2011

Initial Testing Reduces Eyewitness Suggestibility for Faces

Jessica Ann Lapaglia
Iowa State University

Initial testing reduces eyewitness suggestibility for faces

by

Jessica Ann LaPaglia

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Psychology

Program of Study Committee:
Jason C. K. Chan, Major Professor
Robert West
Gary L. Wells

Iowa State University

Ames, Iowa

2011

Table of Contents

List of Tables

List of Figures

List of Appendices

Acknowledgments

I would like to thank Jason Chan for being an excellent mentor and for the enormous amount of time, effort, and resources that he put into assisting me on this project. I would also like to thank Rob West, Gary Wells, Veronica Dark, Doug Bonnett, Miko Wilford, and Matt Erdman for their helpful comments regarding the design of the experiments, statistical analyses, and interpretation of the results. Special thanks to my supportive husband, Jonathon LaPaglia. Finally, I would like to thank the following lab assistants for their help with data collection and coding on this project: Kim Foss, Sabrina Fruechtenicht, Jenn Piatak, Jeremy Schulz, and Bret Wojcik.

Abstract

Past research has demonstrated that verbally recalling the appearance of a perpetrator after witnessing a crime can hinder one's ability to identify that perpetrator in a subsequent lineup (verbal overshadowing; Schooler & Engstler-Schooler, 1990). A recent study by Chan, Thomas, and Bulevich (2009) revealed that taking an initial memory test for an *event* increases one's susceptibility to later misleading information. These findings contradict those from the testing effect literature, which indicate that initial testing should enhance memory performance (Roediger & Karpicke, 2006). In the current study I investigated the effects of verbally describing a *face* on eyewitness suggestibility to later misinformation. Subjects witnessed a simulated crime and then either took a test over their memory for the perpetrator of the crime or performed a distractor task. Following a short delay, subjects heard misleading information about the perpetrator or only correct information. All subjects then took a final test over their memory for the perpetrator. Experiment 1 examined memory for the perpetrator using a free recall and a cued recall test. Experiments 2a and 2b examined witness identification performance with a target-present and a target-absent lineup, respectively. Three major findings emerged. First, initial testing increased correct recall probabilities and decreased misinformation recall probabilities in Experiment 1. Second, initial testing increased the likelihood of making a correct identification in the target-present lineup. Third, testing reduced identifications of individuals who matched the description of the misinformation when subjects were forced to make an identification (i.e., a biased lineup procedure). Implications for eyewitness testimony are discussed.

Chapter 1. Introduction

After witnessing a crime, a person is usually asked to describe the perpetrator or to recall details of the event. A witness can encounter misinformation from a police investigator, the news media, or other witnesses prior to or following a description of the perpetrator or event. During later testimony, the witness may recall the misinformation instead of the correct event details. Research on the misinformation effect has revealed that memory for events and faces is malleable (Loftus, 1979b; Loftus & Greene, 1980). Chan, Thomas, and Bulevich (2009) have recently found that recalling a witnessed event can increase people's susceptibility to misinformation about that event—despite the testing effect literature's indication that testing is a powerful memory enhancer (Spitzer, 1939). In the current experiments, I examined how initial testing and misinformation can affect memory for a *face*.

Eyewitness Suggestibility for Events

In a typical misinformation experiment, subjects witness an event (e.g., the car stopped at a *stop sign* prior to an accident) and are later presented with misleading information (e.g., the car stopped at a *yield sign*) in a narrative or through misleading questions. Subjects who have been exposed to misleading information are more likely to recall the misinformation than those who had not been exposed to misleading information. Loftus, Miller, and Burns (1978) postulated that the misinformation effect occurs because the more recent misinformation replaces the original memory. This memory impairment hypothesis has been rigorously debated (McCloskey & Zaragoza, 1985; Zaragoza, McCloskey, & Jamis, 1987), and it is now clear that the misinformation effect is based on a combination of factors. Memory impairment, misinformation

acceptance (i.e., people report the misinformation because the original information was never properly encoded or because people accept the misinformation as correct; Belli, 1989), and retroactive interference may all be responsible for the misinformation effect (Loftus & Hoffman, 1989). The misinformation effect may also be the result of source misattributions. People may have memory for both the original detail and the misinformation, but incorrectly attribute the source of the misinformation to the original event (Johnson, Hashtroudi, & Lindsay, 1993).

Although the Loftus paradigm has been influential, it lacks a key component that is often present in a real-life situation—namely, an immediate recall test following the event. A witness may talk to a 911 operator or police investigator about the event immediately following its occurrence. Research has shown that taking an initial memory test, as opposed to additional studying, can enhance one's retention of the studied material (Roediger & Karpicke, 2006). Because testing is such a powerful memory enhancer, it may inoculate one from later misinformation. However, experiments examining the effects of testing on memory for an event and for a target face have yielded mixed results.

Chan et al. (2009) examined how an initial memory test can affect one's suggestibility to later misinformation. It was predicted that initial testing would enhance memory for the original event, thereby reducing eyewitness suggestibility. In a series of experiments, subjects watched an episode of the Fox television program "*24*" and were given an immediate memory test following the video or they were given a distractor task. Following a short delay, subjects listened to an audio narrative that contained some misleading information and then took a final memory test. Surprisingly, Chan et al.

found that testing made subjects *more* susceptible to the misinformation. It is uncertain how this phenomenon, termed retrieval-enhanced suggestibility (RES), operates, but Chan et al. provided several possible explanations.

Chan et al. (2009) suggested that RES could be the result of an enhancement of the learning of the misinformation following retrieval. For example, Tulving and Watkins (1974) found that when subjects learned paired associates (A-B) and were later asked to retrieve the target given its cue (A- __), subsequent learning of similar paired associates (A-D) was greater relative to subjects who had not been initially tested. Applied to an eyewitness situation, initial testing over the original event may increase the later learning of related, but misleading, information. The initial recall test may also draw attention to specific parts of the narrative, thus enhancing encoding of the misinformation presented in the narrative. For example, in Chan et al.'s experiments, subjects were asked about the vehicle that the main character drove. Later, subjects heard misleading information about this detail (i.e., a pick-up as opposed to an SUV). When subjects were asked about the vehicle in the initial memory test, this question may have inadvertently drawn their attention to the misleading detail in the narrative, thereby increasing misinformation recall on the final test.

In addition, the RES effect may be the result of insufficient reconsolidation. Recently recalled information may undergo a reconsolidation process, during which the memory becomes particularly malleable and vulnerable to interference (see Hardt, Einarsson, & Nader, 2010, for a review). Therefore, if misinformation is presented during this reconsolidation process, it may produce greater interference than if one had not recalled the event recently. In two experiments, Chan and Langley (2011) found that

RES occurs even when the misinformation is presented long after the completion of the reconsolidation period. Thus they concluded that disruption in reconsolidation is not necessary for RES to occur, although it may augment the RES effect in some situations (e.g., when misinformation is presented soon after the initial test). The current experiments investigated how initial retrieval affects eyewitness suggestibility for faces as opposed to events. Because faces are processed differently than events, RES may or may not occur with faces.

Face Processing verses Event Processing

Eyewitness memories can encompass memory for a perpetrator's appearance, objects in the environment, and the witnessed event as a whole. Faces and events are thought to be processed quite differently. A recent review by Kurby and Zacks (2008) explored how people perceive and remember events. People typically parse events into smaller segments and actions with distinct boundaries. This segmentation of events is an automatic process that is ultimately beneficial to memory for events. Rather than encoding every small detail of an event, one can encode the event into "chunks" of actions and details. In contrast to parsing an event into smaller segments for encoding, faces are considered to be processes as a single object, which is often referred to as holistic processing.

Tanaka and Farah (1993) postulated that faces are recognized holistically. They define a holistic representation as one without an internal part structure. In other words, faces are recognized based on the whole face rather than individual features or component parts such as eyes, nose, and mouth. In a series of three experiments, Tanaka and Farah asked subjects to memorize several whole and scrambled faces. Immediately

following this study phase, subjects were given a recognition test that measured accuracy for features presented in whole faces and in isolation. Subjects were better at identifying features from normal, whole, faces than at identifying facial features presented in isolation. Tanaka and Farah used inverted faces in addition to upright faces as the stimuli in Experiment 2. Once again, there was a benefit of presenting the feature in the context of the entire face, but this benefit was found only for upright faces. In the third and final experiment, researchers examined whether a holistic representation was specific to faces or if the same was true of other objects, such as houses. Results indicated that the benefits of holistic processing were in fact specific to faces.

Note that Tanaka and Farah's (1993) holistic theory is not the only account of face processing. Mauer, Le Grand, and Mondloch (2002) argue that face recognition depends on three types of configural processing: first-order relations (i.e., a face typically consists of two eyes above a nose above a mouth), second-order relations (the spacing among facial features), and holistic processing. Regardless of how exactly faces are processed and remembered, it appears safe to suggest that vast differences exist between event and face processing and it is possible that each is affected by testing and misinformation differently. I now review studies that have examined suggestibility for faces; I then review the effects of verbal descriptions on later perpetrator identifications.

Eyewitness Suggestibility for People

Although memory for faces and events are thought to involve different processes, several researchers have found both to be susceptible to suggestion and interference. For example, Christiaansen, Sweeney, and Ochalek (1983) showed that after witnessing an event, subjects' estimates of a suspect's weight varied widely depending on whether the

suspect was described as a dancer or a truck driver. In addition, several researchers (e.g., Loftus & Greene, 1980; Searcy, Bartlett, & Memon, 2000) have examined the effects of misleading postevent information on memory for faces. In an experiment described by Loftus and Greene (1980), subjects watched a video of a simulated crime. Following a 20 min delay, subjects read a description of the suspect supposedly written by a professor. This description included either all correct information or some misinformation (e.g., that the perpetrator had a moustache when in fact he did not). All subjects then wrote out a description of the perpetrator. Those who heard the misinformation were more likely to include the misleading detail in their description than those who did not hear any misleading information.

In a second experiment, Loftus and Greene (1980) examined whether misinformation could also affect an eyewitness' ability to select a target in a 12-person, simultaneous, target-absent lineup. Subjects who heard that the suspect had a moustache, when in fact he did not, were significantly more likely to choose a person with a moustache than those who did not hear the misinformation. Searcy, Bartlett, and Memon (2000) also found that encountering a misleading detail about a target face led to an increase in choosing a photo that included the incorrect detail. These experiments indicate that memory for a person, like events, is malleable.

Witness Descriptions of the Perpetrator

Eyewitness descriptions of a perpetrator can be a crucial part in apprehending a suspect. The most common technique in questioning witnesses about a suspect is free recall followed by probing questions (i.e., cued recall) to fill in any missing information. This technique, however, can lead to false recall (Meissner, Sporer, & Schooler, 2007).

Police investigators may press witnesses to generate descriptions of an event or perpetrator (similar to forced recall), which can cause witnesses to generate highly confident false memories (Lane & Zaragoza, 2009). Critically, confidently held memories are typically judged accurate by jurors even when they are incorrect (Wells, Lindsay, & Ferguson, 1979).

Eyewitness descriptions of a perpetrator are invaluable in apprehending a suspect; but are they accurate and to what extent? Van Koppen and Lochen (1997) compiled 2,299 witness descriptions and found that people were able to give accurate, though very general, descriptions of a target. Their descriptions typically included information such as gender, age, height, build, race, and hair color, but little information was offered for facial characteristics. When witnesses did describe facial features, they typically only elaborated on the upper half of the face, such as the eyes and hair.

For cases with no clear suspects, face composites are used to help identify the perpetrator. Tools such as Photofit, Identikit, Mac-a-Mug, and the FACES program have been developed whereby witnesses can compile a face from a database of component features (i.e., eyes, lips, noses, hair, etc.; Davies & Valentine, 2007). However, these face composites tend to bare little likeness to the person they attempt to model (Christie & Ellis, 1981; Wells & Hasel, 2007). Wells, Charman, and Olson (2005) investigated how building these face composites can affect later identification in a lineup. They found that building a face composite reduced the probability of correctly identifying the perpetrator in the lineup. Wells and Hasel (2007) suggested that the detrimental effects of building a face composite can be attributed to a mismatch in processing strategies between face encoding and composite building. Specifically, composite building is done

at a featural level whereas faces are typically encoded and recognized in a holistic manner.

Information provided by co-witnesses can also affect eyewitness memory (Leippe & Eisenstadt, 2007; Luus & Wells, 1994). When witnesses view an inaccurate face-composite purportedly built by another witness that includes a misleading detail, witnesses are more likely to identify a suspect with that misleading detail (Jenkins & Davies, 1985). In sum, building and viewing face composites can be detrimental to one's later face recognition accuracy. In fact, Christie and Ellis (1981) found that verbal descriptions actually provide more useful information about a person than face composites. However, verbally describing a face can also alter one's memory for a perpetrator.

Verbal Overshadowing and Facilitation

Research has indicated that verbally recalling specific features of a face can hinder the subsequent accurate recognition of a face. In their seminal study, Schooler and Engstler-Schooler (1990) termed this phenomenon verbal overshadowing. In one experiment, subjects watched a video of a bank robbery and either provided a description of the perpetrator from memory (i.e., a memory test) or participated in an additional distractor task. Subjects in the test condition were given 5 min to write out a detailed description of the robber's face and were encouraged to describe each facial feature in detail. All subjects were then shown an eight-person, simultaneous, target-present lineup with the option of selecting no photo. Schooler and Engstler-Schooler found that recalling features of a face reduced the likelihood of correct identifications in the lineup task.

In a recent review, Chin and Schooler (2008) identified three potential explanations for verbal overshadowing: self-generated misinformation (Meissner, Brigham, & Kelley, 2001), a criterion shift (Clare & Lewandowsky, 2004), and recoding interference (Schooler & Engstler-Schooler, 1990). Verbal overshadowing may occur because of self-generated misinformation. When asked to elaborate on a person's face, one may recall inaccurate details thus resulting in lower accuracy on a subsequent lineup (Meissner et al., 2001). Clare and Lewandowsky (2004) postulated that verbal overshadowing could be the result of a criterion shift (i.e., an increased reluctance to choose a target), because the verbal overshadowing effect disappeared when subjects were forced to choose a target from a target-present lineup.

Schooler and Engstler-Schooler (1990) postulated that verbal overshadowing may be due to recoding interference whereby verbalizing a visual memory results in an incorrect representation of the target face in memory. No interference occurs when people are asked to visualize a visual memory or verbalize a verbal memory. It is only when the memory and test have an incongruent processing requirement that overshadowing occurs. When people verbally describe a face, they break down the face into component parts—such as eyes, nose, and mouth—eliciting a focal or featural processing strategy. This featural processing differs from the way faces are normally processed (i.e., holistically) and such a mismatch in face processing may harm face recognition accuracy.

Macrae and Lewis (2002) showed that priming different types of processing (i.e., global or local) can affect identification accuracy in a lineup. In their experiment, subjects watched a simulated crime video and engaged in a distractor activity (control

group) or a letter identification task. In the letter identification task, subjects either reported the global or local identity in a series of Navon letters, which are pictures of letters made up of smaller letters (see Figure 1). If subjects were asked to report the global identity, they were to report the larger letter made up of the smaller letters, which is thought to prime holistic processing. If subjects were asked to report the local identity, they were to report the smaller letters that make up the large letter. This task is designed to prime featural processing. Subjects in the featural processing condition were less likely to identify the perpetrator and those in the holistic processing condition were more likely to identify the perpetrator compared to controls—indicating a benefit of holistic processing.

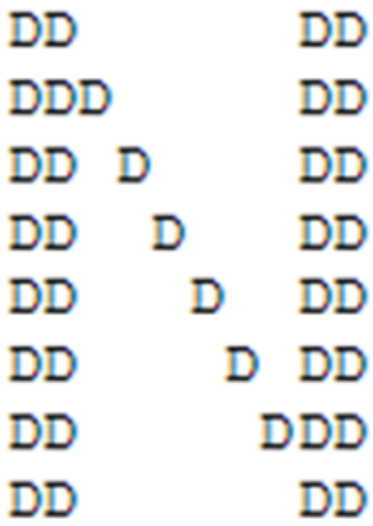


Figure 1. A Navon letter: A global N made up of local Ds.

Recalling a person's face does not always harm later identification performance. In a meta-analysis, Meissner and Brigham (2001) found that the verbal overshadowing effect is small but reliable (d = .12 over 29 studies). Moreover, the effect appears to be quite specific. When people are asked to *elaborate* on a face, they are more likely to show verbal overshadowing. Further, a short delay (under 30 min) between the initial test and the lineup typically results in verbal overshadowing, but delays longer that 30 min typically result in verbal facilitation (i.e., initial testing *increases* the likelihood of

correctly selecting the perpetrator; see Schooler & Engstler-Schooler, 1990, Experiment 5, for an exception). Short delays, however, are unlikely in real-life eyewitness situations. It is unrealistic for a person to witness an event, verbally describe the person, and see a lineup immediately afterward. In fact, show-ups (i.e., when the suspect is caught soon after the crime, the witness may be brought to the scene of the arrest to identify the perpetrator) are used more often in these situations (Dysart & Lindsay, 2007).

Since Meissner and Brigham's (2001) review of verbal overshadowing, several studies have found facilitatory effects of verbalization on later correct identifications. In a study conducted by Meissner et al. (2001), subjects viewed a target face for 10 s and were randomly assigned to one of four testing conditions following a 5 min distractor activity: forced recall, standard recall, warning recall, and control (no recall). Subjects in the standard recall instructions mirrored those of previous verbal overshadowing experiments. Subjects then saw a lineup either immediately or following a 30 min delay. There was verbal overshadowing in the standard condition with no delay. However, when the delay was increased to 30 min, there was a verbal facilitation effect. Specifically, subjects who verbally described the perpetrator chose the target more often (.57) than the control subjects (.37). Warning subjects that they should only describe features for which they are certain also resulted in significantly higher hit rates than controls. However, forcing subjects to generate elaborate descriptions of a target (forced recall) significantly decreased accuracy regardless of delay.

More recent investigations into verbal facilitation have uncovered some of the specific circumstances in which facilitatory effects are found. In a series of experiments conducted by Brown and Lloyd-Jones (2005), subjects studied multiple faces and, after

each face, described the face for 15 s or completed a distractor activity. An old/new recognition task immediately followed this study/recall phase. Brown and Lloyd-Jones found verbal facilitation with standard recall instructions, when subjects were asked to describe similarities and differences between faces, and when asked to provide both holistic and featural descriptors during the initial recall phase. Brown, Gehrke, and Lloyd-Jones (2010) also utilized this same procedure with standard recall instructions and found the verbal facilitation effect was greater for upright and unfamiliar faces than for inverted or familiar faces. In a recent review, Meissner, Sporer, and Susa (2008) explained that several factors, in addition to those described in Meissner and Brigham (2001), might contribute to verbal overshadowing and facilitation. First, verbal facilitation is more likely to occur when there are multiple targets whereas verbal overshadowing is more likely with a single target. Second, the extent to which subjects are provided with the opportunity to generate a verbal description can affect later correct identifications. For example, when descriptions are brief and precise, verbal facilitation is typically seen.

There are conflicting findings about whether verbal overshadowing occurs in a target-absent lineup. Clare and Lewandowsky (2004) and Memon and Rose (2001) found verbal facilitation with a target-absent lineup with no delay between verbalization and the lineup. However, Yu and Geiselman (1993) and Meissner (2002; Experiment 1) found verbal overshadowing with a target-absent lineup. Yu and Geiselman had a 48 hour delay and Meissner included a 5 min delay between verbalization and lineup identification. Recently, Sauerland, Holub, and Sporer (2008) examined choosing rates in a target-absent lineup following a 1 week delay between the description and the lineup.

They found no difference in correct rejections between the no description and description only conditions. It was only when subjects re-read their earlier descriptions that verbal interference occurred. Because of the mixed findings, the current study included both a target-present and target-absent lineup.

The Current Study

The current study examined whether verbally recalling a face immediately following the witnessed event enhances or reduces later eyewitness suggestibility for faces. Chan et al. (2009) found that initial retrieval can enhance suggestibility for events, but it is unknown whether this effect will generalize to faces. Faces are processed holistically whereas an event is processed sequentially as it unfolds over time. Therefore, the effects of testing on susceptibility to misinformation may be quite different for faces relative to events. Because witnesses are often questioned soon after a crime, it is important to understand how this initial test can affect one's later memory for the perpetrator—especially in the face of misleading information.

The overall design of the first experiment was similar to Experiment 1 of Loftus and Greene (1980) except that an initial test condition was included. In the first experiment, subjects watched a simulated crime and then described the perpetrator's face (initial test condition) or performed a distractor task (no initial test condition). Following a 20 min delay, subjects listened to a narrative describing the perpetrator. The narrative included either an erroneous detail (misleading condition) or only correct information (control condition). After an additional 10 min delay, all subjects then provided a verbal description of the suspect (i.e., a final recall test). In Experiments 2a and 2b, the final

recall test was replaced by a simultaneous, six-person, target-present and target-absent lineup, respectively.

It was hypothesized that initial testing would result in greater misinformation recall on the final test. Such a pattern can be considered consistent with the transfer-appropriate processing framework (Fisher & Craik, 1977; Morris, Bransford, & Franks, 1977). Specifically, the initial recall test forces subjects to break down the target's face into featural components and the narrative provided later is a featural description of the target. Thus, the initial test may better integrate the misinformation with the original memory. Alternatively, initial testing may reduce susceptibility to misinformation. Testing protects against forgetting, and it, therefore, may make subjects more resistant to misinformation. Witnesses are typically less suggestible to misinformation when the event has been encoded particularly well (Loftus, 1979a; Marche, 1999). Moreover, the current study included a 30 min delay between the verbal description and the lineup identification task, thus increasing the likelihood of verbal facilitation, such that the verbal description may protect the memory of the face from misinformation.

Chapter 2. Experiment 1

Method

Subjects and design. One hundred thirty-eight students at Iowa State University participated in this experiment for partial course credit. Ten subjects were excluded from analyses because English was not their primary language. Therefore, all analyses were based on the remaining 128 subjects (66 female, 62 male). The experiment used a 2 (test type: initial test vs. no initial test) X 3 (postevent information: control, facial hair, mole) between-subjects design. Sixty-four subjects heard the control narrative, 32 heard the facial hair detail, and 32 heard the mole detail.

Materials and procedure. The study was run on individual computer terminals separated by dividers. Up to eight subjects participated simultaneously. Subjects viewed two foil videos and then a simulated crime video. They were told that the video clips may be used in future studies and that the videos were being pilot tested to ensure that people are able to see and hear everything in the videos adequately. Subjects were further told that following the videos they would rate each clip on its video and sound quality. These instructions were designed to encourage incidental encoding of the critical event. The first two videos were 60 s long and featured neutral material with no people. The first video showed a Hawaiian beach and the second video showed a rabbit performing tricks.

The critical event video was approximately 45 s in length. This video showed a male student studying in a room (purportedly in the library) about 15 feet from the camera. The student answers a phone call and leaves the room. A man then approaches the desk, searches through the student's backpack for a wallet, takes the wallet and a

laptop computer, and quickly leaves. The perpetrator is a white male in his early 20s with short, brown hair and no other distinguishing characteristics. He is in full view for 15 s of the video. Afterwards, subjects completed an audio/visual rating task. In this task, subjects rated each video clip on its video and audio quality on a scale of 1 (*very poor*) to 7 (*excellent*). This rating task was followed by a demographic questionnaire (see Appendix A).

Subjects then either took the initial tests or played Tetris for 10 min as a distractor activity. Subjects in the initial test condition were first given 5 min to type out a description of the target. They were asked to be as detailed as possible in their description (see Appendix B for instructions given). Following this free recall test, subjects were administered a cued recall test, which included 12 questions that asked for descriptions of specific features of the target (see Appendix C for the list of questions). The questions ranged from more general details, such as clothing and build, to more detailed questions about the man's face. Subjects were given 25 s to answer each question.

Following the initial test/distractor phase, all subjects completed the computerized Operation Span (OSPAN) task, a test of working memory capacity (WMC; Unsworth, Heitz, Schrock, & Engle, 2005). In the OSPAN task, subjects memorized letter strings while solving simple math problems. The OSPAN was included to prevent rehearsal of the target event and to introduce a retention interval. Because each person completes this task at different times, the experimenter moved on to the next phase of the experiment only after all subjects tested during the same session were finished. This was the case for

all span tasks used in the current experiment. Most subjects completed this task within 20 min. No subject took more than 30 min for this task.

After subjects completed the OSPAN task, they listened to an audio narrative that included one piece of misinformation or no misinformation (control narrative). There were two different misinformation narratives—one mentioned that the perpetrator had facial hair on his chin; the other indicated that he had a mole on his left cheek (see Appendix D for narrative scripts). Whether a subject heard the control narrative or one of the misleading narratives was counterbalanced across subjects. Similar to Loftus and Greene (1980), subjects were told that a professor wrote out a description of the perpetrator immediately after watching the same video and that a research assistant read and recorded the description in the audio narrative. They were asked to listen to the narrative carefully, but were not given any further instructions.

Next, subjects completed the computerized Symmetry Span (SSPAN) task (Unsworth et al., 2005). In this task, subjects were to remember spatial locations while determining whether block shapes were symmetrical. This distractor phase lasted approximately 10 minutes. Afterwards, subjects took part in the final test phase, which included the same free recall and cued recall test as the initial test phase. Subjects were then asked, retrospectively, if they encoded the crime video intentionally (across Experiments 1 and 2, 25% of subjects reported that they had intentionally encoded the critical event video). Further, they were asked whether they noticed any incorrect information in the narrative (11% reported that they had noticed incorrect information in the postevent narrative).

In order to collect additional WMC data, subjects completed the computerized Reading Span (RSPAN) task (Unsworth et al., 2005). In the RSPAN task, subjects memorized letter strings while deciding whether sentences made sense. Scores on the three SPAN tasks were combined to produce a single estimate for subjects' WMC. Several researchers (Jaschinski & Wentura, 2002; Roediger & Geraci, 2007) have found executive functioning (as estimated by a battery of neuropsychological tests from Glisky, Polster, and Routhieaux, 1995) to be negatively correlated with misinformation recall for an eyewitness event. However, it is unknown whether individual differences in WMC will affect susceptibility to misinformation about faces.

Results and Discussion

Responses were classified as either Correct or Misinformation. The free recall test was coded using the first 11 questions on the cued recall test as the criteria for correct recall. For example, if a subject reports the man's shirt color and hair color correctly, but gives no other details, then that subject would have a free recall accuracy probability of .18 (2 out of 11 correct). Correct recall probabilities did not include question 8 (the question about the misinformation) in the final test. One researcher coded half of the free recall and half of the cued recall tests; a second researcher coded the other half. In addition, they each coded a subset of each other's already coded data. Their inter-rater reliability was high, $r(126) = .95$, $p < .01$.

Logistic regression analyses were used to examine interaction effects for binary data. Pearson's chi-square tests were used to examine simple and main effects for binary data. Partial eta squared (η_p^2) indicates effect size for analysis of variance (ANOVA). Cohen's *d* indicates effect size for t-tests. Phi (φ) indicates effect size for chi-square

tests. Working memory capacity (as measured by a principal component analysis[1]) was not related to any dependent measures of eyewitness memory in all experiments and therefore will not be mentioned further.

Initial Tests. See Table 1 for results from the initial test. Spontaneous reporting of the misinformation was rare in both free recall (M = .06) and in cued recall (M = .09). When a person recalled the misinformation in the initial test, it was always about the facial hair detail.

Table 1

Mean probabilities of correct and misinformation recall in the initial test phase in Experiment 1

	Free Recall	Cued Recall
Correct	.24 (.11)	.59 (.14)
Misinformation	.06 (.24)	.09 (.29)

Note. Standard deviations are presented in parentheses.

Final Test. Separate analyses were conducted for free and cued recall tests. See Table 2 for correct recall probabilities. A 2 (test type: no test, test) X 3 (postevent information: control, mole, facial hair) ANOVA revealed no significant interaction for correct recall in either the free or cued recall tests, Fs < 1. There was, however, a significant testing effect in both the free recall test, $F(1, 122) = 13.79$, $p < .01$, $\eta_p^2 = .10$, and the cued recall test, $F(1, 122) = 7.13$, $p < .01$, $\eta_p^2 = .06$. Subjects who took an initial test had a greater correct recall probability (M = .51 in free recall and M = .76 in cued recall) on the final

tests than those who did not take an initial test (M = .40 and M = .67). Note, however, that the testing effect in the free recall test is not a pure testing effect. The free recall tests were scored based on the cued recall questions and the tested subjects had been exposed to these questions prior to the final test. Therefore, tested subjects were at an advantage as they could use the initial cued recall questions could guide their subsequent free recall attempt.

Table 2

Mean probabilities of correct recall on the final cued and free recall tests as a function of postevent information condition in Experiment 1

		Free Recall Test		Cued Recall Test	
		No Initial Test	Initial Test	No Initial Test	Initial Test
Control	(N = 64)	.40 (.15)	.55 (.20)	.70 (.19)	.75 (.16)
Facial Hair	(N = 32)	.39 (.13)	.47 (.16)	.66 (.12)	.71 (.20)
Mole	(N = 32)	.40 (.16)	.51 (.10)	.66 (.11)	.81 (.14)

Note. Standard deviations are presented in parentheses.

Although the testing effect in the free and cued recall tests is not surprising, some previous studies (e.g., Chan and Langley, 2011; Chan et al., 2009) found no such testing effect with a 30 min delay between retrieval and the final memory test. This suggests that the information in the critical event video was susceptible to rapid forgetting—likely

due to the short length of the critical event video and the instructions given to promote incidental encoding of the video.

See Table 3 for the percentage of subjects reporting the misinformation on the final test. A 2 (test type) X 3 (postevent information) logistic regression analysis revealed a significant interaction for the free recall test, $\chi^2(2, N = 128) = 17.58$, $p < .01$, and the cued recall test, $\chi^2(2, N = 128) = 19.46$, $p < .01$. Specifically, testing reduced misinformation recall in the facial hair narrative condition, but not in the control or mole narrative condition. There was a significant misinformation effect in both free and cued recall. In the free recall test, subjects were far more likely to report the misinformation if they had heard the facial hair detail (53%; $\chi^2(1, N = 96) = 33.60$, $p < .01$, $\varphi = .59$) or the mole detail (63%; $\chi^2(1, N = 96) = 42.58$, $p < .01$, $\varphi = .67$) than those who heard the control narrative (3%). The cued recall data mirrored those from the free recall test; subjects were more likely to report the misinformation if they had heard the facial hair detail (63%; $\chi^2(1, N = 96) = 28.06$, $p < .01$, $\varphi = .54$) and the mole detail (69%; $\chi^2(1, N = 96) = 33.82$, $p < .01$, $\varphi = .59$) than if they had not (11%).

Most relevant for current purposes is the finding that initial testing reduced misinformation recall. Specifically, initial testing reduced the likelihood of subjects reporting the facial hair detail in the cued recall test, $\chi^2(1, N = 32) = 4.80$, $p = .03$, $\varphi = .39$. The data for subjects who heard the mole detail were less clear. To increase power, the data for the facial hair group and the mole group were collapsed, which revealed a significant testing effect such that initial testing reduced misinformation recall on the final cued recall test from 78% to 53%, $\chi^2(1, N = 64) = 4.43$, $p = .04$, $\varphi = .26$. However, no difference was found in the free recall test, $\chi^2 < 1$, $p = .80$.

Table 3

Percentage of subjects reporting the misinformation in the final test in Experiment 1

		Free Recall Test		Cued Recall Test	
		No Initial Test	Initial Test	No Initial Test	Initial Test
Control	(N = 64)	3	3	16	6
Facial Hair	(N = 32)	56	50	81	44
Mole	(N = 32)	56	69	73	63

Chapter 3. Experiment 2

In addition to providing person descriptions to help identify a suspect, witness memory of a perpetrator is often tested in a lineup identification task. Therefore, it is important to understand how the results from Experiment1 translate to a situation in which a witness must identify a perpetrator in a lineup. A target-present lineup and target-absent lineup were used as the final test in Experiments 2a and 2b, respectively. To facilitate comparisons between the two lineups, Experiments 2a and 2b were conducted simultaneously.

Experiment 2a: Target-Present Lineup

Method

Subjects and Design. Two hundred fifty-seven students at Iowa State University participated in this experiment for partial course credit. A total of 17 subjects were excluded from analyses: thirteen because English was not their primary language, two did not follow instructions, and two because of a computer error. Therefore, all analyses were based on the remaining 240 subjects (122 females, 112 males, 6 chose not to answer). The experiment used a 2 (test type: initial test vs. no initial test) X 2 (postevent information: control vs. misinformation) between-subjects design. There were 60 subjects in each condition.

Materials and Procedure. The materials and procedure for Experiment 2a were identical to those of Experiment 1 except the misinformation narrative that included the mole detail was not used. Instead, only the control narrative and the misleading narrative that included the facial hair detail (referred to as the misleading narrative from now on) were used because pilot testing indicated that subjects ignored the mole detail in the

lineup photos. This may have occurred because the mole detail was too small to be clearly visible on a face, but including a mole that is large enough might look unnatural.

The most significant change in Experiment 2a was the inclusion of the target-present lineup identification task (instead of a recall test). Subjects saw a six-person, simultaneous, target-present lineup (see Appendix E for sample lineup and instructions). Photos included in the lineup were obtained from a database of photos from students at Iowa State University. The six foil photos used in the lineup were chosen based on extensive pilot testing to ensure similar choosing rates across the photos (range: 5% to 24% on a forced-choice lineup). All foils and the target had no facial hair; however, each foil photo was altered using Photoshop so that there was a "clean-faced" version of the foil and a version in which the foil had facial hair on the chin (see Appendix F for altered and unaltered foil photos). In addition, all photos were altered to include a white collar. Five of the six foils were randomly chosen for each target-present lineup (all six were used in the target-absent lineup in Experiment 2b), such that each lineup contained three photos that matched the description of the control narrative (one target photo and two control foils) and three photos that matched the description of the misleading narrative (misleading foil). Whether a foil was consistent with the control narrative or had facial hair was randomized. If a subject chose a foil with facial hair, it was scored as a misinformation identification.

Subjects were asked to look at each photo carefully and to identify the perpetrator by pressing the number that appeared underneath his photo (1-6). They were also given the option of selecting no photo by pressing "N", which indicated that the target was not present in the lineup. The experimenter emphasized that the person in the video may or

may not be in the photo lineup. Subjects who indicated that the target was not in the lineup were shown the same lineup immediately afterwards and were asked to choose the person that looked most like the target from the video. With this procedure, it is possible to examine whether non-choosers would select a misleading foil in a forced-choice, biased lineup procedure[2].

Following the lineup, subjects were asked a series of questions obtained from Wells and Bradfield (1998; see Appendix G). This task was self paced. Afterwards, the final screen displayed all six unaltered foil photos and the target. Subjects were asked if they recognize any of the people on the screen from *outside of the experiment* (e.g., a friend, classmate, etc.) and to type in the corresponding number of the people they recognize. No subject recognized any person from the lineup.

Results and Discussion

Responses for the initial free and cued recall tests were classified as either Correct or Misinformation. Because inter-rater reliability in Experiment 1 was high, only one researcher coded the responses in the free and cued recall tests in Experiment 2. Lineup identification was coded as either Correct Identification, Misleading Foil Identification, Control Foil Identification (i.e., identification of a nontarget without facial hair), or No Identification.

Initial Test. See Table 4 for results from the initial test in Experiments 2a and 2b. Baserate false recall probability was low in the free recall and cued recall tests (M = .03 for both).

Table 4

Mean probabilities of correct and misinformation recall in the initial test phase in Experiment 2

		Free Recall	Cued Recall
Experiment 2a	Correct	.27 (.12)	.55 (.15)
	Misinformation	.03 (.18)	.03 (.16)
Experiment 2b	Correct	.28 (.14)	.57 (.13)
	Misinformation	.06 (.24)	.03 (.17)

Note. Standard deviations are presented in parentheses.

Lineup Identification. Lineup identification data are presented in Table 5. For correct identifications, a 2 (test type) X 2 (postevent information) logistic regression analysis revealed no significant interaction, $\chi^2 < 1$, $p = .56$. There was, however, a significant main effect of postevent information, $\chi^2(1, N = 240) = 5.66$, $p = .02$, $\varphi = .15$, such that misinformation reduced correct identifications from 23% to 12%. Consistent with previous research that used at least a 30 min delay between initial testing and lineup identification (Meissner et al., 2001), there was a significant testing effect, $\chi^2(1, N = 240) = 5.66$, $p = .02$, $\varphi = .15$. Specifically, subjects who took an initial test were more likely to select the target (23%) than those who were not initially tested (12%)—a verbal facilitation effect. What is new about the current study is that initial testing appeared to enhance correct identifications even after subjects encountered misinformation. However, planned comparisons revealed that testing significantly increased correct identification when there was no misinformation, $\chi^2(1, N = 120) = 4.66$, $p = .03$, $\varphi = .20$,

but this testing effect was diminished after subjects heard the misinformation, $\chi^2(1, N = 120) = 1.29$, $p = .26$, $\varphi = .10$.

Table 5

Percentage of subjects identifying the target, a misleading foil, a control foil, and making no identification in Experiment 2a

Postevent Information	Initial Test Condition	Identification of target	Identification of misinformation	Identification of control foil	No Identification
Control	No Initial Test	15	27	27	32
	Initial Test	32	28	10	32
Misleading	No Initial Test	8	42	8	42
	Initial Test	15	43	10	32

For misinformation identifications, there was no significant interaction between initial testing and postevent information, $\chi^2 < 1$, $p = .98$. There was, however, a significant misinformation effect, $\chi^2(1, N = 240) = 5.93$, $p = .01$, $\varphi = .16$. Misinformation increased the likelihood of selecting a misleading foil from 28% to 43%. There was no significant effect of initial testing on misinformation choosing. Therefore, testing increased correct identifications without reducing the misinformation effect.

For control foil identification, there was a significant interaction between initial testing and postevent information, $\chi^2(1, N = 240) = 10.39$, $p < .01$. Testing reduced control foil identifications in the control condition, but not in the misleading condition. There was also a marginally significant main effect of initial testing; subjects who took an initial test were less likely to choose a control foil (10%) than those who did not take an initial test (18%), $\chi^2(1, N = 240) = 2.85$, $p = .09$, $\varphi = .11$. Further, there was a significant main effect of postevent information; subjects who heard the control narrative were more likely to choose a control foil (18%) than those who heard the misinformation (9%), $\chi^2(1, N = 240) = 4.25$, $p = .04$, $\varphi = .13$.

Dunning and Perretta (2002) reported that subjects are more likely to make a correct identification if they responded within 10 to 12 seconds on a lineup identification task. Weber, Brewer, Wells, Semmler, and Keast (2004), however, found that the 10 to 12 second rule does not always apply. I examined the effects of response time (RT) on choosing rates in a time-boundary analysis (Figure 2). The time boundary analysis

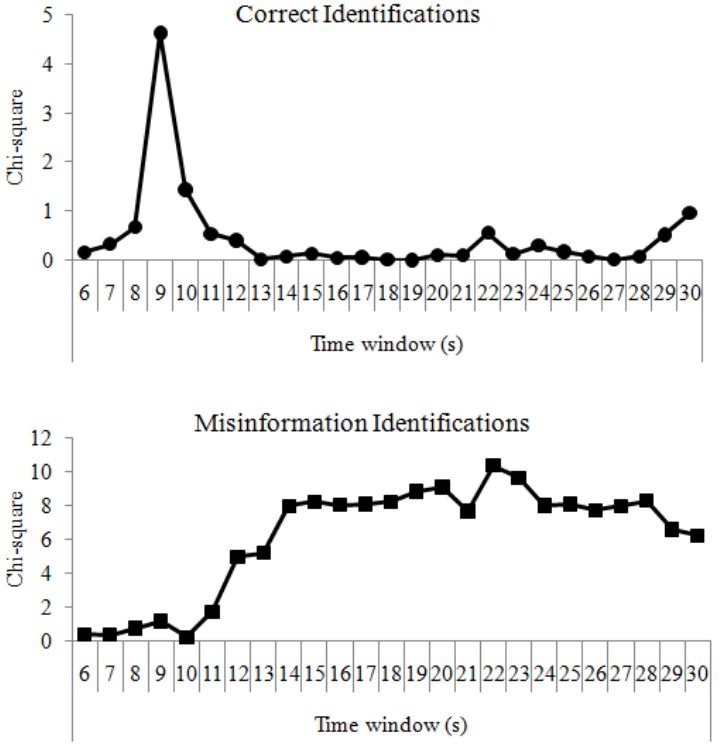


Figure 2. Plots of chi-square by time boundary for correct identifications and misinformation identifications.

mirrored that utilized by Dunning and Perretta and Weber et al. Accuracy was compared for subjects above and below a time boundary (faster or equal vs. slower) in a series of chi-square tests with the boundary set at 1 s time intervals from 6 s to 30 s (i.e., 6 s, 7 s, 8 s, and so on). All subjects were included in the analysis (including nonchoosers). The peak in the series of chi-squares indicates the time boundary discriminating between correct and incorrect subjects and between misinformation choosers and non misinformation choosers for the top and bottom figures respectively. For target identifications, subjects who responded in 9 sec or less were more accurate (33%; N = 24) than those with longer response latencies (18%; N = 216), $\chi^2(1, N = 240)$ 4.63, $p = .03$, $\varphi = .14$. In contrast, for misinformation idenitifications, subjects who responded within 22 sec were less likely to choose a misleading foil (27%; N = 142) than their slower counterparts (47%; N = 98), $\chi^2(1, N = 240) = 10.38$, $p < .01$, $\varphi = .21$.

Subjects who made no identification were shown the same lineup a second time and forced to select a photo that most resembles the man from the video (see Table 6 for choosing rates). A 2 (test type) X 2 (postevent information) logistic regression analysis revealed no significant interaction for correct identifications, $\chi^2 < 1$, $p = .52$, but there was a marginally significant testing effect (30% target identifications for tested subjects and 14% for nontested), $\chi^2(1, N = 81) = 3.14$, $p = .08$, $\varphi = .20$. There was also a marginally significant main effect of postevent information, $\chi^2(1, N = 81) = 3.14$, $p = .08$, $\varphi = .20$, such that misinformation reduced correct identifications (30% for controls and 14% for misled subjects).

For misinformation identifications, there was no significant interaction between initial testing and postevent information, $\chi^2(1, N = 81) = 1.19$, $p = .28$. Subjects who took

an initial test were less likely to choose a misleading foil (41%) than those who were not tested initially (64%), $\chi^2(1, N = 240) = 4.31$, $p = .04$, $\varphi = .23$. Moreover, those who listened to the misleading narrative were more likely (66%) to choose a misleading foil than control subjects (39%), $\chi^2(1, N = 240) = 6.63$, $p = .01$, $\varphi = .28$.

Table 6

Percentage of subjects identifying the target, a misleading foil, and a control foil under forced identification in Experiment 2a

Postevent Information	Initial Test Condition	Identification of target	Identification of misinformation	Identification of control foil
Control	No Initial Test (N = 19)	26	42	32
	Initial Test (N = 25)	33	33	33
Misleading	No Initial Test (N = 18)	4	80	16
	Initial Test (N = 19)	26	47	26

Although witnesses are never forced to make an identification in real-life, they may feel pressured to make an identification in a biased lineup procedure. To examine the effects of initial testing and misinformation in a biased lineup, I examined the combined identification rates of the initial lineup and the forced-choice lineup (see Table 7). There was no significant interaction between initial testing and postevent information

for correct identifications, $\chi^2 < 1$, $p = .60$. There was, however, a significant testing effect, $\chi^2(1, N = 240) = 9.09$, $p < .01$, $\varphi = .20$. That is, testing increased correct identifications from 17% to 33%. There was also a significant main effect of postevent information, $\chi^2(1, N = 240) = 7.37$, $p < .01$, $\varphi = .18$, such that the control group was more likely to select the target (33%) than the mislead group (17%).

Table 7

Percentage of subjects identifying the target, a misleading foil, and a control foil when the identification rates from the regular and forced-choice lineups were combined in Experiment 2a

Postevent Information	Initial Test Condition	Identification of target	Identification of misinformation	Identification of control foil
Control	No Initial Test	23	40	37
	Initial Test	42	38	20
Misleading	No Initial Test	10	75	15
	Initial Test	23	58	18

For total misinformation identifications, there was a significant interaction between initial testing and postevent information, $\chi^2(1, N = 240) = 5.26$, $p = .02$. Specifically, the size of the misinformation effect was reduced by nearly half (from 35% to 20%) if subjects had been tested. There was also a significant misinformation effect, $\chi^2(1, N = 240) = 18.21$, $p < .01$, $\varphi = .28$. Subjects in the misleading group chose a

misleading foil at a higher rate (67%) than controls (39%). For subjects who heard the misleading narrative, there was a significant testing benefit, $\chi^2(1, N = 240) = 3.76$, $p = .05$, $\varphi = .18$; initial testing reduced misinformation identifications, thus replicating the findings in Experiment 1 but in a lineup identification task.

Follow-up Questions. Following the lineup identification task, subjects were asked several follow-up questions. See Table 8 for results. These data were collected for exploratory purposes only. Compared to nontested subjects, the tested subjects rated themselves as paying less attention to the perpetrator's face, $t(238) = 2.89$, $p < .01$, $d = .37$, and less able to pick out details from his face in the video (marginally significant; $t(238) = 1.80$, $p = .07$, $d = .23$). However, initial testing enhanced subjects' ability to estimate the length of time that the target's face was visible, $t(238) = 1.97$, $p = .05$, $d = .26$. Specifically, subjects who took an initial test more closely approximated the length of time the target's face was visible ($M = 12.99$) than those who did not take an initial test ($M = 10.44$); the target's face had been in view for 15 sec.

Subjects who correctly identified the target reported greater confidence (68%) and had a greater willingness to testify (42%) than those who were incorrect (46% and 28%), $t(234) = 3.78$, $p < .01$, $d = .65$, and $t(234) = 2.80$, $p < .01$, $d = .45$, respectively. Compared to incorrect subjects, the correct subjects more closely approximated the length of time the target's face was in view, rated themselves as having a better view of the target, found the lineup task easier, and estimated that they spent less time on the lineup, ts > 2.68, ps < .01, ds > .42. Compared to subjects who did not choose the misinformation, misinformation choosers rated themselves as paying less attention to the

Table 8

Subjects' ratings on the follow-up questions in Experiment 2a with standard deviations in parentheses

	No Initial Test		Initial Test	
	Control	Misleading	Control	Misleading
Goodness of view	4.86 (2.11)	5.27 (1.91)	5.07 (1.62)	4.78 (1.60)
Seconds target's face was in view	11.75 (11.58)	9.37 (6.57)	13.91 (10.89)	12.73 (9.23)
Ability to make out facial features	4.44 (2.07)	4.61 (1.85)	4.00 (1.59)	4.05 (1.69)
Distance from the camera	8.12 (3.99)	8.64 (3.77)	9.41 (5.06)	8.42 (4.63)
Attention paid to target's face	4.24 (1.76)	4.65 (1.99)	3.96 (1.84)	3.64 (1.73)
Confidence in lineup decision	47.27 (26.00)	50.91 (27.75)	48.98 (28.22)	46.67 (26.76)
Difficulty of the lineup	6.25 (2.89)	5.79 (2.12)	6.26 (2.08)	6.24 (2.10)
Time spent making a decision	5.53 (1.99)	5.43 (2.26)	5.59 (2.08)	5.73 (2.34)
Willingness to testify	28.64 (27.90)	32.14 (31.41)	27.48 (26.79)	29.91 (29.04)
Trust in another witness' decision	4.41 (2.28)	4.75 (2.08)	4.43 (2.19)	4.82 (2.07)
Basis to make a decision	4.32 (1.88)	4.39 (1.87)	4.31 (1.80)	4.02 (1.90)

target's face, $t(238) = 2.02$, $p = .04$, $d = .13$, and were poorer at estimating the length of time the target's face was in view, $t(238) = 2.32$, $p = .02$, $d = .35$.

Experiment 2b: Target-Absent Lineup

Method

Subjects and Design. Two hundred and twelve students at Iowa State University participated in this experiment for partial course credit. A total of 12 subjects were excluded from analyses: ten because English was not their primary language, one did not follow instructions, and one because of a computer error. Therefore, all analyses were based on the remaining 200 subjects (101 females, 96 males, 3 did not report their sex). The experiment used a 2 (test type: initial test vs. no initial test) X 2 (postevent information: control vs. misinformation) between-subjects design, with 50 subjects included in each condition.

Materials and Procedure. The materials and procedure for Experiment 2b were identical to those of Experiment 2a except that the lineup did not include the target. The position of the six foils was randomized and whether a foil was presented with or without facial hair was also randomized.

Results and Discussion

Initial Test. See Table 4 for initial test correct and misinformation recall probabilities. Baserate false recall probability was low in free recall ($M = .06$) and cued recall ($M = .03$).

Lineup Identification. Lineup identification data for Experiment 2b are presented in Table 9. No significant interactions or main effects were found for no identification rates. There was no significant benefit of initial testing for no identifications (i.e., correct

rejections). In fact, no identifications were identical for the no initial test and initial test groups. Very few studies have examined verbal overshadowing in a target-absent lineup. Of the ones that did, the results are mixed, with two finding verbal overshadowing (Meissner, 2002, Experiment 1; Yu & Geiselman, 1993), two finding verbal facilitation (Clare & Lewandowsky, 2004; Memon & Rose, 2001), and one finding neither verbal overshadowing nor facilitation (Sauerland et al., 2008). Unfortunately, the current experiment only adds to this confusing state of affairs. It might be that the effects of initial testing on a subsequent target-absent lineup are not particularly robust.

Table 9

Percentage of subjects making no identification, identifying a misleading foil and a control foil in Experiment 2b

Postevent Information	Initial Test Condition	No Identification	Identification of misinformation	Identification of control foil
Control	No Initial Test	44	22	34
	Initial Test	44	34	22
Misleading	No Initial Test	50	42	8
	Initial Test	50	34	16

For misinformation identifications, there was a marginally significant interaction between initial testing and postevent information, $\chi^2(1, N = 200) = 3.56$, $p = .06$. Misinformation exposure increased misinformation identifications for the nontested subjects, but not for the tested subjects. There were no significant main effects.

The effect of RT on choosing rates was examined in a time-boundary analysis (Figure 3). Compared to slower subjects (18%; N = 38), those responding within 20 sec were more accurate (36%; N = 162), $\chi^2(1, N = 200) = 4.51$, $p = .03$, $\varphi = .15$. Further, subjects who responded within 11 sec were less likely to choose a misleading foil (35%; N = 82) than those with longer response latencies (55%; N = 118), $\chi^2(1, N = 200) = 7.55$, $p < .01$, $\varphi = .19$.

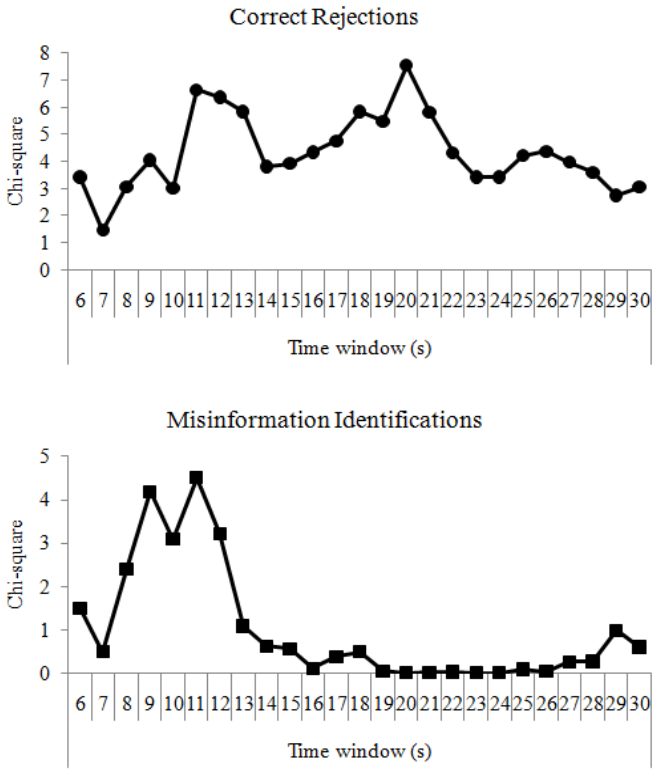


Figure 3. Plots of chi-square by time boundary for correct rejections and misinformation identifications.

Subjects who made no identification were shown the same lineup again and forced to make an identification. See Table 10 for choosing rates in this forced-choice lineup. There was a marginally significant main effect of testing for misinformation

Table 10

Percentage of subjects identifying a misleading foil and a control foil in the forced-choice lineup in Experiment 2b

Postevent Information	Initial Test Condition		Identification of misinformation	Identification of control foil
Control	No Initial Test	(N = 22)	68	32
	Initial Test	(N = 22)	45	55
Misleading	No Initial Test	(N = 25)	64	36
	Initial Test	(N = 25)	48	52

identifications, $\chi^2(1, N = 94) = 3.50$, $p = .06$, $\varphi = .19$. That is, initial testing reduced the likelihood of selecting a misleading foil from 66% to 49%. Total choosing rates in a biased lineup were also examined (see Table 11). For total misinformation identifications, there was a significant interaction between initial testing and postevent information, $\chi^2(1, N = 200) = 5.36$, $p = .02$. For nontested subjects, exposure to misinformation increased the probability that one would identify the misleading foils from 52% to 74%, $\chi^2(1, N = 100) = 5.19$, $p = .02$, $\varphi = .23$. There was also a marginally significant testing benefit—for those who heard the misleading narrative, initial testing reduced the likelihood that a subject would choose a misleading foil (from 74% to 58%), $\chi^2(1, N = 100) = 2.85$, $p = .09$, $\varphi = .17$. Although this effect was only marginally

significant, it has important applied implications that are discussed in the general discussion.

Table 11

Percentage of subjects identifying a misleading foil and a control foil when the identification rates from the regular and forced-choice lineups were combined in Experiment 2b

Postevent Information	Initial Test Condition	Identification of misinformation	Identification of control foil
Control	No Initial Test	52	48
	Initial Test	55	45
Misleading	No Initial Test	74	26
	Initial Test	58	42

Follow-up Questions. Data for the follow-up questions are presented in Table 12. Subjects who were initially tested were more accurate at estimating the length of time the target's face was visible (M = 13.15 s) than those who did not take an initial memory test (M = 9.64 s), $t(198) = 2.98$, $p < .01$, $d = .42$ (the target's face had been in view for 15 s). Compared to subjects who correctly rejected the lineup, choosers rated the lineup as being less difficult (M = 5.70 versus 6.39; $t(198) = 2.25$, $p = .02$, $d = .34$) and claimed to

take less time on the lineup identification task (M = 5.32 versus 6.29; $t(198) = 3.57$, $p < .01$, $d = .51$).

In a meta-analysis, Sporer, Penrod, Read, and Cutler (1995) found a positive relationship between confidence and accuracy (also see Brewer & Wells, 2006; Lindsay, Read, & Sharma, 1998). In fact, eyewitness confidence is often treated as compelling evidence in court (Wells, Olson, & Charman, 2002). Similar to these previous findings, the data from the target-present lineup showed a positive confidence-accuracy relationship. However, in the target-absent lineup, subjects who correctly rejected the lineup were *less* confident (34%) and *less* willing to testify (23%) compared to those who were incorrect (58% and 35%), $t(198) = 6.70$, $p < .01$, $d = .94$, and $t(198) = 2.96$, $p < .01$, $d = .42$, respectively.

Although witness confidence can be a deciding factor in courtroom decisions, previous research has indicated that certain factors, such as postidentification feedback, may inflate eyewitness confidence (Wells & Bradfield, 1998). In the target-absent lineup of the current study, encountering misinformation increased subjects' confidence in their lineup decision if they had not been tested (but no effects were found for tested subjects); this interaction, however, was only marginally significant, $F(1, 196) = 3.02$, $p = .08$, $\eta_p^2 = .02$. Surprisingly, subjects who heard the misleading narrative rated themselves as having a greater basis to make an identification (M = 4.56) than control subjects (M = 4.03), $t(198) = 2.05$, $p = .04$, $d = .29$. Moreover, subjects who chose a misleading foil rated the lineup as being easier (M = 5.54) compared to subject who did not select a misleading foil (M = 6.26), $t(198) = 2.38$, $p = .02$, $d = .37$.

Table 12

Subjects' ratings on the follow-up questions in Experiment 2b with standard deviations in parentheses

	No Initial Test		Initial Test	
	Control	Misleading	Control	Misleading
Goodness of view	5.46 (1.69)	5.53 (1.67)	5.12 (1.61)	5.14 (1.57)
Seconds target's face was in view	10.46 (7.06)	9.02 (6.70)	11.92 (7.99)	14.38 (10.79)
Ability to make out facial features	4.23 (1.55)	4.39 (1.30)	4.08 (1.71)	4.50 (1.53)
Distance from the camera	8.73 (3.99)	8.98 (5.20)	8.06 (3.61)	7.78 (3.34)
Attention paid to target's face	4.04 (1.86)	4.33 (1.89)	4.14 (1.63)	3.78 (1.82)
Confidence in lineup decision	41.15 (27.95)	55.80 (28.77)	44.60 (27.31)	44.76 (29.82)
Difficulty of the lineup	6.23 (2.02)	6.00 (1.98)	6.22 (1.92)	5.72 (2.23)
Time spent making an decision	6.08 (1.78)	5.73 (2.06)	5.76 (1.90)	5.52 (2.14)
Willingness to testify	30.31 (28.81)	30.92 (30.20)	25.20 (25.78)	31.16 (30.14)
Trust in another witness' decision	4.31 (2.12)	5.06 (1.91)	4.44 (1.93)	4.38 (2.12)
Basis to make a decision	3.98 (1.95)	4.71 (1.67)	4.04 (1.67)	4.42 (2.06)

Chapter 4. General Discussion

The present study examined the effects of taking an initial memory test on later perpetrator descriptions and identifications when one encounters misleading information. Two important findings emerged from these experiments. First, in both experiments, misinformation hindered performance on subsequent recall and identification tasks. Specifically, subjects who heard misinformation had lower accurate recall probabilities and higher false recall probabilities in Experiment 1. Moreover, misinformation reduced correct identifications and increased misinformation identifications in the target-present lineup. Second, initial testing had positive effects on later perpetrator descriptions and identifications. Initial testing improved later accurate descriptions of the perpetrator and reduced later false recall in Experiment 1. In Experiment 2, initial testing improved correct identifications in the target-present lineup and reduced misinformation identifications in the biased, target-present and target-present lineups. This finding contradicts the RES effect. I now discuss these main findings in detail.

The Adverse Effects of Misleading Information

The misinformation effect was found in Experiments 1 and 2a, thus replicating the findings of Loftus and Greene (1980; note, however, that Loftus and Greene examined misinformation effects in a target-absent lineup and the current study found a misinformation effect in the target-present condition). Compared to subjects who heard the control narrative, mislead subjects in the target-absent lineup experiment who received no initial test were more confident in their selection and reported a greater willingness to testify. Not only did encountering misinformation increase false recall

probabilities and misinformation identifications and inflate eyewitness confidence, it also decreased correct identifications.

Misinformation can be encountered anywhere (e.g., the news media or other witnesses). Therefore, what is perhaps most alarming is that misinformation greatly reduced correct identifications in the target-present lineup. Overall, correct identifications dropped from 30% in the control condition to a mere 14% in the misleading condition. Another important finding was that misinformation reduced the benefits of initial testing in Experiment 2a. Without misinformation, initial testing increased correct identifications by 17%, but when misinformation was presented, the testing benefit was reduced to 7%. Therefore, misinformation can have damaging effects on eyewitness identification—even reducing the beneficial effects of initial testing.

Given these findings, one may ask how the effects of misinformation can be minimized. Providing people a warning about the credibility of a source can reduce, and even eliminate, the misinformation effect for events—especially when the warning is specific (Greene, Flynn, & Loftus, 1982; Lindsay, 1990; Wright, 1993). However, specific warnings are nearly impossible in the real-world. Therefore, eyewitnesses should be wary of the sources of information they are exposed to. That is, they should consider the validity of information provided by the news media or overheard from other witnesses at the scene of the crime.

Another way to reduce eyewitness suggestibility for faces may be to change the type of lineup used. Sequential lineups have been shown to reduce false positives compared to simultaneous lineups (see Steblay, Dysart, Fulero, & Lindsay, 2001, for review). This is because in simultaneous lineups, people typically choose the person who

looks most like the perpetrator by making relative judgments. In sequential lineups, however, people see one person at a time and are more likely to make an absolute judgment by comparing each person to their memory for the perpetrator. Therefore, using a sequential lineup may reduce misinformation identifications. Unfortunately, most law enforcement agencies in the United States still use the simultaneous lineup, according to the National Institute of Justice (NIJ; www.ojp.usdoj.gov/nij).

The Benefits of Initial Testing

Experiment 1 showed that initial testing improved later correct descriptions of a face. This finding is consistent with the extensive body of literature emphasizing the benefits of testing on memory retention. Experiment 2a also revealed a benefit of initial testing for correct identifications in a target-present lineup. Instead of finding verbal overshadowing, Experiment 2a showed verbal facilitation. That is, taking an initial test increased later corrected identifications. Perhaps more important for current purposes, however, is that initial testing reduced the harmful effects of misinformation.

In Experiment 1, false recall probabilities dropped dramatically from 78% in the no initial test condition to 53% in the initial test condition. In Experiments 2a and 2b, there was no such reduction in misinformation identifications in the first, unbiased lineup. However, when examining total output, the benefits of initial testing emerged. For those who heard the misleading narrative, misinformation identifications dropped from 75% in the no initial test group to 58% in the initial test group (averaged across Experiments 2a and 2b). Examining the effects of initial testing and misinformation in a biased lineup procedure has real-world implications. Although reforms have been implemented in

Wisconsin, North Carolina, New Jersey, and several large cities, biased lineup administration is still common practice in most precincts (www.innocenceproject.org).

The benefits of initial testing on witness memory for people are clear from these data. Intriguingly, this pattern seems to contradict those reported by Chan et al. (2009), who found that testing increased eyewitness suggestibility. Why, then, did the current experiments result in a testing benefit instead of retrieval-enhanced suggestibility? I propose two possible explanations for these disparate findings. First, the materials and procedure used in the current experiments differed from those utilized by Chan and colleagues. Second, faces and events are processed differently. Therefore, the effects of testing and misinformation may be different for memory for faces, relative to events, because of this processing difference.

There is an obvious difference in materials used in the current experiments relative to those used by Chan and colleagues (2009). Chan et al.'s witnessed event lasted ~40 min whereas the current study's event lasted 45 s. For a drawn out 40 min event, Chan et al. found little forgetting between the event video and the final test, and so initial testing did not produce a significant testing effect. But when an eyewitness event is very short, an initial test may be more effective at protecting against forgetting and misinformation because better encoded information is typically less susceptible to misinformation (Loftus, 1979a; Marche, 1999). Further, the current study used instructions to promote incidental encoding of the critical event; Chan et al. specifically told subjects that they would take a test over their memory for the video. Intentional encoding instructions may increase the likelihood of rehearsal. Taking these two factors into consideration, it is not surprising that a significant testing effect was not found in

Chan et al.'s study. In contrast, a significant testing effect was found in Experiment 1, suggesting that initial testing had enhanced retention of the fleeting witnessed event. Another difference was the instructions given to subjects for the audio narrative. The instructions used mirrored those from Loftus and Greene (1980) in which the experimenter told subjects that the narrative was written by a professor. This may have acted as a warning for tested subjects. Tested subjects were well aware of the difficulty of the initial test and therefore may have put little trust into the memory of another witness.

Despite the different materials used in the present study compared to those utilized by Chan et al. (2009), the main purpose was to determine whether the RES effect would generalize to faces, and not whether RES would occur with a shorter event video and incidental encoding instructions. Memory for faces could react differently to testing and misinformation than memory for events because faces and events are processed differently. As mentioned in the introduction, faces are thought to be processed holistically whereas events are processed sequentially. Although holistic processing appears to be an efficient method for remembering and recognizing faces, Wilford and Wells (2010) recently found that holistic processing has one striking disadvantage. Namely, people are poorer at localizing changes to a person's face than to other objects such as houses. When subjects saw a face that had been altered (e.g., with a different nose), they were better at detecting that a change had occurred, but were worse at detecting *what* had changed about the face, relative to altered houses. This finding suggests that holistic processing, relative to featural processing, impairs subjects' ability to pinpoint what had changed between two faces.

Relating this finding to the current study, subjects who had been tested initially may have been better at localizing the change than those who had not taken an initial test. When subjects took the initial test, they were forced to break the face down into component parts (i.e., they described the eyes, nose, mouth, etc.). This may have allowed them to localize the change in the misleading faces (i.e., the facial hair) better because their memory of the target face incorporated both a holistic and a featural representation. With both a holistic and featural representation in mind, tested subjects might have an advantage in rejecting the misleading detail in the narrative and the misinformation faces in the lineup. Without the initial test, the memory for the face might have been preserved in a holistic format that made localization of change difficult. Indeed, because of the short duration of the critical event and the incidental encoding instructions, this holistic representation of the perpetrator may have been particularly difficult to recover for the nontested subjects.

Concluding Remarks

Attempts were made to make the current study resemble real-life eyewitness situations. For example, the eyewitness event was something that a student may actually witness—a theft in the library. Moreover, the format of the questioning was analogous to how an investigator might question a witness (i.e., free recall followed by probing questions). However, some aspects of the study may limit the generalizability of the results. For example, the delay was only 30 min between initial retrieval and the final recall/recognition test. Such a short delay is unlikely for real-life eyewitness events. In fact, it may take weeks or even months for police investigators to apprehend a suspect and put together a lineup. Future experiments may examine how a longer delay affects

lineup identification in the current paradigm. But because testing tends to protect against long-term forgetting, it is likely that greater testing benefits would emerge (see Chan & Langley, 2011, among others).

In sum, the current study provided further evidence of verbal facilitation. More importantly, the current study uncovered a new and surprising finding given RES: initial testing *reduces* suggestibility for faces. This was true in both recall and in biased-lineup identifications. The present study has identified an important boundary condition for RES. Though, one needs to be cautious about the generality of this boundary condition because it is unclear whether the effect would persist with different instructions for the audio narrative, and for a longer witnessed event with intentional encoding instructions.

Footnotes

[1] The principal component analysis was used to develop a single WMC score based on all the WMC tasks combined.

[2] In a biased lineup procedure, witnesses are not told that the perpetrator may or may not be included in the lineup and are not told that choosing no person is an option.

Reference

Belli, R. F. (1989). Influences of misleading postevent information: Misinformation interference and acceptance. *Journal of Experimental Psychology: General, 118,* 72-85.

Brewer, N., & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness idenfication: Effects of lineup instructions, foil similarity and target-absent base rates. *Journal of Experimental Psychology: Applied, 12,* 11-30.

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

Brown, C., Gehrke, J., & Lloyd-Jones, T. J. (2010). A visual and semantic locus to beneficial effects of verbalization on face memory. *American Journal of Psychology, 123,* 51-69.

Chan, J. C. K., & Langley, M. (2011). Paradoxical effects of testing: Retrieval enhances both accurate recall and suggestibility in eyewitnesses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, in press.*

Chan, J. C. K., Thomas, A. K., & Bulevich, J. B. (2009). Recalling a witnessed event increases eyewitness suggestibility: The reversed testing effect. *Psychological Science, 20,* 66-73.

Chin, J. M., & Schooler, J. W. (2008). Why do words hurt? Content, process, and criterion shift accounts of verbal overshadowing. *European Journal of Cognitive Psychology, 20,* 396-413.

Christiaansen, R. E., Sweeney, J. D., & Ochalek, K. (1983). Influencing eyewitness descriptions. *Law and Human Behavior, 7,* 59-65.

Christie, D. F., & Ellis, H. D. (1981). Photofit constructions versus verbal descriptions of faces. *Journal of Applied Psychology, 66,* 358-363.

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

Davies, G. M., & Valentine, T. (2007). Facial composites: Forensic utility and psychological research. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Handbook of Eyewitness Psychology: Vol. 2. Memory for people* (pp. 59-83). Mahwah, NJ: Erlbaum.

Dunning, D., & Perretta, S. (2002). Automaticity and eyewitness accuracy: A 10-to-12 second rule for distinguishing accurate from inaccurate positive identifications. *Journal of Applied Psychology, 87,* 951-962.

Dysart, J. E., & Lindsay, R. C. L. (2007). Show-up identifications: Suggestive technique or reliable method? In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Handbook of Eyewitness Psychology: Vol. 2. Memory for people* (pp. 137-153). Mahwah, NJ: Erlbaum.

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

Fisher, R. P., & Craik, F. I. M. (1977). Interaction between encoding and retrieval operations in cued-recall. *Journal of Experimental Psychology: Human Learning and Memory, 3,* 701-711.

Glisky, E. L., Polster, M. R., & Routhieaux, B. C. (1995). Double dissociation between item and source memory. *Neuropsychology, 9,* 229-235.

Greene, E., Flynn, M. S, & Loftus, E. F. (1982). Inducing resistance to misleading information. *Journal of Verbal Learning and Verbal Behavior, 21,* 207-219.

Hardt, O., Einarsson, E. O., & Nader, K. (2010). A bridge over troubled water: Reconsolidation as a link between cognitive and neuroscientific memory research traditions. *Annual Review of Psychology, 61*, 141-167.

Jaschinski, U. & Wentura, D. (2002). Misleading postevent information and working memory capacity: An individual differences approach to eyewitness memory. *Applied Cognitive Psychology, 16,* 223-231.

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

Johnson, M. K., Hashroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114,* 3-28.

Kurby, C. A. & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences, 12,* 72-79.

Leippe, M., & Eisenstadt, D. (2007). Eyewitness confidence and the confidence-accuracy relationship in memory for people. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Handbook of Eyewitness Psychology: Vol. 2. Memory for people* (pp. 377-426). Mahwah, NJ: Erlbaum.

Lindsay, D. S. (1990). Misleading suggestions can impair eyewitnesses' ability to remember event details. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 1077-1083.

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

Loftus, E. F. (1979a). Reactions to blatantly contradictory information. *Memory & Cognition, 7,* 368-374.

Loftus, E. F. (1979b). The malleability of memory. *American Scientist, 67,* 312-320.

Loftus, E. F., & Greene, E. (1980). Warning: Even memory for faces may be contagious. *Law and Human Behavior, 4,* 323, 334.

Loftus, E. F. (1989). Misinformation and memory: The creation of new memories. *Journal of Experimental Psychology: General, 118,* 100-104.

Loftus, E.F., Miller, D.G., & Burns, H.J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 19-31.

Luus, C. A., & Wells, G. L. (1994). The malleability of eyewitness confidence: Co-witness and perseverance effects. *Journal of Applied Psychology, 79,* 714-724.

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

Malpass, R. S., & Devine, P.G. (1981). Eyewitness identification: Lineup instructions and the absence of the offender. *Journal of Applied Psychology, 66,* 482-489.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6,* 255-260.

Marche, T. A. (1999). Memory strength affects reporting of misinformation. *Journal of Experimental Child Psychology, 73,* 45-71.

McCloskey, M., & Zaragoza, M. (1985). Misleading postevent information and memory for events: Arguments against memory impairment hypotheses. *Journal of Experimental Psychology: General, 114,* 1-16.

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

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

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

Meissner, C. A., Sporer, S. L., & Schooler, J. W. (2007). Person descriptions as eyewitness evidence. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Handbook of Eyewitness Psychology: Vol. 2. Memory for people* (pp. 1-34). Mahwah, NJ: Erlbaum.

Meissner, C. A., Sporer, S. L., & Susa, K. J. (2008). A theoretical review and meta-analysis of the description-identification relationship in memory for faces. *European Journal of Cognitive Psychology, 20,* 414-455.

Memon, A., & Rose, R. (2001). Identification abilities of children: Does a verbal description hurt face recognition? *Psychology, Crime, & Law, 0,* 1-14.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16,* 519-533.

Roediger, H. L., & Geraci, L. (2007). Aging and the misinformation effect: A neuropsychological analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 321-334.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long term-retention. *Psychological Science, 17,* 249-255.

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,* 497-528.

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

Searcy, J., Bartlett, J. C., & Memon, A. (2000). Influence of post-event narratives, line-up conditions and individual differences on false identification by young and older eyewitnesses. *Legal and Criminal Psychology, 5,* 219-235.

Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology, 9,* 641-656.

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

Steblay, N., Dysart, J., Fulero, S., & Lindsay, R. C. L. (2001). Eyewitness accuracy rates in sequential and simultaneous lineup presentations: A meta-analytic comparison. *Law and Human Behavior,* 25, 459-473.

Tanaka, J. W., & Farah, M. J. (2003). The holistic representation of faces. In M.A. Peterson & G. Rhodes (Eds.), *Perception of faces, objects and scenes* (pp. 53-740). Oxford, England: Oxford University Press.

Thomas, A. K., Bulevich, J. B., Chan, J. C. K. (2010). Testing promotes eyewitness accuracy with a warning – Implications for retrieval enhanced suggestibility. *Journal of Memory and Language, 63,* 149-157.

Tulving, E., & Watkins, M.J. (1974). On negative transfer: Effect of testing one list on the recall of another. *Journal of Verbal Learning and Verbal Behavior, 13,* 181-193.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37,* 498-505.

Van Koppen, P. J., & Lochun, S. K. (1997). Portraying perpetrators: The validity of offender descriptions by witnesses. *Law and Human Behavior, 21,* 661-685.

Weber, N., Brewer, N., Wells, G. L., Semmler, C., & Keast, A. (2004). Eyewitness identification accuracy and response latency: The unruly 10-12-second rule. *Journal of Experimental Psychology: Applied, 10,* 139-147.

Wells, G. L., & Bradfield, A. L. (1998). "Good, you identified the suspect:" Feedback to eyewitnesses distorts their reports of the witnessing experience. *Journal of Applied Psychology, 83,* 360-376.

Wells, G. L., Charman, S. D., & Olson, E. A. (2005). Building face composites can harm lineup identification performance. *Journal of Experimental Psychology: Applied, 11,* 147-156.

Wells, G. L., & Hasel, L. E. (2007). Facial composite production by eyewitnesses. *Current Directions in Psychological Science, 16,* 6-10.

Wells, G. L., Lindsay, R. C. L., & Ferguson, T. J. (1979). Accuracy, confidence, and juror perceptions in eyewitness identification. *Journal of Applied Psychology, 64,* 440-448.

Wells, G. L., Olsen, E. A., & Charman, S. D. (2002). Eyewitness identification confidence. *Current Directions in Psychological Science, 11,* 151-154.

Wilford, M. M. & Wells, G. L. (2010). Does facial processing prioritize change detection? Change blindness illustrates costs and benefits of holistic processing. *Psychological Science, 21,* 1611-1615.

Wright, D. B. (1993). Misinformation and warnings in eyewitness testimony: A new testing procedure to differentiate explanations. *Memory, 1,* 153-166.

Yu, C. J., & Geiselman, R. E. (1993). Effects of constructing identi-kit composites on photospread identification performance. *Criminal Justice and Behavior, 3,* 280-292.

Zaragoza, M. S., McCloskey, M., and Jamis, M. (1987). Misleading postevent information and recall of the original event: Further evidence against the memory impairment hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 36-44.

Appendix A

Demographic questionnaire used in Experiments 1 and 2.

1. How old are you?
2. What is your sex?
3. What is your race/ethnicity?
4. How many years of secondary education have you completed (including your current year)?

Appendix B

Free recall instructions for the initial test in both experiments and the final test in Experiment 1.

You will now type a detailed description of the man in the video that stole the wallet and laptop. Please try and be as detailed as possible in your description. Try and describe his clothing, hair, eyes, nose, mouth, and any other distinguishing features. You will have 5 minutes to type in your description. If you finish before the 5 minutes is up, try to picture him in your mind and see if you can remember any other details. If you have any questions, please ask your experimenter now. Otherwise, please let the experimenter know that you are ready to begin.

Appendix C

Cued recall questions and the correct answers. If answer has more than one correct response, only one was necessary to be scored as correct. Only question 8 was scored for misinformation recall.

1. What color was the man's shirt? [Correct Answer: Black]
2. What color were the man’s pants? [Correct Answer: Gray]
3. Approximately how tall was he? [Correct Answer: 5’10’’; response must be within 2 inches to be scored as correct]
4. Approximately how much did he weigh? [Correct Answer: 165 pounds; response must be within 10 pounds to be scored as correct]
5. What color was his hair? [Correct Answer: Brown, dirty blonde]
6. Briefly describe his hairstyle. [Correct Answer: Short, buzz-cut]
7. What shape would you say his face was? (i.e., round, oval, square, heart-shaped, etc.) [Correct Answer: Oval]
8. Describe any distinguishing characteristics that he may have had. [Correct Answer: No distinguishing characteristics; Misinformation: facial hair, mole]
9. Briefly describe his eyes (i.e., color, size, shape). [Correct Answer: Blue]
10. Briefly describe his nose (i.e., size, shape). [Correct Answer: Thin bridge, rounded tip]
11. Briefly describe his mouth (i.e., shape, thickness). [Correct Answer: Thin to average lips]
12. Describe any other details you can remember about what he looks like. [Item was not scored]

Appendix D

The audio narrative scripts used in the present experiments. All three narratives were used in Experiment 1. Only the Control Narrative and the Facial Hair Narrative were used in Experiment 2. The misinformation is in italics.

Control Narrative

"A man walked into the room after a student had left and stole the student's laptop and wallet. He was wearing a long-sleeved, black shirt and gray cargo pants. He was of average build. He was about 5 foot 10 and weighed approximately 165 pounds. I got a pretty good look at his face. He had short, brown hair and blue eyes. He had no real distinguishing features. He had a pretty average face."

Facial Hair Narrative

"A man walked into the room after a student had left and stole the student's laptop and wallet. He was wearing a long-sleeved, black shirt and gray cargo pants. He was of average build. He was about 5 foot 10 and weighed approximately 165 pounds. I got a pretty good look at his face. He had short, brown hair and blue eyes. *His only real distinguishing feature was that he had some facial hair—just some hair on his chin.* He had a pretty average face."

Mole Narrative

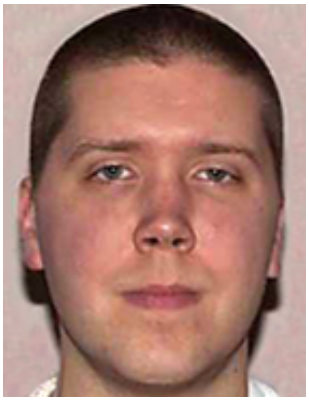
"A man walked into the room after a student had left and stole the student's laptop and wallet. He was wearing a long-sleeved, black shirt and gray cargo pants. He was of average build. He was about 5 foot 10 and weighed approximately 165 pounds. I got a pretty good look at his face. He had short, brown hair and blue eyes. *His only real distinguishing feature was that he had a mole on his left cheek.* He had a pretty average face."

Appendix E

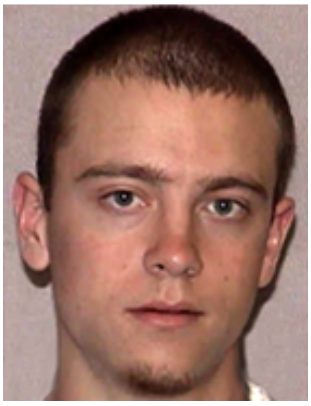
Sample target-present lineup and instructions for the lineup identification task in Experiment 2.

“Type the number of the photo of the person that you believe to be the man who stole the wallet and laptop in the video. If none of the photos match the person you saw, type "N" for none.”

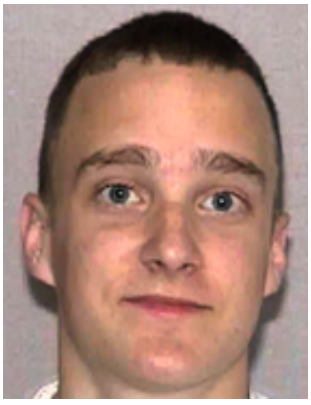
1



2

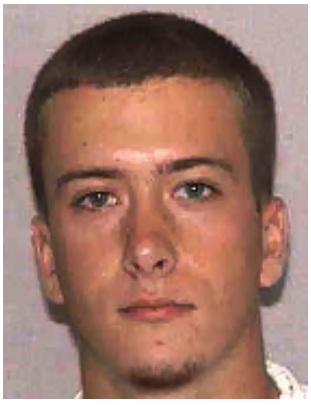


3



4

5



6

Appendix F

Foil photos used in the lineup identification task in Experiment 2. The top row includes the original photos; the bottom row includes the altered photos.

Appendix G

Follow-up questions from Experiment 2. Bracketed words are a reference to the items in Tables 8 and 12.

1. How good of a view did you get of the perpetrator? Rate from 1 (*very poor*) to 9 (*very good*) [**Goodness of view**]
2. How many seconds would you estimate that the perpetrator's face was in view? [**Seconds target's face was in view**]
3. How well were you able to make out specific features of his face from the video? [**Ability to make out facial features**]
4. What would you estimate was the distance (in feet) between the camera-eye view and the perpetrator's face? [**Distance from camera**]
5. How much attention were you paying to his face while viewing the video? Rate from 1 (*none*) to 9 (*my total attention*). [**Attention paid to target's face**]
6. At the time that you identified the person in the photospread, how certain were you that the person you identified from the photos was the man that stole the wallet and laptop from the video? Rate your confidence from 0 - 100% confident. [**Confidence in lineup decision**]
7. How easy or difficult was it for you to figure out which person in the photos was the perpetrator? Rate from 1 (*extremely easy*) to 9 (*extremely difficult*). [**Difficulty of the lineup**]
8. After you were first shown the photos, how long do you estimate it took you to make an identification? Rate from 1 (*I needed almost no time to pick him out*) to 9 (*I had to look at the photos for a long time to pick him out*). [**Time spent making a decision**]
9. On the basis of your memory, how willing would you be to testify in court that the person you identified was the man in the video? Rate your willingness from 0 - 100% willing to testify. [**Willingness to testify**]
10. Assume that an eyewitness had about the same view of the perpetrator that you had from the video. Do you think that an identification by this eyewitness ought to be trusted? Rate from 1 (*definitely should not be trusted*) to 9 (*definitely should be trusted*). [**Trust in another witness' decision**]

11. To what extent do you feel that you had a good basis (enough information) to make an identification? Rate from 1 (*no basis at all*) to 9 (*a very good basis*). [**Basis to make a decision**]