Never forget a face: Verbalization facilitates recollection as evidenced by flexible responding to contrasting recognition memory tests

Dawn R. Weatherford [1,2] · Mitchell A. Meltzer [3,4] · Curt A. Carlson [2] · James C. Bartlett [3]

Abstract

Verbal facilitation occurs when describing a face improves its subsequent recognition; but there are several theoretical explanations debated in the literature. The results of the present studies support a relatively unrestricted, parsimonious theory that verbal facilitation occurs because describing a face supports recollection of several different facets of the face-viewing experience. This recollection is then demonstrated by flexibly responding to two competing types of recognition task demands. Participants studied a list of faces and, following each face, performed a nonverbalization task (Experiment 1) or described its features or traits (Experiment 2). Two subsequent recognition tests included intact faces, new faces, and conjunctions (each of which recombined features of two studied faces). *Inclusion* test instructions emphasized featural information: respond "yes" to both intact and conjunction faces (both of which contained studied features), but "no" to new faces. *Exclusion* test instructions emphasized configural information: respond "yes" only to intact faces (which were the only test items that matched studied configurations), and "no" to both conjunctions and new faces. Both yes/no responses and confidence ratings supported our hypothesis that verbalization improved discrimination between (a) conjunctions and new faces in the inclusion test, and (b) intact faces and conjunctions in the exclusion test. Additional secondary responses about face type elucidated that verbalization at study improves the ability to recollect either featural or configural information, depending on which type of response the recognition test required. We discuss these findings about practical applications of improved face memory in real-world contexts.

Keywords Verbal facilitation · Facial description · Recognition memory · Recollection · Conjunctions

Verbal descriptions play a vital role in many criminal cases, where eyewitnesses need to provide an account of the events of the crime and the criminal. Putting aside information about transient aspects of the criminal (e.g., style of clothes), describing the perpetrator's face may influence the police's ability to apprehend a suspect. Findings across many different types of stimuli, however, support that the act of facial description can fundamentally alter memory (e.g., Nakabayashi, Burton, Brandimonte, & Lloyd-Jones, 2012a).

Although describing a face can produce memory interference (e.g., Meissner, Sporer, & Susa, 2008; Mickes & Wixted,

2015), opposing evidence supports that some situations produce facilitative effects (e.g., Brown, Gehrke, & Lloyd-Jones, 2010; Brown & Lloyd-Jones, 2005). For example, old/new recognition is improved when participants verbally describe each face during and/or immediately after its presentation on a study list compared with when they perform a control task (e.g., counting backwards). However, the source of this verbalization benefit is still a matter of debate. In the interest of more fully understanding the mechanisms of this verbal facilitation effect, so that they may be exercised in practicable circumstances, such as eyewitness memory preservation shortly after a crime, we examined the effects of description on both encoding and retrieval of faces in a traditional facial recognition paradigm.

Verbalization facilitates recollection

Theorists have attributed verbal facilitation—wherein recognition memory is improved when faces are followed by a description task rather than a nondescription task—to (1) distributed visual scanning during initial face learning, which

✉ Dawn R. Weatherford
    dweatherford@tamusa.edu

1   Department of Science & Mathematics, Psychology Program, 1
    University Way, Texas A&M University-San Antonio, San
    Antonio, TX 78224, USA

2   Texas A&M University-Commerce, Commerce, TX, USA

3   University of Texas at Dallas, Richardson, TX, USA

4   University of Texas at Arlington, Arlington, TX, USA

enhances processing and memory of diagnostic facial *features* (Brown et al., 2010; Nakabayashi, Lloyd-Jones, Butcher, & Liu, 2012b; Winograd, 1981); (2) enhanced processing and/or retrieval of some form of "*configural*," "holistic," or "global"[1] information (Brown et al., 2010; Brown & Lloyd-Jones, 2005); (3) beneficial effects of *semantic* elaboration (Bower & Karlin, 1974; Brown & Lloyd-Jones, 2006); and (4) enhanced processing that improves *recollection* of information about the face-viewing experience in a subsequent test (Brown et al., 2010; Brown & Lloyd-Jones, 2006). No single account is empirically favored, and they are not mutually exclusive. Here, we test a more unrestricted, novel hypothesis that combines aspects of several accounts by more clearly examining how *any and all* of these types of information might be used during a recognition memory test.

Our proposal assumes that recognition memory relies on two processes—familiarity and recollection. Familiarity is a general sense of resemblance between a test item and a previously seen stimulus that varies only in strength. Crucially, it provides no contextual information about an encoding event (e.g., how you were feeling when you first encountered a stimulus). On the other hand, recollection, when elicited above a certain threshold by a test item (cf. Mickes, Wais, & Wixted, 2009; Slotnick & Dodson, 2005), provides contextual information about when a stimulus was previously encountered (e.g., Heathcote, Raymond, & Dunn, 2006). By this interpretation, we view recollection as the conscious reexperiencing of a prior event, and assume that one important consequence of such conscious reexperiencing is to enhance control of "old" and "new" judgments in recognition memory tests (e.g., Moscovitch, 2008; Yonelinas & Jacoby, 2012). Specifically, in tests of face recognition, we assume that recollection can be used to control the occurrence of false recognition errors in response to new faces that are perceived as familiar due to their strong resemblance to previously studied faces.

The idea is not that recollection reduces the perceived familiarity of high-resemblance new faces. Rather, we propose that recollection allows the rejection of such faces through a conscious comparison to recollected information, *despite* the high resemblance. This view of recollection is in line with recent neuropsychological conceptions, which hold that the hippocampus works as a "comparator" in the detection of novelty in stimuli that, despite being new, evoke activation of previously encoded information based upon their informational overlap (Kumaran & Maguire, 2007; Olsen, Moses, Riggs, & Ryan, 2012).

According to our interpretation, verbalization improves the ability to selectively recollect useful information about the

previous face-viewing experience (e.g., Brown et al., 2010; Brown & Lloyd-Jones, 2006), depending upon the task demands. Recollected details can include any and all of the information as described in the multiple types of description-identification theories outlined above: featural, configural, semantic, and descriptive. Under such conditions that recollecting any and all of these types of information is advantageous, verbalization shortly after encoding allows observers to use all manner of information flexibly, in such a way that meets task demands.

Poststimulus verbalization

To explore this more expanded recollection hypothesis, we started by considering a previous study in which Jones, Armstrong, Casey, Burson, and Memon (2013) examined the effects of *poststimulus* verbalization (i.e., the effects of verbally describing a face after, as opposed to during, its presentation). Their approach has the benefit of controlling for verbalization effects on perceptual encoding processes occurring when a face is in view (e.g., subvocal descriptions, hedging). After the face is removed from view, the researcher can cue what type of judgment should be made so as to prevent participants from using goal-oriented attentional mechanisms to encode the face in a manner consistent with the type of judgment they are making (i.e., focusing on features when they know they need to make a featural judgment). In their study, participants viewed a single block of study faces, one at a time, and were either prompted to verbally describe or perform a nondescription control task (either counting backwards or visualization) after each face. The recognition test included studied (intact) faces, entirely new faces, and conjunctions that recombined the inner and outer features of two studied faces. Notably, each intact face was both a featural and configural match with a previously studied face, whereas conjunctions were a featural match in the absence of a configural match (Bartlett, Shastri, Abdi, & Neville-Smith, 2009; Jones & Bartlett, 2009). Likewise, new faces had neither studied features nor configurations.

In the first Jones et al. (2013) experiment, verbalization increased correct recognition of intact faces, but had no effect on false recognition of conjunctions, which suggested that verbalization increased intact-conjunction (IC) discrimination. Intact faces and conjunctions are both featural matches to studied faces. However, only intact faces are configural matches to studied faces. Therefore, IC discrimination can be considered as a metric to measure the use of configural information. In contrast, conjunction-new (CN) discrimination can be viewed as a metric to measure the use of featural information, as both conjunctions and new faces lack a configural match to any studied face, but only conjunctions contain previously

---

[1] Although we acknowledge that these three terms are not universally agreed to be interchangeable, we use these terms in similar ways throughout the paper for ease of reader understanding and brevity.

studied features. Because IC discrimination, but not CN discrimination, was affected by verbalization, the authors argued that conjunction faces did not elicit recollection in such a way that would reduce conjunction false alarms because conjunctions are poor cues for recollection (Shapiro & Penrod, 1986).

To conceptually replicate their results and combat the possibility that the backwards-counting task harmed face recognition, leading to perceived verbalization benefits in contrast, participants in a follow-up experiment only *visualized* each study face. Results supported that the visualization and nondescription counting task conferred a similar, and not beneficial, impact on memory. Both experiments led the authors to propose that poststimulus verbalization facilitates recollection of the verbal description itself, thereby increasing the correct recognition of intact faces without affecting false recognition of conjunctions. Subsequently, this *description recollection* proposal can be viewed as more restricted version of the recollection hypothesis we described above, with two stipulations added. First, only verbal information is recollected. Second, such information is not recollected in response to conjunctions.

A more recent study used a similar method to support the same hypothesis about the role of verbalized content (Jones, Robinson, & Steel, 2018). Across five studies, the authors investigated how facial description (which should theoretically increase recollection and familiarity) would interact with facial repetition (which should theoretically only increase familiarity; e.g., Bruce, Dench, & Burton, 1993; Jones & Bartlett, 2009). Participants either viewed or described faces that they had studied either one or two times, followed by a recognition test using 1–6 confidence-based recognition decisions. As a direct measure of recollection, participants who indicated having previously studied a face (i.e., responded 4, 5, or 6) were additionally prompted to type any descriptive contents that they could recall. When participants received study-task cues (either view or describe) after the face was removed from view, description benefits were less pronounced than when study-task cues appeared before the face was presented. The authors hypothesized that this effect was due to a combination of hedging (i.e., a strategic attempt to covertly describe each face during its short presentation, in preparation for typing a description when prompted) and configural enhancement brought about by preparing to describe (i.e., a shift from passive to active visual processing in preparation for typing a description when prompted; see also, Nakabayashi, Lloyd-Jones, et al., 2012b). While Jones et al. (2013) argued that hedging would affect verbalization, and thus recollection, configural enhancement may only exert its influence on familiarity.

Using recollection to flexibly respond to recognition test instructions

We saw no reason to question the Jones et al. (2013; Jones et al., 2018) results, and expected to replicate the verbalization benefit on memory for intact faces in the present research. However, in both sets of studies, participants were instructed to only respond "yes" to previously studied faces if they were a featural and configural match to a studied face. For an old/new recognition paradigm (e.g., Jones et al., 2018), those directions are pretty straightforward. However, for an old/conjunction/new paradigm, participants are unable to demonstrate any lingering memory for a featural match in the absence of a configural match (as would be revealed by CN discrimination) if they are always instructed to reject conjunctions.

In the current set of studies, we sought to more closely examine task performance when participants are required to use recollection to accept conjunctions as previously studied. Our findings could then provide additional evidence with which to reinterpret the predictions of the *description recollection* account, which argues that conjunction faces are poor cues for recollecting the verbalized content upon which verbal facilitation relies. By combining two types of instructions (i.e., exclusion and inclusion) in a conjunction face paradigm, we can more closely examine how verbalization might allow participants to flexibly use recollected information, even in response to conjunctions. *Exclusion* instructions, as were used in Jones et al. (2013), instruct participants to recognize only intact faces as old, while rejecting both conjunctions and new faces as new. By contrast, *inclusion* instructions, which were not adopted by Jones et al. (2013), instruct participants to recognize both intact faces and conjunctions as old, while rejecting only new faces as new. Essentially, exclusion and inclusion task demands vary only in how participants are told to respond to conjunctions. Therefore, to flexibly respond to exclusion versus inclusion instructions, participants must use recollection to consciously identify conjunctions as recombinations of parts of studied faces. However, since a conjunction was never described as a single contextual unit, recollecting descriptive contents may not necessarily be as helpful in rejecting or accepting a conjunction (depending upon test instructions).

In the absence of verbalization, findings of one study (Jones & Bartlett, 2009) suggest that participants are largely incapable of responding to these two different types of test instructions in a flexible way. Participants in Jones and Bartlett (2009) studied a list of faces, without any verbalization task, and then made recognition decisions under *inclusion* and *exclusion* instructions. Importantly, Jones and Bartlett found no reliable difference in the pattern of "old" judgments between the two instructional conditions. Participants in both conditions displayed the typical conjunction paradigm pattern: they responded "old" more often to intact faces than to

conjunctions, and more often to conjunctions than to new faces. This failure to adhere to the differing exclusion/inclusion task demands suggests that familiarity, rather than recollection, supports conjunction responses to faces under nonverbalization conditions.

Although exclusion instructions emphasize recollecting configural information and inclusion instructions emphasize recollecting featural information, actually meeting those demands requires a more robust encoding experience that allows for such flexibility (as opposed to an experience that only increases familiarity strength without sufficient recollection of contextual details). Our primary assertion is that the flexibility is not contingent upon the exact verbalization experience. Instead, participants encode various types of information as a function of the description task. It is the use of that information that is flexible. When the task demands featural information, verbalization makes such information accessible. Similarly, when the task demands configural information, verbalization allows the same type of accessibility.

In other words, simply remembering only featural, configural, semantic, or descriptive information (as suggested by various interpretations of the previous theoretical accounts currently debated in the literature) would be insufficient to perform optimally on both the inclusion and exclusion intact/conjunction/new recognition tests. Instead, we propose an *expanded recollection* account: Verbalization facilitates recollection *of many aspects of the face-viewing experience* by supporting the recognition of studied faces as intact and, *when task demands require it*, the rejection of conjunction faces as new. Further, our recollection hypothesis is not contingent upon what participants prepare to verbalize (e.g., hedging) or actually verbalize during the poststimulus task.

The current research

Before testing this hypothesis directly, we aimed to replicate the results of Jones and Bartlett (2009), who found no marked differences between exclusion and inclusion task performance, by using a more direct measure of recollection against which the effects of verbalization can be examined. We wanted to strengthen our position that recollection, as opposed to alternative explanations, would be responsible for accepting or rejecting a conjunction in line with test instructions when such cues were available. In our initial experiment, we predicted that a backwards counting task (which Jones et al., 2013, confirmed produced similar outcomes as visualization), would not promote recollection in this way. Instead, we expected to find responses that were largely based on familiarity.

Because many studies suggest that recognition and metacognitive memory processes adapt in response to differing task demands (e.g., Diana & Reder, 2004), Experiment 1

tested participants' performance in response to (a) a nondescription (i.e., backwards counting) task at study and (b) exclusion and inclusion instructions that required up to three different types of responses for each face. As in Jones and Bartlett (2009), participants made yes/no recognition decisions (hereafter referred to as primary yes/no responses) to intact, conjunction, and new faces under either exclusion or inclusion instructions. Next, participants rated the confidence in their primary recognition responses. Last, we added a third type of response to this paradigm that encouraged participants to interrogate their memory to make decisions that ostensibly relied on recollection (hereafter referred to as secondary item-type responses). In our modified paradigm, when a participant responded "no" to a face under exclusion instructions or "yes" to a face under inclusion conditions (regardless of their confidence rating), they were prompted to identify the face as conjunction or new (in the exclusion condition) and intact or conjunction (in the inclusion condition). Presumably, only recollection—namely, the *conscious* discrimination of conjunctions from other types of test items—would allow participants to correctly recognize a conjunction as a recombination of studied face parts.

Our predictions for this initial experiment are guided by previous conjunction paradigms adopted in the absence of verbalization (e.g., Jones & Bartlett, 2009; Reinitz, Lammers, & Cochran, 1992). Regarding primary yes/no responses, we predicted that participants would demonstrate the classic pattern of intact > conjunction > new as indicated by proportion of "yes" responses and d' (a measure of discriminability from which we can calculate IC and CN discrimination). However, we did not expect a Test (inclusion, exclusion) × Face Type (intact, conjunction, new) interaction that would suggest the use of recollection to flexibly respond to task demands. Regarding primary yes/no responses and confidence ratings, we predicted that receiver operating characteristic (ROC) curves and their corresponding area under the curve (AUC) would again support participants' ability to use familiarity to distinguish between intact, conjunction, and new faces. Statistically, however, we expected no significant differences in responding across varying levels of confidence by test. Lastly, we predicted that participants' ability to identify each face type with secondary item-type responses would suggest little to no involvement of recollection. Instead, familiarity should facilitate identification of entirely new faces (for which memory strength should be weak or absent), but that distinguishing between intact faces and conjunctions should be more challenging. However, because no paradigm of which the authors are aware has used this additional secondary item-type response, we saw it as an opportunity to establish a baseline for this more precise type of response against which performance can be compared with chance (if participants

were guessing among the three face types without any useful memories upon which to draw) and subsequent verbalization trials (to be explored in Experiment 2).

Experiment 1

Method

Participants

Native English-speaking undergraduate students (N = 36) from Texas A&M University–Commerce participated in the experiment (M_{age} = 24.31 years; 24 females). As evidence suggests that race and verbalization independently influence memory (Nakabayashi, Lloyd-Jones, et al., 2012b), we did not select based on participant race. Self-reported race reflected a diverse sample (six Black/African American, 24 White/Caucasian, three Hispanic/Latino, one Asian/Pacific Islander, and two other). All participants reported normal or corrected-to-normal vision.

Design

We held the nondescription task (backwards counting) performed by all participants constant and varied test face type (intact, conjunction, new) and test instructions (inclusion, exclusion) completely within participants.

Materials

Participants viewed front-facing grayscale photographs (see Fig. 1) of the head and tops of the shoulders of Caucasian males and females (taken, with permission, from Bartlett, 2009). Pairs of individuals of the same gender and approximate age comprised four-face sets of two original and two recombination faces. Original faces represented unaltered versions of photographed persons. Each recombination face was composed of the interior region of one original face and the exterior region of another original face (i.e., internal facial features such as eyes, nose, and mouth were always kept intact and together).

Both original and recombination faces appeared in the study and test phases, but varied across participants. For example, if a participant studied two recombination faces from the same four-face set, the test may either include those same two recombination faces as intact faces or the two unstudied original faces from the same four-face set as conjunctions. To reduce any effects driven by the stimuli themselves, another participant would have studied the original faces from the same set, and then see a test that may have represented the same original faces as intact or unstudied recombination faces as conjunctions. This fully counterbalanced design ensured

that each stimulus served equally often as an intact, conjunction, and new face across all participants. Within participants, no stimulus was presented as multiple face types (i.e., if a participant was shown an intact face as test, no recombination of its features was ever presented as another test face).

Procedure

After providing informed consent, participants were instructed that they would view a series of faces, one at a time, and perform a backwards counting task between each face's presentation. During the backwards counting task, participants viewed a randomly selected three-digit number and mentally counted backwards in intervals of three for 25 s, followed by a 5-s period to type the number to which they counted. After four practice trials, the study phase included 36 trials (two primacy and two recency buffers with 32 critical faces). Each face was displayed for 2 s, preceded by a 2.5-s white screen. After the study phase, participants performed a Sudoku puzzle task for 5 minutes.

After the puzzle task, participants performed two self-paced recognition tests, each with a different set of instructions. Before reading instructions for either test, participants viewed pictorial examples taken from the practice trials and/or unstudied stimuli with statements about the nature and composition of each of the three face types (all instructions adapted from Jones & Bartlett, 2009). The intact face example was an exact match to a face studied in a practice trial. The conjunction example included two practice trial faces beside their recombination. The new face example was an unstudied face.

Following the examples, tests were blocked such that instructions directly preceded each respective test, with order counterbalanced across participants. The inclusion test instructed participants to respond "yes" to both intact and conjunction faces and "no" to new faces. The exclusion test instructed participants to respond "yes" to intact faces and "no" to conjunction and new faces. Each test contained two buffer, eight intact, eight conjunction, and eight new faces. Within participants, no face appeared in both tests. For each trial, participants made a yes/no response by pressing the "y" and "n" keys followed by a confidence rating on a scale of 1 to 3, with 1 labeled *not at all confident* and 3 labeled *very confident*.[2] After the confidence rating, participants in the inclusion condition were probed to identify each face they had accepted with a "yes" decision as intact or conjunction. Similarly, participants in the exclusion condition were probed to identify each face they had rejected with a "no" decision as

---

[2] Although some studies adopt a blended response of decision and confidence (e.g., respond from 1 to 6 with 1 representing *definitely no* and 6 representing *definitely yes*), we saw no reason to consider the cognitive mechanisms associated with a split approach to be any different. Further, as we asked participants to make multiple types of confidence/decision judgments, we wanted to avoid confusion between judgments.

Fig. 1 Example four-face set where features from the two original faces in the top row were recombined to create two additional faces in the bottom row. If participants studied the two faces in the top row, either of the faces in the bottom row would be the conjunction. However, if participants studied the two faces in the bottom row, either of the faces in the top row would be the conjunction

conjunction or new. Upon completion of the test, participants provided demographic information, were debriefed, thanked, and dismissed.

Results and discussion

Prior to data analysis, we excluded data from $n = 3$ participants for failing to comply with instructions (e.g., typing no or obviously incorrect numeric responses during the counting trials of the study phase, responding "yes" to all faces during the test phase). The final sample size for analysis was $n = 33$. We approached the data in two ways by analyzing (1) the primary yes/no judgments and confidence ratings, and (2) the secondary item-type judgments that followed "yes" decisions under inclusion and "no" decisions under exclusion, as a more direct measure of recollection.

Recognition accuracy by response type

Primary yes/no responses

Proportion of "yes" responses. Before detailing our more analytically nuanced approach, we started by considering the mean proportion of "yes" responses by face type and

instructions (see Table 1). As is typically observed in the conjunction paradigm literature, we found a main effect of face type, $F(2, 31) = 54.16$, $p < .001$, $\eta_\rho^2 = .78$. Follow-up analysis confirmed that participants responded "yes" more often to intact than to conjunction and new faces, and more often to conjunction than to new faces. Also replicating previous findings, we observed a main effect of test type, $F(1,32) = 73.41$, $p < .001$, $\eta_\rho^2 = .70$. Follow-up analyses confirmed that participants followed instructions by responding "yes" more often in the inclusion than exclusion condition. Critically, however, we did not observe an interaction, $F(2, 31) = .70$, *ns*.

Discriminability. Having replicated the classic intact > conjunction > new pattern using the proportion of "yes" responses, we next considered values that more clearly differentiated between the roles of featural and configural processing in recognition responses. As established in previous work with conjunction face paradigms (Jones et al., 2013; Jones & Bartlett, 2009), we treated participants' ability to distinguish between intact and conjunction faces (i.e., IC discrimination) as a measure of configural memory. In contrast, we treated participants' ability to distinguish between conjunction and new faces (i.e., CN discrimination) as a measure of featural memory. We examined IC and CN discrimination based on

Table 1 Experiment 1 means and standard deviations for response proportions by test and face type

	Inclusion			Exclusion		
Primary yes/no responses						
	Count		Unstudied	Count		Unstudied
Test face		*Yes*			*Yes*	
Intact	**.72(.18)**			**.45(.22)**		
Conj	**.61(.21)**			.39(.18)		
New			.35(.18)			.14(.12)
Secondary item-type responses						
Test Face	*Intact*	*Conj*	*New*	*Intact*	*Conj*	*New*
Intact	**.34(.21)**	.37(.17)	.28(.18)	**.45(.23)**	.35(.20)	.19(.18)
Conj	.18(.14)	**.43(.23)**	.39(.21)	.39(.18)	**.37(.18)**	.24(.18)
New	.11(.12)	.24(.14)	**.65(.18)**	.14(.13)	.35(.19)	**.51(.21)**

Note. To ease comparison, proportion of correct responses are in boldface. For primary responses, proportion of "no" responses equal 1 minus proportion of "yes" responses. For secondary responses, proportions within a given test face category sum to 1, within rounding

yes/no recognition data (regardless of confidence level) using d' scores, where zero indicates chance levels of discrimination between test face types.

As shown in Fig. 2, d' was higher for CN compared with IC discrimination, as evidenced by a main effect of discrimination type, $F(1, 32) = 10.47$, $p = .003$, $\eta_\rho^2 = .25$, and this pattern was nearly identical for the exclusion and inclusion tests, as evidenced by no main effect of test type or interaction between test type and discrimination type.

Primary yes/no responses and confidence judgments

Receiver operating characteristic curves. After successfully replicating previous patterns with primary yes/no responses, we incorporated participants' confidence ratings to reveal additional patterns. Previous verbalization studies have established that criteria may shift across a range of dimensions

Fig. 2 Mean d' scores for intact-conjunction discrimination and conjunction-new discrimination in Experiment 1. Error bars represent 95% confidence intervals calculated using the Cousineau–Morey correction for within-subject comparisons (see Cousineau, 2005; Morey, 2008)

in response to the verbalization task (e.g., Chin & Schooler, 2008; Sauerland, Holub, & Sporer, 2008; Wilson, Seale-Carlisle, & Mickes, 2017) and awareness that conjunction faces were included as test items (Reinitz & Loftus, 2017). However, ROC analysis allows us to straightforwardly measure discriminability that is not conflated with response bias (e.g., Gronlund, Wixted, & Mickes, 2014; Mickes, 2015). ROC curves may be constructed by varying the x-axis and y-axis to support comparisons for both intact-conjunction (IC) discrimination, a measure of configural processing, as well as conjunction-new (CN) discrimination, a measure of featural processing.

To create our ROC curves, we converted the yes/no and 3-point confidence ratings participants made to each test item into a 1–6 scale in which 1 represented a *no* response made with high confidence and 6 represented a *yes* response made with high confidence. The proportions of "1," "2," "3," "4," "5," and "6" responses made by each participant were converted to recognition rates at five different criterion levels from highest (proportions of "6" responses), to next highest (cumulative proportions of "6" and "5" responses), and so on through the lowest (cumulative proportions of "6," "5," "4," "3," and "2" responses). These recognition rates were then calculated for each item type and plotted as ROC curves representing aggregate intact-conjunction (IC) and conjunction-new (CN) discrimination (see Fig. 3).

Exclusion and inclusion CN ROCs are approximately equidistant from the diagonal line representing chance-level discrimination, suggesting that CN discrimination was unaffected by test instructions. This outcome replicates a key finding of Jones and Bartlett (2009), who focused on CN discrimination in their ROC analysis, and suggests that participants are inflexible in responding to exclusion versus inclusion task demands with upright faces. While the IC ROCs also support
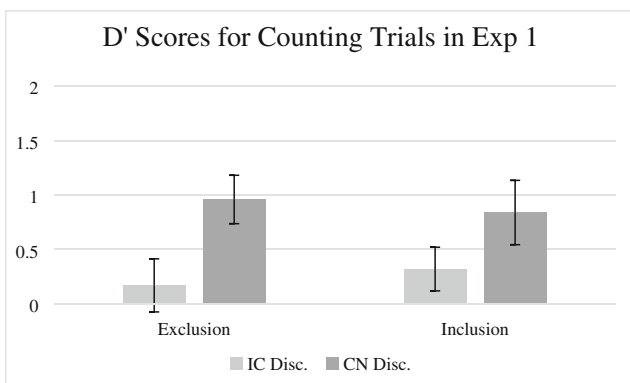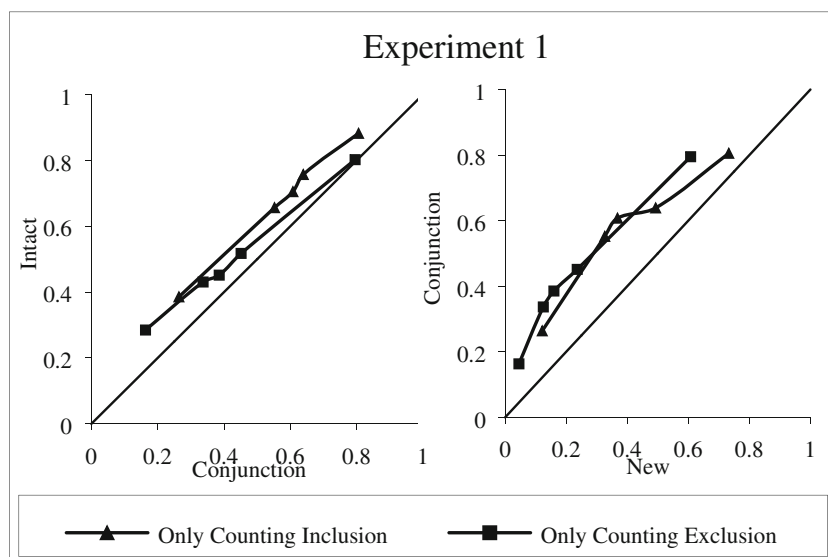
Fig. 3 Receiver operating characteristics (ROCs) in Experiment 1 contrasting intact and conjunction (IC discrimination) and conjunction and new (CN discrimination) performance for inclusion and exclusion tests

this conclusion, performance was quite low on both the exclusion and inclusion tests and the resulting possibility of a floor effect makes these data more challenging to interpret.

Area under the curve. To quantitatively evaluate these observations, we calculated area-under-the-curve (AUC) scores. Although several methods of calculating these scores exist, some involve extrapolating the left-most and right-most data points of each ROC to the 0/0 and 1/1 coordinates on the plot. Among other things, this is problematic for directly comparing two ROCs, which do not perfectly overlap along the x-axis, as different degrees of extrapolation are needed for each (e.g., Ma, Bandos, Rockette, & Gur, 2013). Therefore, we used the pROC toolbox (Robin et al., 2011) in R to compute partial AUC scores for exclusion discrimination and inclusion discrimination within subportions of the aggregate ROCs that overlapped along the x-axis of the ROC plots. Crucially, neither IC discrimination (exclusion: .55, inclusion: .63) nor CN discrimination (exclusion: .67, inclusion: .65) differed across test types (both ps > .10). This pattern is consistent with the pAUCs, and was supported by a 2 × 2 factorial repeated-measures analysis of variance (RMANOVA; Test Type [exclusion, inclusion] × Measure [IC discrimination, CN discrimination]), which revealed only a significant main effect of measure such that CN discrimination exceeded IC discrimination across test types, $F(1, 32) = 10.47$, $MSE = 14.16$, $p = .003$, $\eta_\rho^2 = .25$.

Secondary item-type responses

The exclusion/inclusion equivalence in IC and CN discrimination suggests that recollection contributes little to configural and featural recognition for faces. However, in search of more direct evidence for this conclusion, we examined participants' accuracy using the secondary item-type responses. We had two goals in mind with these data: (1) characterize participants' accuracy as a baseline, in the absence of any verbal description task, and (2) generate item-type accuracy against which we could compare our counting and description trials in Experiment 2.

To characterize participants' accuracy, we first began by considering chance-level performance for the secondary item-type responses as they differed from primary yes/no responses. Despite the instructions varying between inclusion and exclusion, the ratio of intact, conjunction, and new faces remained equally weighted across test types. In other words, if participants were merely guessing on the primary responses, a "yes" response would be correct 66% of the time for the inclusion test and 33% of the time for the exclusion test. However, for the secondary item-type responses, participants would have a 33% chance of producing the correct response for any of the three face types, regardless of the test instructions. Table 1 includes the proportion of all responses by face type.

For intact faces, participants correctly recognized 34% ($SD = 29\%$) on the inclusion test and 45% ($SD = 22\%$) on the exclusion test. Only exclusion test accuracy exceeded the chance value of 33%; $t(32) = 3.15$, $p = .004$, Cohen's $d = .52$. For new faces, participants correctly recognized 65% ($SD=18\%$) on the inclusion test and 55% ($SD = 28\%$) on the exclusion test. Both types of accuracy exceeded chance: inclusion, $t(32) = 10.17$, $p < .001$, Cohen's $d = 1.77$; exclusion $t(32) = 4.43$, $p < .001$, Cohen's $d = .78$. These values establish that participants, even in the absence of verbalization, were able to use memory cues to successfully identify entirely new faces and, at least under exclusion instructions, intact faces.

Having established basic old/new accuracy, we considered responses to conjunctions, upon which our predictions about the use of recollection chiefly rely. Crucially, participants correctly recognized 43% (SD = 23%) of conjunctions on the inclusion test and 37% (SD = 18%) of conjunctions on the exclusion test. Statistically, only the inclusion recognition rate exceeded chance performance, $t(32) = 2.59$, $p = .014$, Cohen's $d = .43$. Like the exclusion/inclusion data from the primary responses, these values do not provide strong evidence for the involvement of recollection in configural and featural information to identify faces.

Discussion

Consistent with previous findings (Jones & Bartlett, 2009), Experiment 1 revealed that participants were unable to flexibly follow contrasting exclusion versus inclusion task instructions under nonverbalization conditions. Furthermore, their performance on the secondary item-type responses suggest that although participants were able to identify new faces as such, findings for intact faces and conjunctions were more mixed and dependent upon test instructions. Collectively, the results suggest difficulty in using configural and featural information to make the primary and secondary responses. Nevertheless, participants did not exhibit floor effects that may obscure our ability to detect meaningful patterns in the data for this and future experiments.

Having established the utility of this new paradigm while also replicating previous findings under nonverbalization conditions, we turned to the research questions at hand: Will poststimulus verbalization at the time of study improve performance using these modified inclusion and exclusion instructions? Further, will verbalization enable subsequent recollection of configural and featural information about these faces depending on the test type, regardless of different types of description tasks at study? If both answers are "yes," then two patterns should emerge. First, poststimulus verbalization will increase discrimination between intact faces and conjunctions (i.e., IC discrimination) in the exclusion test, replicating the findings of Jones et al. (2013). Second, poststimulus verbalization will increase discrimination between conjunctions and new faces (i.e., CN discrimination) in the inclusion test, a finding that previous literature has not addressed.

Results to support these predictions can be derived from each of the three responses (primary yes/no, confidence rating, and secondary item-type) by poststimulus task (description or backwards counting). We expect that backwards counting trials should largely follow the pattern observed in Experiment 1, in that backwards counting did not confer a robust ability to flexibly respond to test instructions in such a way as to produce interactions. However, for description trials, we expected to see such interactions emerge. For primary yes/no responses,

this prediction would be supported by a significant interaction between face type and test type for the proportion of "yes" responses for d' for trials followed by the description, but not backwards counting, task. For primary yes/no responses and confidence ratings, ROCs and AUCs should support differences by discrimination type and test type, again with patterns for IC and CN discrimination being selectively affected by the description task under such conditions that recollection supported accurately responding to competing test instructions. Because secondary item-type responses encouraged participants to identity one of the three face types as such, we expected a different pattern of results than the primary yes/no responses. Specifically, we did not make strong predictions about interactions by test instructions here. Instead, we predicted main effects that would support verbal facilitation by way of superior identification performance for intact faces and conjunctions, regardless of test type, as these stimuli that would benefit the most strongly from the recollection facilitated by the description task.

Finally, we addressed an additional question concerning whether verbalization benefits may depend on descriptive content and attentional mechanisms guided by the verbalization instructions. Nearly all of the theoretical accounts of verbal facilitation assert some special role for either featural or configural information/processing, especially as facilitation may be distinguished from overshadowing (e.g., Fallshore & Schooler, 1995; Nakabayashi, Lloyd-Jones, et al., 2012b; Winograd, 1976). Subsequently, several studies have assessed the verbalization of facial features versus personality traits suggested by the faces (e.g., Brown et al., 2010; Brown & Lloyd-Jones, 2006; Wickham & Lander, 2008). However, no previous studies have compared verbalization instructions using the poststimulus verbalization paradigm introduced by Jones et al. (2013). To rectify this gap in the literature and more fully explore alternatives to the expanded recollection account, half of our participants were instructed to describe traits, while the rest were instructed to describe features. Our expanded recollection account would not make predictions about a selective benefit of one description type over the other. Instead, it would predict that both types of description would confer benefits to test performance following largely the same pattern of results.

Experiment 2

Method

Participants

Native English-speaking undergraduate students ($N = 73$) from Texas A&M University–Commerce participated in the experiment ($M_{age} = 25.07$ years; 52 female). Self-reported race

reflected a diverse sample (13 Black/African American, 44 White/Caucasian, nine Hispanic/Latino, four Asian/Pacific Islander, and three other). All participants reported normal or corrected-to-normal vision.

Design and procedure

We explored the effect of description by including an additional within-participants factor of poststimulus task (describe vs. count) during the study phase. Except for the description task, all other elements were identical to Experiment 1.

Participants were randomly assigned to the featural or trait description condition. Participants in the featural description condition were instructed to "describe each face in terms of its features, such as eyes, nose, and mouth." Participants in the trait description condition were instructed to "describe each face in terms its personality type, honesty, and attractiveness." Both groups were encouraged to "be as descriptive as possible" in the time allowed. To control for processes independent of the description task itself (e.g., preparing to describe, as in Jones et al., 2013), participants viewed a single block of faces and performed either a description or counting task after each face (see also, cue-after condition; Jones et al., 2018, interpretation saved for the General Discussion).

Results

We excluded data from $n = 4$ participants ($n = 2$ featural, $n = 2$ trait) who failed to comply with instructions (e.g., typing no or obviously incorrect numeric/verbal responses during the counting or description trials of the study phase, responding "yes" to all faces during the test phase), resulting in a final sample size of $n = 69$. Additionally, we pooled the data from both description conditions (featural vs. trait) for the following analyses unless otherwise noted. This facilitated comparison with the results of Jones et al. (2013), who gave their participants no direction as to the content of their descriptions at encoding. For analyses collapsed across description type, we followed the same analytic structure as in Experiment 1.

Recognition accuracy by response type, comparing description to counting trials

Primary yes/no responses

Proportion of "yes" responses. As in Experiment 1, we first considered the mean proportion of "yes" responses by all face types (intact, conjunction, new) and test types (inclusion, exclusion; see Table 2). We again observed a main effect of face type, $F(2, 67) = 246.36$, $p < .001$, $\eta_\rho^2 = .88$, and a main effect of test type, $F(1, 68) = 259.34$, $p < .001$, $\eta_\rho^2 = .79$. However, unlike Experiment 1, we also observed an

interaction, $F(2, 67) = 10.68$, $p < .001$, $\eta_\rho^2 = .24$. Importantly, this two-way interaction (which included new faces for which participants did not perform a poststimulus task) was qualified by a three-way interaction between test type (inclusion, exclusion), face type (intact, conjunction), and poststimulus task (description, counting), $F(1, 68) = 9.10$, $MSE = .42$, $p = .004$, $\eta_\rho^2 = .12$. While promising, we scrutinized the meaning of this interaction further by reincorporating new faces using d' values as a more sensitive measure of discriminability.

Discriminability. We calculated d' scores for IC discrimination and CN discrimination for each participant (see Fig. 4). Turning first to the exclusion data, IC discrimination for description trials exceeded that for counting trials. By contrast, CN discrimination was virtually unchanged across the trial types. The inclusion data, on the other hand, showed the opposite pattern—higher levels of CN discrimination for description compared with counting trials with no corresponding effect on IC discrimination. Consistent with this pattern, a $2 \times 2 \times 2$ factorial ANOVA (Test Type [exclusion vs. inclusion] × Poststimulus Task [description vs. counting] × Discrimination Type [intact conjunction vs. conjunction new]) revealed a significant three-way interaction, $F(1, 68) = 7.45$, $MSE = 18.99$, $p = .008$, $\eta_\rho^2 = .10$. Simple effects analyses revealed that description increased IC discrimination, but only in the exclusion condition: exclusion, $t(68) = 2.53$, $p = .014$; inclusion, $t(68) = 0.58$, $p = .566$, and increased CN discrimination, but only in the inclusion condition: exclusion, $t(68) = 0.55$, $p = .582$, inclusion, $t(68) = 4.68$, $p < .001$. These corresponding effects on IC and CN discrimination, which were limited to tests where each was encouraged, suggest that verbalization facilitates recollection of both configural *and* featural information in face memory.

Primary yes/no responses and confidence judgments

Receiver operating characteristic curves. Group-level aggregated ROCs are shown in Fig. 5. Consistent with our predictions, the aggregate exclusion IC ROC for description trials lies above the one for counting trials, indicating that verbalization increased exclusion IC discrimination and replicating a key finding of Jones et al. (2013). By contrast, for the inclusion test, the counting trials IC ROC lies slightly above, but close to, the IC ROC for the description trials, indicating little effect of trial type on IC discrimination.

Also consistent with our predictions, the aggregate inclusion CN ROC for the description trials lies above the one for the counting trials, indicating that verbalization increased inclusion CN discrimination as well. By contrast, for the exclusion test, the CN ROCs for the description and counting trials

Table 2 Experiment 2 means and standard deviations for proportion of responses by test, face, and task

	Inclusion			Exclusion		
Primary yes/no responses						
	Count	Describe	Unstudied	Count	Describe	Unstudied
Test Face		*Yes*			*Yes*	
Intact	**.77(.21)**	**.91(.14)**		**.49(.30)**	**.64(.25)**	
Conj	**.51(.27)**	**.72(.24)**		.29(.25)	.29(.22)	
New			.29(.17)			.13(.16)
Secondary item-type responses						
Count						
Test Face	*Intact*	*Conj*	*New*	*Intact*	*Conj*	*New*
Intact	**.47(.30)**	.29(.25)	.23(.21)	**.49(.29)**	.38(.28)	.13(.18)
Conj	.18(.21)	**.32(.22)**	.49(.21)	.29(.25)	**.41(.27)**	.30(.27)
Describe						
Intact	**.54(.29)**	.37(.28)	.08(.14)	**.64(.25)**	.24(.22)	.11(.16)
Conj	.29(.25)	**.43(.27)**	.28(.24)	.29(.22)	**.43(.26)**	.27(.26)
Unstudied						
New	.05(.08)	.24(.16)	**.72(.17)**	.13(.16)	.26(.19)	**.61(.25)**

Note. To ease comparison, proportion of correct responses are in boldface. For primary responses, proportion of "no" responses equal 1 minus proportion of "yes" responses. For secondary responses, proportions within a given test face category sum to 1, within rounding

lie virtually on top of each other, indicating similar levels of CN discrimination for the two poststimulus tasks.

Area under the curve. Supporting these observations, pAUC was greater for description trials than for counting trials for exclusion IC discrimination (description: .74, counting: .63), $p < .001$, but not for inclusion IC discrimination (description: .72, counting: .77), $p = .309$. Likewise, pAUC was greater for description trials than for counting trials for inclusion CN discrimination (description: .78, counting: .61), $p < .001$, but not for exclusion CN discrimination (description: .52, counting: .53), $p = .915$. This represents the predicted discrimination Type × Test Type interaction, wherein description

facilitates IC discrimination on the exclusion test but CN discrimination on the inclusion test.

Secondary item-type responses

Having established a baseline in Experiment 1, which confirmed that participants do have *some* ability to perform the identification task under nondescription conditions, we were interested in considering how the poststimulus description task may improve identification performance. A central distinction between the primary yes/no responses and the secondary-item type responses is that primary yes/no response are contingent upon test instructions. However, identification
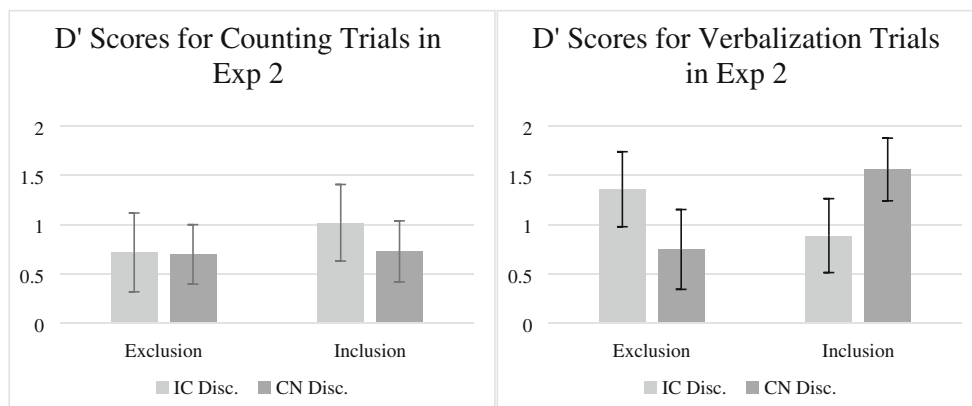
Fig. 4 Mean *d'* scores for intact-conjunction discrimination and conjunction-new discrimination in Experiment 2. Error bars represent 95% confidence intervals calculated using the Cousineau–Morey correction for within-subject comparisons (see Cousineau, 2005; Morey, 2008)
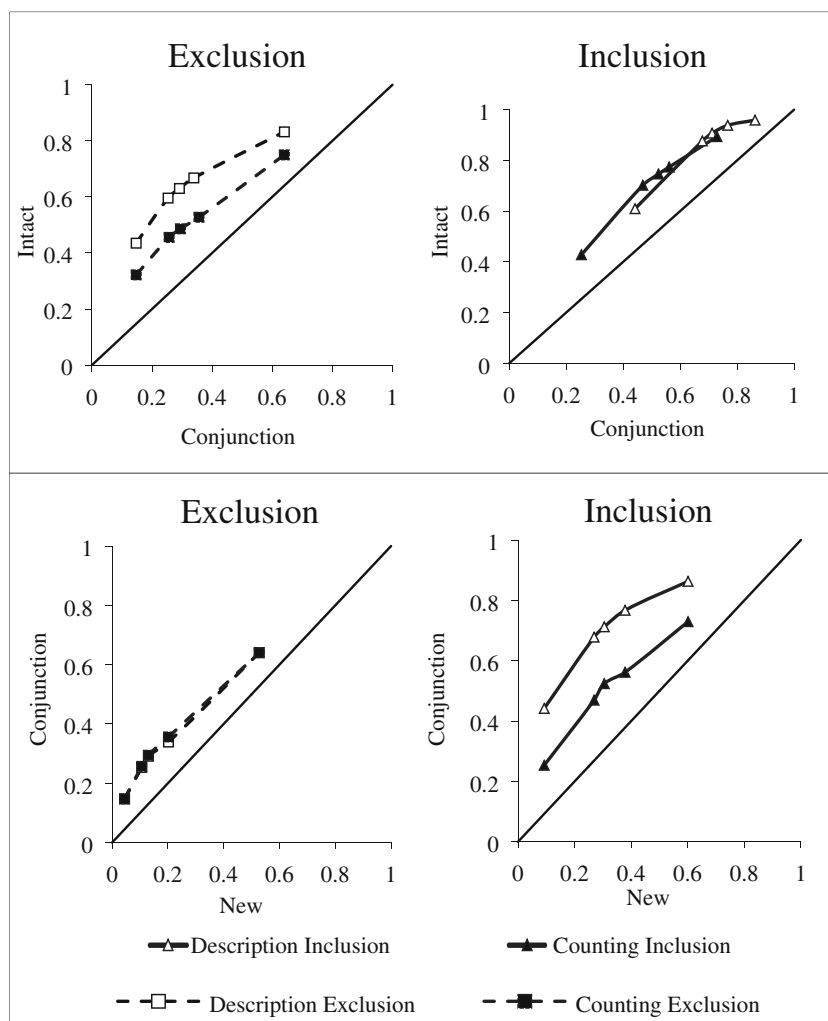
Fig. 5 Receiver operating characteristics (ROCs) in Experiment 2 contrasting intact and conjunction (IC discrimination) and conjunction and new (CN discrimination) performance for inclusion and exclusion tests

by item-type was not. In other words, participants in both types of tests would need to successfully identify each of the three face types as such in these secondary item-type responses, regardless of whether the face appeared in the inclusion or exclusion test.

Therefore, we subjected the identification rates for intact and conjunction faces, respectively, to a 2 × 2 factorial ANOVA (Test Type [exclusion vs. inclusion] × Poststimulus Task [description vs. counting]).[3] For intact faces, we observed no main effect of test type $F(1, 68) = 2.21$, ns, a main effect of poststimulus task, $F(1, 68) = 15.43$, $MSE = .90$, $p < .001$, $\eta_\rho^2 = .19$, but no interaction, $F(1, 68) = 2.12$, ns. As expected, these results confirm verbal facilitation by way of higher identification rates for intact

faces that were followed by description than by counting, which was not contingent upon test type. For conjunctions, we again observed no main effect of test type $F(1, 68) = 2.47$, ns, a main effect of poststimulus task, $F(1, 68) = 4.71$, $MSE = .29$, $p = .033$, $\eta_\rho^2 = .07$, but no interaction, $F(1, 68) = 2.21$, ns. Although conjunction identification rates were admittedly weaker than those of intact faces, as would be the case for the arguably more difficult task of identifying conjunctions, we still see evidence of verbal facilitation. Both findings provide continued support for our expanded recollection account.

Recognition accuracy differences by description type

Finally, as mentioned above, there is considerable interest in whether description content moderates the effects of verbalization on face memory (e.g., Brown et al., 2010; Brown & Lloyd-Jones, 2006). Therefore, we reanalyzed our data by taking into account whether participants described feature or trait attributes of faces in the poststimulus task. In the interest

---

[3] Since new faces were unstudied, and therefore had no poststimulus task, we did not include them in these analyses. However, descriptive values demonstrate improved performance compared with the new face identification rates of Experiment 1.

of brevity, we have included descriptive information about our analyses here. The interested reader can access supplementary materials including exact verbal descriptions by participant and stimulus, description coding, figures, and tables on the open source website listed below.

Accuracy as predicted by description quality and quantity. We engaged in a rigorous, multirater verbalization coding process. Despite high interrater reliability and sufficient within and between participant variability, we found no significant relationship between any of the descriptors (by either type or amount) and accuracy on any of the three responses (primary, confidence rating, or secondary).

Discriminability by description condition. The data are strikingly similar across content types, and, consistent with this observation, description condition (trait vs. featural) did not qualify the Poststimulus Task × Test Type × Discrimination Type interaction reported above, $F(1, 67) = .01$, $MSE = .01$, $p = .95$, $\eta_\rho^2 < .001$, for the four-way interaction, when entered into the model. Conjunction recognition accuracy was also quite similar across test types, except that it was above chance on the exclusion test for counting trials in the trait condition, $t(33) = 2.59$, $p = .014$, but not in the featural condition, $t(34) = 0.92$, $p = .365$. Substantiating this observation, description condition did not qualify the Test Type × Poststimulus Task interaction reported above, $F(1, 67) = .94$, $MSE = .05$, $p = .34$, $\eta_\rho^2 = .01$, for the three-way interaction.

ROCs and pAUCs by description condition. We separately aggregated exclusion and inclusion IC and CN ROCs for participants in the feature and trait description conditions. For both description conditions, the description trial exclusion IC ROC lies above that for the counting trials, particularly towards the right of the plot, indicating a description > counting advantage in IC discrimination. By contrast, the inclusion IC ROCs for the description and counting trials lie close to one another, indicating little impact of encoding condition on such discrimination. Likewise, the description trial inclusion CN ROC lies above that for the counting trials indicating a description > counting advantage in CN discrimination, yet no such pattern is present in the exclusion data, where the CN ROC virtually lie on top of each other. All these observations were supported quantitatively by an analysis of pAUC scores. Collectively, the pAUC score, d', and conjunction recognition accuracy data suggest little to no influence of description content on verbal facilitation in face memory.

Discussion

Across all three types of responses (primary yes/no, confidence rating, and secondary item type), our results support the expanded recollection account. Although the data corresponding to the poststimulus backwards counting task were somewhat similar to those of Experiment 1 (especially when considering the ROC/AUC values that capture discriminability without conflating it with response bias), the poststimulus description task produced different patterns. Participants were consistently able to capitulate to test demands. When the test instructions encouraged acceptance of conjunctions, verbalization allowed for such flexibility. When it discouraged such acceptance, verbalization allowed for that, too. Nevertheless, secondary item-type responses suggest that participants did so intentionally, using recollection, as opposed to simply basing all responses merely upon familiarity.

Having adopted this new paradigm, the question remains whether simply asking participants to make the item-type judgments, without preceding those instructions with inclusion and exclusion primary yes/no responses, would have produced a different set of results. Our data between the two experiments suggest that when memory is weaker, as would be the case when participants primarily rely on familiarity strength to make their judgments, the secondary-item type responses are influenced by the primary yes/no instructions (as evidenced by interactions between face type and test type in Experiment 1). However, the poststimulus description task (whether feature or trait) provided a stronger inoculation against such reactivity (as evidenced by no such interaction). Future research could explore that simpler variant of the task, and the extent to which it supports poststimulus verbal facilitation. The predictions of the expanded recollection account would remain the same: Verbalization would facilitate participants' ability to follow different recognition test instructions. For the time being, our data across these three different response types allowed us bridge the broader literature by replicating previously observed patterns with yes/no responses from old/conjunction/new paradigms (Jones, Bartlett, & Wade, 2006; e.g., McKone & Peh, 2006; Reinitz & Loftus, 2017), ROCs that combine yes/no responses and confidence ratings (e.g., Jones & Bartlett, 2009; Meltzer & Bartlett, 2019), and extending to clarify predictions and theory regarding the source of verbal facilitation (Jones et al., 2013).

General discussion

Across two experiments, our data support that describing faces (in terms of features or traits) improves recognition accuracy that relies on featural or configural information, depending upon the test instructions. In Experiment 1, participants who did not describe any faces, and instead performed a

counting task after each face, were unable to capitulate to task demands by flexibly responding under inclusion and exclusion instructions for a recognition test composed of intact, conjunction, and new faces. In contrast, participants in Experiment 2, who performed a description or counting task after each study face, demonstrated their recognition flexibility in two ways. First, participants showed markedly higher CN discrimination for faces followed by description than counting under inclusion instructions. Further, participants showed markedly higher IC discrimination for faces followed by description than counting under exclusion instructions. Secondary item-type responses under both types of test instructions similarly confirmed that participants' recollection of face type (intact, conjunction) was enhanced for faces followed by the description task.

These findings contribute to an important, but fractured, literature regarding the description-identification relationship. Previous studies have appealed to verbalization enhancing facial recognition by way of reliance upon featural information (e.g., visual scanning; Winograd, 1981), holistic/configural information (Nakabayashi, Lloyd-Jones, et al., 2012b), semantic elaboration (e.g., Bower & Karlin, 1974; Brown et al., 2010), or memory for verbal information (e.g., description-recollection hypothesis; Jones et al., 2013; Jones et al., 2018). Our interpretation of the current findings needs no such specific claims. Instead, our data support a broader conclusion that is not limited by the mode (visual or verbal), type (featural or configural), or amount (single judgments vs. multiple statements) of recollected information. This more parsimonious approach shows the relatively unrestricted benefit of facial description even when the participant is unaware of what type of poststimulus task (description or counting) or recognition test (inclusion or exclusion) may be performed.

Different measures of recollection influence different uses of recollected information

Admittedly, conjunction faces are never encountered in real-world contexts. As an internally valid tool, conjunction faces allow researchers to test associative memory strength under such circumstances when facial recombinations can be controlled. Realistically, though, it is not uncommon for two different individuals to look incredibly similar yet distinct from one another based on featural similarities (e.g., same nose shape) set inside unique configurations (e.g., differing distances between nose and ears). These facial recombinations, therefore, provide a valuable tool for testing the degree to which facial information is represented in memory and used to make recognition decisions. In an exclusion/inclusion paradigm, conjunction faces allowed us to make and support specific predictions about the role of recollected featural and configural information, without having to rely on verbal recall at test (e.g., Jones et al., 2018). However, if we were to apply a

testing format that prompted participants to type their verbal recall after making a positive recognition decision, our predictions and interpretations would remain unchanged. Our expanded recollection account holds that participants may not experience descriptive recollection in response to conjunctions, but that verbal facilitation might still be demonstrated through the ability to accept conjunction faces under inclusion test instructions.

Similarly, the results of Experiment 2 do not directly speak to hedging, as we did not cue participants to recall any of their verbal contents. However, we find evidence that describing, and not just preparing to describe, produces a benefit that cannot be explained with the description-recollection hypothesis alone: description allowed participants to recognize facial recombinations with featural similarities and unique configurations under such circumstances when it was encouraged (i.e., inclusion), and also reject them under such circumstances when it was discouraged (i.e., exclusion). When the test instructed participants to use information other than the description content on which to make their decision, as was the case in Jones et al. (2018), they were able to successfully do so. The same logic that supports that recollection would increase the descriptive recall in a follow-up judgment would also support that recollection would increase face-type identification (i.e., intact, conjunction) in a follow-up judgment.

Nevertheless, future studies would need to more closely examine these possibilities to unify the fractured state of theory in the description-identification relationship. These studies can approach this problem by (a) combining approaches we have adopted with measures of recollection to more directly contrast predictions restricted to one type of recollected information (e.g., verbal information) to more unrestricted types of recollected information (e.g., any type of contextual information from the face-viewing experience that contributes to successful recognition decisions), and (b) measuring verbalization benefits using more externally valid paradigms.

Bringing verbal facilitation benefits to the real world

In real-world settings, it would be considered rare to ask someone to remember several faces in a series that are closely related and described in time. Often, individuals would need to apply description benefits to one important face, such as when an eyewitness views a suspect committing a crime or when a person meets a previously unknown individual at a social gathering. However, for the description to be beneficial, it must provide recollective flexibility. Suppose, for instance, that an eyewitness describes a suspect's face shortly after the commission of a crime, without constraining such description to features or perceived traits. For memory to be valuable during a lineup or showup, the eyewitness's memory must contain enough fidelity to recognize the perpetrator (as a function of enhanced memory strength by way of description), but

not misidentify similar-looking individuals who might bear strong resemblance to the perpetrator in some, but not all, ways (as would be the case if memory strength, but not precision, were enhanced uniformly).

To our knowledge, only one verbal facilitation paradigm has ever tested memory for a single face using both target-present and target-absent lineups (Sporer, Kaminski, Davids, & McQuiston, 2015). Participants who reread their descriptions prior to a lineup administration showed the dual benefit of both higher hit rates and lower false alarm rates compared with participants who did not provide a description. Crucially, description quality was positively related to lineup performance. When participants were prompted to provide correct information, in which they held high confidence, those descriptions likely provided stronger contextual cues to recollection.

Instruments and techniques to elicit these types of valuable descriptions that support recollection have been developed. The person description interview (PDI; Demarchi & Py, 2009) encourages participants to adopt strategies that have been known to distinguish between verbal facilitation and overshadowing (Meissner et al., 2008). In their field study, Demarchi and Py (2009) asked participants to provide a description of a single target person after interacting face-to-face or viewing a photograph. In comparison to a control *standard interview* prompt to "Please describe the person you saw. Try to be as complete as possible," participants in the experimental condition were encouraged to adopt a strategy of *specific to general* or *general to specific* (regarding the level of detail in physical and perceived characteristics) and/or *top to bottom* or *bottom to top* (regarding the spatial arrangement of featural information). Both the detail specificity and spatial arrangement strategies increased the amount of correct information and decreased the number of errors in the descriptions compared with control.

Together, Sporer et al. (2015) and Demarchi and Py (2009) may shed some light on a major distinction between verbal facilitation and verbal overshadowing. Both studies suggest that (a) description quality of a single face matters, and (b) description instructions can alter description quality. Why, then, did Jones et al. (2013) and other multiple-face studies find no significant impact of description quality? One possibility is that a multiple-face paradigm might improve memory by naturally encouraging the spontaneous description-enhancement strategies adopted in the PDI. Under such circumstances, descriptions over the course of multiple trials should increase in quality, be more distinctive and precise, and enhance recollection at test. Therefore, multiple-face paradigms that do not find an influence of description quality (e.g., Brown et al., 2010; Brown & Lloyd-Jones, 2006) may be dealing with a large proportion of high-quality descriptions, negating their influence overall. Given the appropriate circumstances, many forms of

contextual information can serve to improve eyewitness memory in theory and practice.

Avoiding the ugly side of the description-identification relationship

For various reasons, verbal overshadowing has received relatively more attention in the literature than verbal facilitation (Alogna et al., 2014; Meissner & Brigham, 2001; Mickes & Wixted, 2015; Wilson et al., 2017), perhaps because it more strongly fits with the increasingly prevalent narrative of the unreliable eyewitness. For instance, studies that disrupt the temporal sequence between encoding the to-be-remembered face (e.g., during a mock crime video or facial recognition sequence) and producing the description are more likely to produce deleterious verbalization effects (e.g., Meissner & Brigham, 2001; Meissner, Brigham, & Kelley, 2001; Schooler, Fiore, & Brandimonte, 1997). Under such circumstances, delayed verbal descriptions are likely based on retrieval from long-term memory, as opposed to generating descriptions from active representations in working memory that have been neither delayed nor contaminated by interference (see also retrospective verbal reports; Ericsson & Simon, 1980). In contrast, our data combines with other papers adopting similar paradigms and results that fit within our broader theoretical interpretation (e.g., Brown et al., 2010; Brown & Lloyd-Jones, 2005; Jones et al., 2013; Jones et al., 2018; Sporer et al., 2015). Verbal facilitation offers a mechanism to redeem eyewitnesses' memory using a description process that is essential for the police to create a lineup and apprehend a suspect. Along with the extant verbalization literature, our data suggest that facial recollection can be improved when descriptions are provided (a) shortly after the encoding period, (b) in a multiface paradigm, and (c) with attentional focus paid to unique facial attributes as opposed to generic information such as height, age, gender, and race. Accordingly, verbalization under such task conditions should be more strongly scrutinized to show the boundary conditions of the effect using more real-world stimuli and externally valid paradigms.

Compliance with ethical standards

References

Alogna, V. K., Attaya, M. K., Aucoin, P., Bahnik, S., Birch, S., Birt A. R., Bornstein, B. H., … Zwaan, R. A. (2014). Registered replication report: Schooler and Engstler-Schooler (1990). *Perspectives on Psychological Science*, *9*(5), 556–578. doi:https://doi.org/10.1177/1745691614545653

Bartlett, J. C. (2009). *Conjunction face stimuli*. University of Texas at Dallas.

Bartlett, J. C., Shastri, K. K., Abdi, H., & Neville-Smith, M. (2009). Component structure of individual differences in true and false recognition of faces. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(5), 1207–1230. doi:https://doi.org/10.1037/a0016368

Bower, G. H., & Karlin, M. B. (1974). Depth of processing pictures of faces and recognition memory. *Journal of Experimental Psychology*, *103*(4), 751–757. doi:https://doi.org/10.1037/h0037190

Brown, C., Gehrke, J., & Lloyd-Jones, T. J. (2010). A visual and semantic locus to beneficial effects of verbalization on face memory. *The American Journal of Psychology*, *123*(1), 51–69. Retrieved from http://www.jstor.org/stable. https://doi.org/10.5406/amerjpsyc.123.1.0051

Brown, C., & Lloyd-Jones, T. J. (2005). Verbal facilitation of face recognition. *Memory & Cognition*, *33*(8), 1442–1456. doi:https://doi.org/10.3758/BF03193377

Brown, C., & Lloyd-Jones, T. J. (2006). Beneficial effects of verbalization and visual distinctiveness on remembering and knowing faces. *Memory & Cognition*, *34*(2), 277–286. doi:https://doi.org/10.3758/BF03193406

Bruce, V., Dench, N., & Burton, M. (1993). Effects of distinctiveness, repetition and semantic priming on the recognition of face familiarity. *Canadian Journal of Experimental Psychology = Revue Canadienne de Psychologie Expérimentale*, *47*(1), 38–60. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8481711

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*(3), 396–413. doi: https://doi.org/10.1080/09541440701728623

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology, 1*(1), 42–45. https://doi.org/10.20982/tqmp.01.1.p042

Demarchi, S., & Py, J. (2009). A method to enhance person description: A field study. In R. Bull, T. Valentine, & T. Williamson (Eds.), *Handbook of psychology of investigative interviewing* (pp. 241–256). New York, NY: Wiley-Blackwell.

Diana, R. A., & Reder, L. M. (2004). Visual versus verbal metacognition: Are they really different? In D. T. Levin (Ed.), *Thinking and seeing: Visual metacognition in adults and children* (pp. 187–201). Cambridge, MA: MIT Press.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *87*(3), 215–251.

Fallshore, M., & Schooler, J. W. (1995). Verbal vulnerability of perceptual expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(6), 1608–1623. doi:https://doi.org/10.1037/0278-7393.21.6.1608

Gronlund, S. D., Wixted, J. T., & Mickes, L. (2014). Evaluating eyewitness identification procedures using receiver operating characteristic analysis. *Current Directions in Psychological Science*, *23*(1), 3–10. doi:https://doi.org/10.1177/0963721413498891

Heathcote, A., Raymond, F., & Dunn, J. (2006). Recollection and familiarity in recognition memory: Evidence from ROC curves. *Journal of Memory and Language*, *55*(4), 495–514. doi:https://doi.org/10.1016/j.jml.2006.07.001

Jones, T. C., Armstrong, R., Casey, A., Burson, R. A., & Memon, A. (2013). Verbal description benefits for faces when description conditions are unknown a priori. *The Quarterly Journal of Experimental Psychology*, *66*(9), 1818–1839. doi:https://doi.org/10.1080/17470218.2013.771688

Jones, T. C., & Bartlett, J. C. (2009). When false recognition is out of control: The case of facial conjunctions. *Memory & Cognition*, *37*(2), 143–157. doi:https://doi.org/10.3758/MC.37.2.143

Jones, T. C., Bartlett, J. C., & Wade, K. (2006). Nonverbal conjunction errors in recognition memory: Support for familiarity but not for feature bundling. *Journal of Memory and Language*, *55*(1), 138–155. doi:https://doi.org/10.1016/j.jml.2006.01.002

Jones, T. C., Robinson, K., & Steel, B. C. (2018). Context retrieval and description benefits for recognition of unfamiliar faces. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *44*(10), 1514–1532. doi:https://doi.org/10.1037/xlm0000537

Kumaran, D., & Maguire, E. A. (2007). Which computational mechanisms operate in the hippocampus during novelty detection? *Hippocampus*, *17*, 735–748. doi:https://doi.org/10.1002/hipo.20326

Ma, H., Bandos, A. I., Rockette, H. E., & Gur, D. (2013). On use of partial area under the ROC curve for evaluation of diagnostic performance. *Statistics in Medicine*, *32*(20), 3449–3458. doi:https://doi.org/10.1002/sim.5777

McKone, E., & Peh, Y. X. (2006). Memory conjunction errors for realistic faces are consistent with configural processing. *Psychonomic Bulletin & Review*, *13*(1), 106–111. doi:https://doi.org/10.3758/BF03193820

Meissner, C. A., & Brigham, J. C. (2001). A meta-analysis of the verbal overshadowing effect in face identification. *Applied Cognitive Psychology*, *15*(6), 603–616. doi:https://doi.org/10.1002/acp.728

Meissner, C. A., Brigham, J. C., & Kelley, C. M. (2001). The influence of retrieval processes in verbal overshadowing. *Memory & Cognition*, *29*(1), 176–186. doi:https://doi.org/10.3758/BF03195751

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*(3), 414–455. doi:https://doi.org/10.1080/09541440701728581

Meltzer, M. A., & Bartlett, J. C. (2019). Holistic processing and unitization in face recognition memory. *Journal of Experimental Psychology: General*, *148*(8), 1386–1406. doi:https://doi.org/10.1037/xge0000640

Mickes, L. (2015). Receiver operating characteristic analysis and confidence-accuracy characteristic analysis in investigations of system variables and estimator variables that affect eyewitness

memory. *Journal of Applied Research in Memory and Cognition*, *4*(2), 93–102. doi:https://doi.org/10.1016/j.jarmac.2015.01.003

Mickes, L., Wais, P. E., & Wixted, J. T. (2009). Recollection is a continuous process: Implications for dual process theories of recognition memory. *Psychological Science*, *20*(4), 509–516. doi:https://doi.org/10.1111/j.1467-9280.2009.02324.x

Mickes, L., & Wixted, J. T. (2015). On the applied implications of the "verbal overshadowing effect". *Perspectives on Psychological Science*, *10*(3), 400–403. doi:https://doi.org/10.1177/1745691615576762

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods in Psychology, 4*(2), 61–64. https://doi.org/10.20982/tqmp.04.2.p061

Moscovitch, M. (2008). The hippocampus as a "stupid," domain-specific module: Implications for theories of recent and remote memory, and of imagination. *Canadian Journal of Experimental Psychology*, *62*(1), 62–79. doi:https://doi.org/10.1037/1196-1961.62.1.62

Nakabayashi, K., Burton, A. M., Brandimonte, M. A., & Lloyd-Jones, T. J. (2012a). Dissociating positive and negative influences of verbal processing on the recognition of pictures of faces and objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(2), 376–390. doi:https://doi.org/10.1037/a0025782

Nakabayashi, K., Lloyd-Jones, T. J., Butcher, N., & Liu, C. H. (2012b). Independent influences of verbalization and race on the configural and featural processing of faces: A behavioral and eye movement study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(1), 61–77. doi:https://doi.org/10.1037/a0024853

Olsen, R. K., Moses, S. N., Riggs, L., & Ryan, J. D. (2012). The hippocampus supports multiple cognitive processes through relational binding and comparison. *Frontiers in Human Neuroscience*, *6*(146), 2–13. doi:https://doi.org/10.3389/fnhum.2012.00146

Reinitz, M. T., Lammers, W. J., & Cochran, B. P. (1992). Memory-conjunction errors: Miscombination of stored stimulus features can produce illusions of memory. *Memory & Cognition*, *20*(1), 1–11. doi:https://doi.org/10.3758/BF03208247

Reinitz, M. T., & Loftus, G. R. (2017). Conjunction faces alter confidence-accuracy relations for old faces. *Journal of Experimental Psychology: Learning Memory & Cognition*, *43*(6), 837–850. doi:https://doi.org/10.1037/xlm0000342

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Mueller, M. (2011). pROC: An open-source package for R and

S+ to analyze and compare ROC curves. *BMC Bioinformatics*, *8*, 12–77. doi:https://doi.org/10.1007/s00134-009-1641-y

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*(3), 497–528. doi:https://doi.org/10.1080/09541440701728417

Schooler, J. W., Fiore, S. M., & Brandimonte, M. A. (1997). At a loss from words: Verbal overshadowing of perceptual memories. In D. L. Medin (Ed.), *The psychology of learning and motivation: Advances in Research and Theory* (pp. 293–334). Academic Press.

Shapiro, P. N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin*, *100*(2), 139–156.

Slotnick, S. D., & Dodson, C. S. (2005). Support for a continuous (single-process) model of recognition memory and source memory. *Memory & Cognition*, *33*(1), 151–170. doi:https://doi.org/10.3758/BF03195305

Sporer, S. L., Kaminski, K. S., Davids, M. C., & McQuiston, D. (2015). The verbal facilitation effect: Re-reading person descriptions as a system variable to improve identification performance. *Memory*, 1–16. doi:https://doi.org/10.1080/09658211.2015.1106561

Wickham, L. H. V., & Lander, K. (2008). The effect of verbal description and processing type on face identification. *European Journal of Cognitive Psychology*, *20*(3), 577–586. doi:https://doi.org/10.1080/09541440701728433

Wilson, B. M., Seale-Carlisle, T. M., & Mickes, L. (2017). The effects of verbal descriptions on performance in lineups and showups. *Journal of Experimental Psychology: General*, *1*(2), 1–12. doi:https://doi.org/10.1037/xge0000354

Winograd, E. (1976). Recognition memory for faces follow nine different judgments. *Bulletin of the Psychonomic Society*, *8*, 419–421.

Winograd, E. (1981). Elaboration and distinctiveness in memory for faces. *Journal of Experimental Psychology: Human Learning & Memory*, *7*(3), 181–190. doi:https://doi.org/10.1037/0278-7393.7.3.181

Yonelinas, A. P., & Jacoby, L. L. (2012). The process-dissociation approach two decades later: Convergence, boundary conditions, and new directions. *Memory & Cognition*, *40*(5), 663–680.