Questionnaire designed to assess the extent of participants’ experience with both Whites and Blacks in business, personal, public, educational, media, and neighborhood settings (Slone, Brigham, & Meissner, 2000). All 16 questions are answered on a 9-point Likert-type scale, with higher numbers indicative of greater contact. Our results showed that for all 16 questions, participants’ reported experience with Blacks was significantly less than that reported with Whites. Interestingly, the highest rating pertaining to contact with Whites was $M = 5.48$ (the highest rating pertaining to contact with Blacks was $M = 3.25$), thus demonstrating that even reported experience with Whites was not great.

These results suggest that Hispanics might demonstrate an ORB in face recall with Blacks but not with Whites. Research by Platz and Hosch (1988) supports this conjecture, in which they examined the ORB among Hispanics,
Whites, and Blacks in a city where Hispanics comprise the majority of the population. They asked convenience store clerks to identify a customer that, 1 h prior, visited the store and paid for merchandise. Their results showed that Hispanic witnesses demonstrated a greater ORB in their recognition of Black customers than White customers. Based on this, using faces of an ethnic group with which Hispanics reportedly have little exposure may be a more valid test of the ORB in recall.

**Experiment 2**

We conducted a second study to assess whether the ORB would emerge when implementing a stronger test using Blacks as the cross-race group, as well as to attempt to replicate the findings of a description-advantage from Experiment 1. The method was identical to Experiment 1. It was hypothesized that (1) Hispanics would display an ORB in the recall of facial information using Blacks as the cross-race, and (2) the description-advantage from Experiment 1 would replicate.

**Method**

**Face Recall Stage**

Participants and Design

Hispanic participants (N = 120) were recruited from psychology classes to serve as witnesses. This sample was 63% female with a mean age of M = 20. A 2 (Recall Method: Description vs. Facial Composite) × 2 (Target Ethnicity: Hispanic vs. Black) between-groups design assessed the impact of recall method and target ethnicity on recall accuracy.

Materials and Procedure

The materials and procedure followed that of Experiment 1, except that Black faces were used instead of White, chosen in an identical manner.

**Evaluation Stage**

The accuracy of the descriptions and composites was assessed in tasks identical to those used in Experiment 1.

**Rating and Matching Tasks**

Forty psychology students (85% Hispanic, 12% White, 2% Black, 1% “other”; 60% female; M = 21 years old) served as rating judges, and an additional 40 students (70% Hispanic, 25% White, 2.5% Puerto Rican, 2.5% “other”; 60% female; M = 20 years old) served as matching judges. The procedures were identical to those used in Experiment 1.

**Results and Discussion**

Similar to Experiment 1, analyses examined whether evaluations of own- and other-race composites varied as a function of judges’ ethnicity on either task. For the rating task, broken down by judges’ ethnicity, each group’s mean accuracy rating for the Hispanic faces and Black faces ranged between M = 6.12–6.73 and M = 6.20–7.10, respectively, with no significant effect of judges’ ethnicity, F(2, 36) < 1, p > .05. For the matching task, judges’ accuracy for the Hispanic and Black faces ranged between 26–29% and 16–17% respectively, with no effect of judges’ ethnicity, F(2, 36) < 1, p > .05. Thus, the following analyses are collapsed across all judges.

A 2 (Type of Recall) × 2 (Target Ethnicity) ANOVA on rating accuracy revealed a significant main effect of Type of Recall, F(1, 116) = 115.36, p < .001; η² = 0.50 (see Table 1). The descriptions were rated as significantly more accurate (M = 6.99, SD = 0.59) than the composites (M = 5.29, SD = 1.06). The main effect of Target Ethnicity was not statistically significant, nor was the interaction, both p values > .05. A 2 (Type of Recall) × 2 (Target Ethnicity) ANOVA on matching accuracy also revealed a significant main effect for Type of Recall, F(1, 116) = 7.42, p < .01; η² = 0.06. The proportion of descriptions matched to the correct target face (26%) was reliably higher than the proportion of composites correctly matched (18%). The main effect of Target Ethnicity was statistically significant, F(1, 116) = 9.27, p < .01; η² = 0.07. The Hispanic recalled information was correctly matched to the target face more often (27%) than was the Black recalled information (17%). The interaction was not statistically significant, p > .05. Again, the relationship between the rating and matching tasks was moderate and significant, r = .35, n = 120, p < .001.

The significant main effect of Target Ethnicity on the matching task was found amongst a sample of mostly Hispanic judges. Thus, to clarify whether this effect was in fact a function of the ORB during recall rather than an ORB in judges’ ability to match the information with the correct face, an additional sample of Black judges (N = 22) completed the matching task as described above, in a similar fashion to the earlier groups of judges. The results revealed that information recalled about Hispanic faces was correctly matched to the target face more often (20%) than was the Black recalled information (10%), t(21) = 5.62, p < .001, thus replicating the ORB effect found with our majority Hispanic judges described above.

As in Experiment 1, composite participants filled out a questionnaire. When asked To what extent do you think your facial composite is similar to the actual face you viewed?, participants’ mean rating was 3.33 (SD = 1.22) on a 7-point scale with larger numbers indicating greater similarity. When participants were asked How easy or difficult was it for you to create a facial composite that accurately represents the actual face you viewed?, participants’ mean rating (1 = extremely easy, 7 = extremely difficult) was 5.05
(SD = 1.39) with the majority of participants endorsing an answer above the scale midpoint. No responses were relat-
ed to the ethnicity of composite face, $p$ values $> .05$. Last, when asked If you thought the process of creating the com-
posite was difficult, what made it so?, 52% endorsed “searching through all the features interfered with my
memory of the face,” 50% endorsed “the facial feature cat-
alog was limited; I could not find the features I was looking
for,” 32% endorsed “it was difficult to position individual
features where I wanted in the facial space,” and 25% en-
dorsed “I couldn’t remember the face well enough to re-
construct it.”

Some support for our first hypothesis was found: A
greater number of both the descriptions and composites
based on Hispanic faces were matched to the correct target
face than the descriptions and composites based on Black
faces, which persisted amongst both Hispanic and Black
judges. This finding of an ORB in face recall among His-
panics for Black faces but not for White faces is similar
to that of Platz and Hosch’s (1988) finding of a greater ORB
among Hispanics in their recognition of Blacks than
Whites. Some caution is warranted in interpreting our find-
ings, however, given the small effect size and the null re-
results on the rating task. Our second prediction was also
confirmed in finding a description advantage over the com-
posites on both the rating and the matching tasks, replicat-
ing Experiment 1. This finding is again supported by par-
ticipants’ general dissatisfaction with the accuracy of their
composites ($M = 3.33$ on a 7-point scale) and their view of
the difficulty of the task ($M = 5.05$ on a 7-point scale).

General Discussion

Two studies investigated face recall performance of His-
panics on a description and composite task, and explored
the ORB as a function of recall. Experiment 1 found an
advantage for the use of descriptions over composites on
both a rating and matching task, but found no evidence of
an ORB using White faces. Experiment 2 attempted to re-
late the description-advantage and provided a stronger
test of the ORB using Blacks as the cross-race group. The
results of Experiment 2 were consistent with the first ex-
periment in finding that the descriptions were more accu-
rate than the composites. Results also showed a significant
ORB on the matching task among a majority-Hispanic
sample of judges; this result was confirmed with a sample
of Black judges ruling out the possibility that the ORB was
caused during matching rather than during recall. However,
since the ORB shown in Experiment 2 was found only on
one evaluation task, and because the associated effect size
was small, caution must be used in interpreting the strength
of this effect among an Hispanic population.

One important finding from this research is that the de-
scriptions were perceived as more accurate representations
of a face than were the composites. Regardless of the eth-
nicity, the descriptions were rated as more accurate than the
composites and also provided a better basis for finding the
intended target face than did the composites. These results
are consistent with previous research showing descriptive
information to be reliably superior to visual likenesses on
similar tasks (Christie & Ellis, 1981; Flin et al., 1989), and
also support some of the literature on face processing in
memory when it comes to the use of verbal versus visual
modalities in externalizing memories (e.g., Shepherd & El-
lis, 1996). This is the first set of studies to replicate previous
findings of a description-advantage using participants and
stimulus faces of ethnic groups other than White, and using
current facial composite methods. Future research should
investigate whether this description-advantage persists
when using standard versus enhanced witness interview
methods (e.g., the cognitive interview, Fisher & Geiselman,
1992) and different description elicitation methods (e.g.,
free recall) that are often used by investigators (Meissner
et al., 2006).

The current findings of an advantage of descriptions
over facial composites are of practical relevance to law en-
forcement procedures; clearly verbal description is an in-
herent part of most crimes involving eyewitness identifica-
tion. These findings may be considered somewhat at odds
with what is known as the ‘verbal overshadowing effect’: 
The phenomenon that verbally describing a face between
presentation of that face and a subsequent identification
attempt can impair identification accuracy (e.g., Schooler
& Engstler-Schooler, 1990). However, Meissner and Brig-
ham’s (2001a) meta-analysis synthesizes the verbal over-
shadowing literature demonstrates that the effect is mini-
mized under some important conditions, including when
the postdescription delay (the amount of time between de-
scription elicitation and identification) is greater than ap-
proximately 30 min, and when more standard description
instructions (i.e., free recall) are used, both of which are
likely procedures used routinely by police in their investi-
gations (e.g., Meissner et al., 2006).

In interpreting the description-composite effect, data
from both studies show what could be a floor effect in terms
of composite accuracy scores (at least with the matching
task), consistent with previous composite research (Davies
& Valentine, 2006). It is also important to note that the
results of the matching task indicate fairly low levels of
description accuracy. Therefore, while the descriptions
might have appeared generally accurate in a side-by-side
comparison of the descriptions and corresponding faces,
the usefulness of those descriptions in selecting the corre-
sponding face amongst a group of faces was a difficult task.
This result is relevant to police as they develop procedures
for enlisting the public’s assistance in searching for crime
suspects.

The second important finding in this research is evidence
that an ORB in the recall of faces may exist; this effect is
certainly worthy of further investigation that more fully ex-
plores eyewitness’ recall tasks using various ethnic groups.
Currently there are several possible explanations for the lack

of broader support for this effect—in contrast with the reliable support for the ORB in recognition. A common explanation involves the possible differential reliance on configural versus featural information for processing own- versus other-race faces, respectively (Rhodes, Brake, Taylor, & Tan, 1989). A recognition task which relies mostly on configural processing might account for the own-race advantage, but description generation, a featural task, might not be expected to reveal differences between own- and other-race faces (e.g., Wells & Turtle, 1987). Alternatively, limitations in our linguistic ability to describe faces might transcend ethnicity (Sporer, 1988). Further, the composite program used requires witnesses to make selections from sets of many features, therefore they must be able to discriminate between features that are embedded among a group of similar others. Memory for features presented in isolation is worse than memory for whole facial configurations (Tanaka & Farah, 1993; Tanaka & Sengco, 1997), so the structure of the composite task may interfere with the original memory image (Comish, 1987; Jenkins & Davies, 1985), thus overriding any effect of ethnicity. Recall performance based on isolated feature presentation versus feature groups, and the role of ethnicity in this task, is unclear but provides an interesting avenue for future research.

This research did not fully cross participant ethnicity by target ethnicity in our investigation of the ORB; we included only Hispanic participants from a Hispanic-majority community because we were interested in the unique nature of their contact with other groups. Some argue in favor of using a fully-crossed design because any effects of the materials themselves can be ruled out in the case of an interaction, otherwise a main effect of stimulus ethnicity could be due to the faces themselves, particularly when a limited set is used (e.g., Wells & Olsen, 2001). Because a fully-crossed design was not implemented here, we were careful in our selection of the faces used and in pilot testing these faces to represent a range of distinctiveness in each set. An additional concern is whether there could be a racial/ethnic bias in judges’ evaluations of the own- and cross-race information presented here. That is, the majority of judges used in the rating and matching tasks in both experiments are Hispanic. Yet, statistical analyses ruled out differences between judges’ evaluations of the recalled information due to their ethnicity (albeit, the sample sizes of the other ethnic groups were small). Also, additional matching data from Experiment 2 using Black judges confirmed the results found using a mostly Hispanic sample. Still, whether a race bias exists in the evaluation of own- and other-race information is a concern and an important topic for future research.

The current research provides support for the use of facial descriptions over composites for recalling a face; in both studies the descriptions were more accurate and provided a better cue to finding the correct face than did the composites. But despite difficulties with facial composite construction demonstrated here and previously, law enforcement often rely upon this tool, therefore we recommend that law enforcement support and consider adopting newer, sophisticated technologies (e.g., Frowd, Hancock, & Carson, 2004; Tre-doux, Rosenthal, da Costa, & Nunez, 1999) that are designed based on psychological theory and face processing in memory that may better assist witnesses in the externalization of memory images. Our results also showed some evidence for an ORB in face recall amongst Hispanic witnesses. Interestingly, the effect was found when the other-race group was of an ethnic minority; the effect was not found when using Whites as the outgroup. Research should investigate whether the ORB in face recall replicates among Hispanics, and would persist among different ethnic groups who have varying degrees of contact with other groups.

Acknowledgments

We thank Rebecca Bautista, Gina Castanon, Daniel Degadillo, Joni Eastman, Andrea Guthrie, and Stephanie Verlander for their assistance with data collection, entry and analysis. These data were presented at the 2003 meeting of the Society for Research in Memory and Cognition and the 2004 meeting of the American Psychology-Law Society, and represent a portion of the first author’s doctoral dissertation.

References


Read, & M. Toglia (Eds.), Handbook of eyewitness psychology (pp. 59–86). Mahwah, NJ: Erlbaum.


Received December 26, 2006
Revision received March 20, 2007
Accepted March 20, 2007

Dawn McQuiston-Surrett

Arizona State University
4701 W. Thunderbird Rd.
Glendale, AZ 85306
USA
E-mail dawn.mcquiston@asu.edu