BRIEF REPORT

Attempted recall of biographical information influences face attractiveness

David Grybinas[1] · **Ian G. Dobbins**[1]

Abstract

We examined affective consequences arising from the kinds of memory retrieval failures that often accompany social interaction. To do so, we measured the influence of cued-recall outcomes for biographical information on the rated attractiveness of faces. The data demonstrate that retrieval of names (Experiment 1a) and professions (Experiment 1b) increases the rated attractiveness of target faces relative to faces that failed to produce recall of associative information. This was predicted by a confirmation of search (COS) model originally developed on verbal memoranda, which assumes that confirmation bias during memory search leads to affective consequences depending upon retrieval's success or failure. The current study extends this model, showing that evaluative judgments of individuals are in part contingent upon the memory retrieval skills of their assessors. We conclude by discussing potential extensions of the COS paradigm to the measurement of implicit attitudes and special populations.

Keywords Memory · Encoding effects · Emotion · Cued recall

Consider the common scenario of encountering a recent acquaintance and trying to remember their name or some other important biographical characteristic (e.g., context of prior meeting, profession, etc.). In the current study, we examine whether the outcome of these types of retrieval attempts influences attractiveness ratings of the acquaintance—an intriguing possibility suggested by recent findings examining the effects of retrieval outcomes on the rated pleasantness of verbal memoranda. We briefly review this research next.

An early and well-replicated interaction between memory and affect is the mere exposure effect (Zajonc, 1968), in which earlier exposure to materials increases their later rated pleasantness and other measures of positive affect. Surprisingly, these increases occur even when subjects are unaware of the prior encounters, leading to the hypothesis that mere exposure effects operate implicitly (Bornstein & D'Agostino, 1992, 1994), perhaps by increasing perceptual fluency (although see Newell & Shanks, 2007). In contrast, the acquaintance

anecdote above focuses on whether the outcome of *explicit* retrieval attempts moderate affect. This was explored in four experiments by Grybinas, Kantner, and Dobbins (2019) spanning recognition, source memory, and cued recall with verbal memoranda. In the case of recognition memory, two initial experiments demonstrated that memoranda correctly (and incorrectly) recognized from prior study were rated as more pleasant than baseline norms (see also Newell & Shanks, 2007). Extending prior research, the experiments also demonstrated that memoranda correctly (and incorrectly) classified as new to the experiment were rated less pleasant than baseline norms. This devaluation accompanying correct rejections is noteworthy because stimuli correctly identified as new during recognition, as well as those rated during pleasantness norming, were both seen for the first time in the experiment context. Given this, the lowered perceived pleasantness of the memoranda during recognition must result specifically from the explicit judgment that they are not recognized. This initial set of findings led Grybinas et al. (2019) to hypothesize that it was the outcome of memory search that was moderating the rated pleasantness of the memoranda, such that when the search for memory evidence supporting the study context was confirmed, there was a positive affective response, and when it was disconfirmed, there was a negative affective response. This COS model assumes that observers misattribute the affective consequences of memory search outcomes

✉ David Grybinas
   davidgrybinas@wustl.edu

✉ Ian G. Dobbins
   idobbins@wustl.edu

1 Department of Psychological & Brain Sciences, Washington
   University in Saint Louis, Saint Louis, MO 63130, USA

(confirmation vs. disconfirmation) to the memoranda, and is motivated by the well-documented tendency of observers to favor confirmation in judgment and decision-making tasks (Nickerson, 1998). An additional, smaller effect on rated pleasantness was linked to the accuracy of judgments—namely, correct judgments led to somewhat higher pleasantness judgments than errors. This much smaller influence of accuracy was assumed to reflect the effort by which the confirmation or disconfirmation was achieved, with the greater effort accompanying errors lowering perceived pleasantness of their associated memoranda (see also Chetverikov, 2014; Lee, 2001). Thus, under the COS model, whether a retrieval search is confirmed or disconfirmed alters affective response to the memoranda, and this effect can be slightly moderated by the ease with which the confirmation or disconfirmation arises (viz., accuracy).

If memory search outcomes influence affective responses to memoranda, then the COS model should generalize to other memory paradigms. To test this, Experiments 3 and 4 of Grybinas et al. (2019) examined verbal source memory and paired-associate cued recall tasks. For source memory, the data demonstrated that equally familiar memory probes drawn from two prior sources (pronounceability and abstractness rating tasks) were rated as more or less pleasant depending on whether the source memory query was confirmed or disconfirmed. For example, items from the pronounceability task were rated as more pleasant when correctly endorsed following the "Pronounceability task?" query (a search confirmation) than when correctly rejected following the "Concrete/Abstract task?" query (a search disconfirmation). The final test of the COS model considered cued-recall using Lithuanian–English paired associates and is the primary motivation of the current report. Following elaborative study using an imagery-based mnemonic (Atkinson & Raugh, 1975), a cued-recall test was given in which subjects rated how pleasant they found each Lithuanian context word prior to attempting to recall its English target word. Critically, all pairs were presented once during study, and hence each Lithuanian cue was matched for exposure when later rated for pleasantness. Furthermore, because subjects had no familiarity with Lithuanian, the cues were effectively meaningless, minimizing any contribution of a priori semantics to the pleasantness ratings. Consistent with the COS model, recall of the paired English word (search confirmation) significantly increased rated pleasantness of the Lithuanian cue, whereas forgetting (search disconfirmation) decreased rated pleasantness. This study thus offered initial evidence that the outcomes of cued-recall searches moderate the affective appraisal of the retrieval cues that elicit the search and raise the possibility these effects may generalize to social contexts and stimuli. More specifically, they suggest that success or failure in searching one's memory for biographical information about faces might influence affective-linked ratings of them, and in

particular, appraisals of attractiveness. Aside from correct recall and forgetting, subjects can also provide incorrect answers during cued recall (viz., intrusions). As noted above, COS effects are assumed to also be moderated by effort and in the case of intrusions, there may be a trade-off between the fact they represent a confirmation yet may be more effortful. Thus, intrusions may be more modest in their contribution to positive affect. Consistent with this, the cued-recall experiment of Grybinas et al. (2019) found that intrusions yielded affective responses in the middle of successful retrievals and trials where subjects did not recall any associate (forgetting). Given this, intrusions are less diagnostic of the COS model than success or complete failure during cued recall.

Despite the robustness of COS effects using verbal memoranda, they may not generalize to face memoranda given the vast literature showing that neural substrates of face and word processing are largely distinct (for review, see Robotham & Starrfelt, 2017), and the fact that verbal and face memoranda differ considerably in the features upon which evaluative judgments can be made. In the case of faces, research suggests that attractiveness ratings depend upon perceptual cues such as symmetry, sexual dimorphism, and averageness, among others (e.g., Rhodes, Proffitt, Grady, & Sumich, 1998; Thornhill & Gangestad, 1999), perhaps because they signal qualities linked to evolutionary fitness. Alternatively, cultural or developmental accounts of facial attractiveness emphasize the role of learning in the acquisition of preferences for certain facial features (e.g., Cooper, Geldart, Mondloch, & Maurer, 2006). Critically however, both approaches assume that directly perceived facial features are the basis of rated attractiveness, and that these same features are used across exemplars for evaluation.

In contrast, pleasantness ratings of meaningful words are instead primarily based on long-term semantic information that can be rich, conflicting, and vary in relevance across tested exemplars. For example, when considering the pleasantness of "diamonds," some may consider their appearance or durability favorably, whereas others may consider their adverse sociopolitical consequences or wealth signaling negatively, and yet others may attempt to weigh these potentially conflicting aspects. Moreover, when evaluating "ice cream" for pleasantness, an entirely different domain of semantic knowledge is relevant. If this subjectivity and variability in sampling the semantic feature space considered relevant for the affective evaluation of words is critical for COS effects to emerge, then these effects would not occur for attractiveness ratings of faces, which are based on a directly perceived and common sets of features. However, if COS effects depend upon generalized reactions to memory search outcomes, then they should generalize to attractiveness rating of faces, even though words and faces differ in the relevant features used for evaluation and demonstrate many differences during encoding and memory retrieval (Bruce & Young, 1986; Coleshill et al.,

2004; Herlitz & Lovén, 2013; Woodhead & Baddeley, 1981; Yonelinas, Kroll, Dobbins, & Soltani, 1999). Finally, it is important to emphasize the only COS effect demonstrated for cued recall to date, used completely meaningless (from the perspective of the participants) Lithuanian word cues. Thus, the inherent semantic ambiguity of these materials may be requisite for the consequences of retrieval outcomes to be observed during cued recall. Since there is no comparable degree of ambiguity in the attractiveness rating of faces, which yield high interrater reliability (Foo, Simmons, & Rhodes, 2017), successful generalization is far from guaranteed.

To critically test the generalizability of COS effects to socially relevant stimuli, retrieval demands, and evaluations, we adapted the cued-recall COS verbal paradigm of Grybinas et al. (2019), focusing on the ability to recall a face's previously learned name (Experiment 1a) or profession (Experiment 1b) and its potential effect on rated attractiveness.

Experiment 1a

Experiment 1a used face–name cued-recall to investigate whether recall of names influences the affective appraisal of their paired faces. In particular, by asking subjects to rate attractiveness of faces prior to recalling their paired name associates we were able to examine whether memory retrieval attempts alter rated attractiveness.[1] Based on the COS model, we predicted accurate recall of associated names (confirmations) would produce higher attractiveness ratings than faces that failed to produce cued recall of names (disconfirmations); because inaccuracy has been shown to dampen COS effects, intrusions (incorrect confirmations) were anticipated to fall somewhere in between.

---

[1] In Experiments 1 and 2 of Grybinas et al. (2019), the order recognition and pleasantness judgments were manipulated such that subjects either rated items for pleasantness either before or after the recognition decision. Neither experiment showed a main effect of judgment order on rated pleasantness, nor did judgment order interact with the outcome of memory judgment on rated pleasantness. This insensitivity of judgment order appears to occur because subjects initiate retrieval attempts before pleasantness ratings even when those ratings are performed first. For example, in Experiment 2 of Grybinas et al. (2019), recognition decision times were almost 3 times faster when they followed pleasantness ratings (mean median 548 ms) than when they preceded the pleasantness ratings (mean median 1,470 ms), $t(1, 152) = 19.5$, $p < .001$, and the same pattern occurred with pleasantness judgments, $t(1, 152) = 6.74$, $p < .001$. This would suggest that the memory decisions and pleasantness ratings were being processed to some degree jointly. In the current study, we chose to have the pleasantness ratings conducted first because it highlights the importance of retrieval outcomes that may be concurrently undertaken during the pleasantness evaluation. Again, however, the findings of Grybinas et al. (2019) indicate similar effects of retrieval outcomes occur regardless of whether memory reporting precedes or follows the pleasantness rating.

Method

Participants

Thirty-five Washington University in St. Louis undergraduates participated in exchange for course credit. Informed consent was obtained in compliance with the Institutional Review Board of Washington University in St. Louis. Testing occurred in groups of one to five people. This sample size was based on Experiment 4 of Grybinas et al. (2019), which demonstrated a large differential effect between successful and failed cued recall on rated pleasantness ($d = 1.25$) using a sample size of 32. Assuming the same effect size would lead to 99% power in the present design.

Materials

Face stimuli were randomly selected from either 105 male or female faces in the Chicago Faces Database (chicagofaces.org; Ma, Correll, & Wittenbrink, 2015), all of which have neutral expressions. For each participant, a subset of 35 faces were randomly selected and held out for initial normative ratings. Following this, 70 different male faces or female faces were used in the two subsequent study–recall cycles. First-name associates were randomly selected from top 70 male or female baby names from 2000–2010, according to the United States Social Security Administration (https://www.ssa.gov/OACT/babynames/). Within each gender of face cues, names were randomly assigned for each subject. To ensure there was no effect of gender confounded with face–name recall, subjects were randomly assigned to receive either all female face–name pairs or all male face–name pairs. Stimuli were presented using PsychoPy software (Peirce, 2007).

Design and procedure

Prior to memory testing, a norming phase was administered in which participants rated 35 faces not used in his or her session. Attractiveness ratings were made on a 1–7 Likert scale (1 = not at all attractive to 7 = extremely attractive) and self-paced. These values were combined with the ratings in the norming phase of Experiment 1b to create a normative baseline for the materials. Following the norming phase, two study–test cycles of the memory experiment were administered each consisting of 35 to-be-encoded face–name associates.

The study phase used an elaborative imagery task (McCarty, 1980) to facilitate face–name encoding with each pair presented for 15 s. During the 15 s, subjects were asked to identify a salient feature (e.g., big eyes) and a word that was phonologically/orthographically similar to the paired name (e.g., *dust* for "Dustin"), and then asked to imagine these two features interacting (e.g., dust going into big eyes).

Following each 15-s trial, subjects were asked to indicate the imaginability of the mnemonic on a 1–4 scale.

Testing began immediately after reading instructions. During test, each of the 35 encoded faces was serially re-presented to subjects. Before attempted recall of name-associate, subjects rated the attractiveness of each face using the same method as in the initial norming phase (1–7 Likert scale of attractiveness, self-paced). Following each attractiveness rating, a space appeared allowing subjects to type in their recalled response (i.e., the paired name). After typing in their response, the "enter" key was pressed to move on to the following trial; if unable to recall the name-associate, subjects were simply asked to press "enter" and advance. Subjects entered recall responses at their own pace.

Data were analyzed using the R statistical language with lme4 (Bates, Mächler, Bolker, & Walker, 2014), lmerTest (Kuznetsova, Brockhoff, & Christensen, 2015), ez (Lawrence, 2016), broom (Robinson, 2017), tidyverse (Wickham, 2017), and sjPlot (Lüdecke, 2018) packages. The project was preregistered at the Open Science Foundation (osf.io/tgh63), where the data and analysis files are also available (osf.io/twsn2).

Results

Recall performance

Subjects correctly recalled names on 28% of the trials (recall). For incorrect trials, subjects failed to produce a response 50% of the time (forgotten) and falsely recalled a name 22% of the time (intrusion). Ratings of imaginability (1–4) during encoding modestly predicted subsequent recall levels. This was tested by correlating imageability ratings and subsequent name recall (1 = recall; 0 = intrusion or forgotten) within each participant. These correlations were then transformed via Fisher's z and submitted to a one-sample t test (mean r = .20), $t(34) = 8.09$, $d = 1.36$, $p < .001$. Thus, subjectively successful imagery anticipates subsequent recall, consistent with the elaborative imagery encoding task.

Recall and face attractiveness

Figure 1 shows the mean attractiveness rating at test (values centered within each subject) for each category of cued-recall outcome: recall, intrusion, and forgotten trials. To test whether recall outcomes influenced the overall pattern of rated attractiveness, a one-way repeated-measures analysis of variance (ANOVA) on mean attractiveness ratings was conducted across the three outcomes, which trended towards significance, $F(2, 68) = 3.07$, $\eta_p^2 = .083$, $p = .053$. Planned pairwise comparisons among the three outcomes were significant for recall versus forgotten trials, $t(34) = 2.64$, $d = .45$, $p = .012$, trended towards significance for intrusions versus forgotten
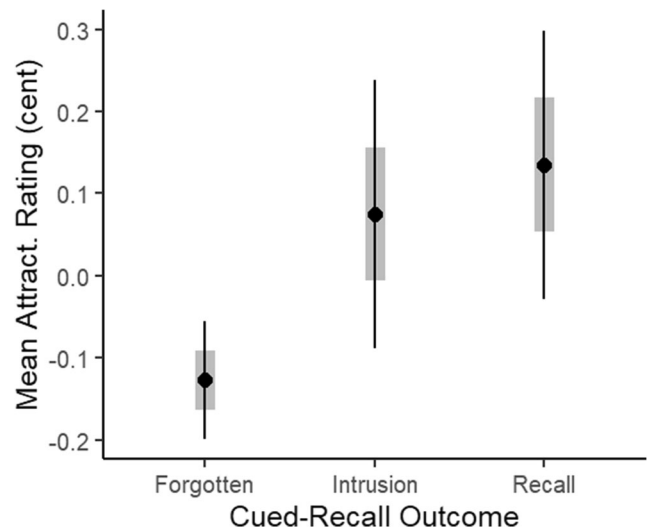

Fig. 1 Relationship between cued-recall outcomes and mean rated attractiveness ratings of faces at test. Points indicate the mean, gray boxes indicate ±1 *SEM*, and lines indicate ±2 *SEM*s. Attractiveness ratings were centered within each subject before plotting

trials, $t(1, 34) = 2.02$, $d = .34$, $p = .052$, and unreliable for recall versus intrusions, $t(1, 34) = .46$, $p = .648$. Consistent with the COS model, successfully recalling names increased rated attractiveness relative to completely forgetting that information, yielding a roughly medium effect size. Unlike prior findings, however, there were no statistical differences between accurate (recall) and inaccurate (intrusions) confirmations.

In contrast to a retrieval search-based phenomenon, one might argue that the Recall > Forgotten effect occurred because faces with greater normative attractiveness engendered better initial encoding of names. This would mean that the key Recall > Forgotten effect was not due to retrieval outcomes moderating affect at test, but instead due to more normatively attractive faces eliciting stronger encoding, perhaps because of increased interest at encoding. To rule out this interpretation, we contrasted the normative attractiveness ratings for the three cued-recall outcomes within participants (recall vs. intrusion vs. forgotten). That is, for each subject, we replaced their attractiveness ratings given at test with the normative attractiveness values calculated from their cohort during the norming phase. Critically, no subject provided norming data for the faces upon which they were tested, so the norms are statistically independent of subjects' attractiveness ratings at test. If normative attractiveness governed name encoding, then the recall, intrusion, and forgotten retrieval categories should yield decreasing normative scores. This did not occur (see Fig. 2), and a one-way ANOVA demonstrated no reliable difference for normative attractiveness scores across the retrieval outcomes, $F(2, 68) = 1.32$, $\eta_p^2 = .037$, $p = .275$. Furthermore, the direct comparison of the recall versus forgotten outcomes was also unreliable, $t(34) = .87$, $d = .15$, $p = .393$, and in fact trended in the *opposite* numerical direction
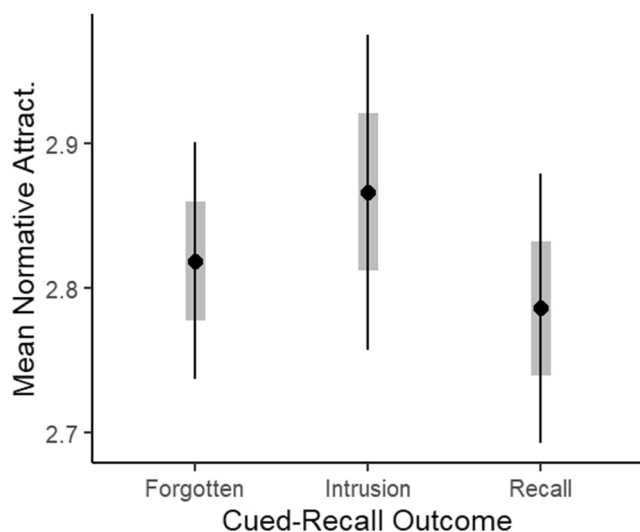
Fig. 2 Relationship between cued-recall outcomes and normative attractiveness. Points indicate the mean, gray boxes indicate ±1 SEM, and lines indicate ±2 SEMs

compared with Fig. 2. A final potential concern with this null finding might be that the norms were simply an unreliable measure of the attractiveness perceived by observers at study and test.

To further explore this, we used linear mixed modeling (LMM) to test whether both the normative attractiveness scores and the retrieval outcomes each contributed to the attractiveness ratings given at test. Each attractiveness rating at test was modeled as a function of the contrast coded retrieval outcome (1 = recalled, 0 = intrusion, −1 = forgotten), and the normative pleasantness rating of the face. As Table 1 shows, the retrieval outcome and normative attractiveness each made

Table 1 Linear mixed model set and results regressing rated test attractiveness on predictors - retrieval outcome and normative attractiveness. Subject intercepts and slopes were modeled as random, uncorrelated effects

Predictors	Rated test attractiveness		
	Estimates	CI	p
Intercept	2.84	[2.39, 3.30]	**<.001**
Retrieval outcome	0.14	[0.07, 0.21]	**<.001**
Norm. attractiveness (cent.)	0.78	[0.65, 0.90]	**<.001**
Random effects			
σ^2	1.03		
$\tau_{00\ \text{Intercept}}$	0.84		
$\tau_{00\ \text{Retrieval Outcome}}$	0.02		
$\tau_{00\ \text{Norm. Attract.}}$	0.13		
ICC	0.45		
N_{subject}	35		
Observations	2,450		
Marginal R^2 / Conditional R^2	0.196 / 0.558		

reliable contributions to predicting the attractiveness judgments at test. Moreover, the contribution of normative attractiveness was quite large, demonstrating that the normative scores are highly predictive of responding at the trial level and therefore must be psychometrically reliable. Thus, overall, it is not the case that the normative attractiveness scores are unreliable indicators of each subjects' perceived attractiveness, but instead that the contribution of normative attractiveness to rated attractiveness at test is separate from the further reliable influence of the retrieval outcomes.

Experiment 1b

Experiment 1b sought to extend the findings of Experiment 1a to a different type of biographical information—namely, professions. Also, because recall performance was somewhat low in Experiment 1a, three study–test phases instead of two were used in Experiment 1b to shorten the length of each study list. This was a minor departure from the preregistered design.

Method

Participants

Thirty-five Washington University in St. Louis undergraduates participated in exchange for course credit. Informed consent was obtained in compliance with the Institutional Review Board of Washington University in St. Louis. Testing occurred in groups of one to five people. One subject was omitted from analysis due to invariant ratings of attractiveness at test, leaving 34 subjects for analysis.

Materials

All face stimuli were drawn from the same 105 male or 105 female faces as in Experiment 1a. Subjects again either received all male or all female faces. The paired professions (e.g., "doctor") were randomly selected from 75 common professions listed in an Online Dictionary (http://learnersdictionary.com/3000-words/topic/jobs-professions).

Design and procedure

The overall design and procedure closely followed that of Experiment 1a. The experiment started with a norming phase in which 30 random hold-out faces were rated for attractiveness (rather than 35). These ratings were combined with the norming ratings of Experiment 1a to create the master list of normative scores. To increase cued-recall performance, we constructed three smaller study–test cycles of 25 face–profession pairs (rather than two cycles of 35) for a total of

75 A–B pairs at test. During encoding, subjects were serially presented with each face–profession pair for 15-s each and asked to identify a salient feature of the face (e.g., large forehead) with a word that was phonologically/orthographically similar to the paired profession (e.g., *bus* for "businesswoman"), and then asked to imagine these two features interacting (e.g., a bus driving across a prominent forehead). After each trial, subjects rated the mnemonic for imaginability (1–4). During test, each of the 25 encoded face cues was rated for attractiveness (1–7 Likert scale) and then queried for recall of the profession-associate.

Results

Cued-recall performance

Subjects correctly recalled 39% of professions, forgot 45%, and committed intrusions for 16% of the trials. In line with the reduced list lengths, the proportion of recall in Experiment 1b was higher than that of 1a, $t(66) = 2.84$, $p = .008$. As in Experiment 1a, ratings of imaginability (1–4) during encoding predicted subsequent recall levels (mean $r = .14$), $t(33) = 5.64$, $d = .97$, $p < .001$, consistent with the elaborative encoding task.

Cued-recall and face attractiveness

Figure 3 shows the mean attractiveness rating at test (centered within each subject) as a function of each cued-recall outcome: recall, intrusion, and forgotten. A one-way repeated-measures ANOVA on mean attractiveness ratings across the three outcomes was significant,
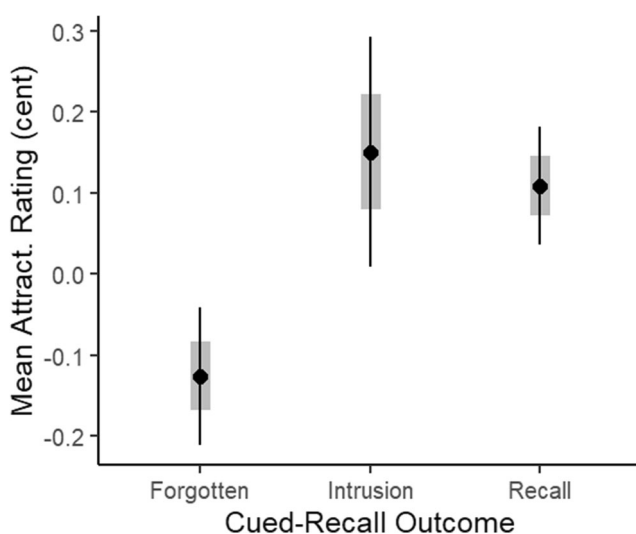

Fig. 3 Relationship between cued-recall outcomes and mean attractiveness ratings of the materials. Points indicate the mean, gray boxes indicate ±1 SEM, and lines indicate ±2 SEMs

$F(2, 64) = 6.88$, $\eta_p^2 = 0.177$, $p = .002$.[2] Planned pairwise comparisons among the three retrieval outcomes were significant for recall versus forgotten trials, $t(33) = 3.54$, $d = .61$, $p = .001$, and intrusion versus forgotten trials, $t(32) = 2.76$, $d = .48$, $p = .009$, but unreliable for recall versus intrusions, $t(32) = .39$, $p = .701$.[1] Thus, consistent with the COS model and Experiment 1a, recalling the associative information increased rated attractiveness relative to completely forgetting that information, yielding a medium to large effect size; however, there was no statistical difference between erroneous confirmations (intrusions) and successful confirmations (recall).

To again test whether the Recall > Forgotten effect might be an artifact of superior encoding for normatively more attractive faces, we replaced each subject's attractiveness test ratings with the normative attractiveness of the cue face, looking for differences across the recall outcomes (see Fig. 4). There were none via a one-way ANOVA, $F(1, 33) = 0.36$, $p = .555$, nor a direct comparison of recall versus forgotten trials, $t(33) = .60$, $p = .555$. Again, the numerical differences across these two retrieval outcomes were numerically opposite of that which would have occurred had normatively attractive faces garnered superior encoding. Thus, higher normative attractiveness does not facilitate the encoding of either names or professions.

Finally, as in Experiment 1a, we modeled test attractiveness ratings at the trial level using predictors of contrast-coded retrieval outcome and normative attractiveness via LMM (see Table 2). Again, both made reliable contributions to attractiveness ratings at test.

Discussion

The mere exposure effect provides evidence that implicit memory retrieval phenomena may influence affective reactions. In contrast, the current study provides further evidence that the outcomes of explicit memory attempts (in this case, cued recall attempts) influence affective reactions towards memoranda. We previously observed a robust analogous phenomenon during verbal cued recall using meaningless Lithuanian cues and unrelated English associates (Grybinas et al., 2019), and the COS hypothesis put forward to explain this and related verbal memory phenomena suggested it may also occur for socially relevant stimuli (faces) and the types of biographical retrieval often attempted for new acquaintances (names and professions), even though these stimuli and this

---

[2] A subject in Experiment 1b did not commit any intrusions and was consequently left out of the one-way ANOVA. This subject was likewise omitted from any pairwise comparison involving intrusion trials, explaining the differences in reported degrees of freedom. There was thus a total of 33 subjects for these analyses.
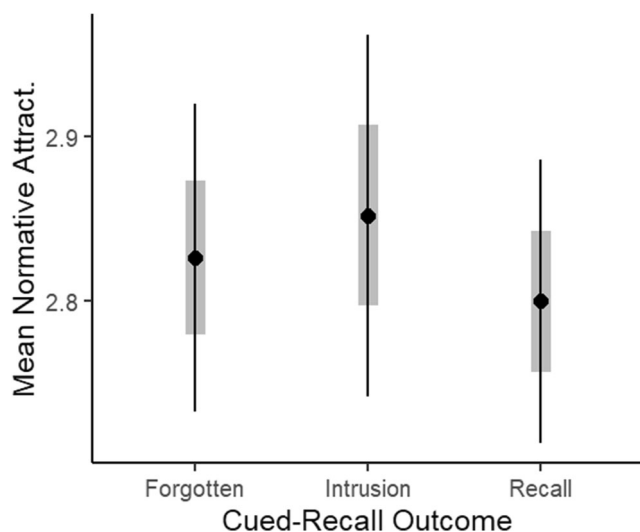
Fig. 4 Relationship between cued-recall outcomes and normative attractiveness. Points indicate the mean, gray boxes indicate ±1 SEM, and lines indicate ±2 SEMs

evaluation are markedly different than the pleasantness ratings of meaningless Lithuanian cues. The current data demonstrate that this form of retrieval-moderated affective misattribution does translate to more socially relevant scenarios, yielding increases or decreases in positive affect towards faces depending upon whether retrieval searches are confirmed or disconfirmed, respectively.

Under the confirmation of search (COS) model the confirmatory status of memory searches produces these affective responses, which are assumed to occur because memory searches are model-based and hence confirmatory in nature. This assumption is present in numerous memory frameworks

Table 2 Linear mixed model set and results regressing rated test attractiveness on predictors - retrieval outcome and normative attractiveness. Subject intercepts and slopes were modeled as random, uncorrelated effects

Predictors	Rated test attractiveness		
	Estimates	CI	p
Intercept	2.85	[2.55, 3.15]	**<.001**
Retrieval outcome	0.14	[0.08, 0.19]	**<.001**
Norm. attractiveness	0.70	[0.63, 0.77]	**<.001**
Random effects			
σ^2	0.89		
$\tau_{00\ subject}$	0.44		
$\tau_{00\ subject.1}$	0.01		
$\tau_{00\ subject.2}$	0.04		
ICC	0.33		
$N_{subject}$	34		
Observations	2,550		
Marginal R^2 / Conditional R^2	0.272 / 0.513		

that assume that mental models guide deliberate retrieval attempts. Although these frameworks use various characterizations for such models, such as retrieval descriptions (Norman & Bobrow, 1979), systematic or heuristic monitoring processes (Johnson, Hashtroudi, & Lindsay, 1993), encoding specificity (Tulving & Thomson, 1973), and transfer appropriate processing (Rajaram, Srinivas, & Roediger, 1998), all assume these models or descriptions are crucial for efficient and selective memory retrieval. The COS model takes this one step further in assuming that subjects have a preference for confirming these models via retrieval in accordance with the well-documented preference of observers to prefer confirmation over disconfirmation in a host of judgment and decision-making situations (Nickerson, 1998). The final element of the COS model is the misattribution of the affective consequences of retrieval outcomes to the evaluation of the memoranda that are the focus of retrieval search—in this case, faces. Thus, the minor elation or minor frustration that results from confirmed or disconfirmed memory search is projected onto the memoranda eliciting the search. Thus, even though rated face attractiveness has been tied to a range of overt perceptual characteristics, such as symmetry, averageness, and sexual dimorphism (e.g., Rhodes et al., 1998), the confirmation or disconfirmation of episodic retrieval searches cued by faces also alters their rated attractiveness.

In the case of cued recall, the COS model predicts the greatest differences between accurate recall and attempts that completely fail to produce associative information, a result that was demonstrated twice in the present study. In the case of intrusions, the model suggests the effects may be intermediate because, although intrusions constitute a confirmation, they may require greater expenditure of effort and uncertainty, thereby undercutting the confirmatory effect. In other words, while intrusions are ostensibly confirmatory, they may yield more muted responses than successful recall because they are more uncertain and effortful. However, despite the reliable influence of accuracy documented in previous studies (e.g., Chetverikov, 2014; Grybinas et al., 2019; Lee, 2001), we observed no statistical differences across successful and unsuccessful confirmations (i.e., recall vs. intrusion). This may in part reflect the small number of trials in which subjects committed intrusions (Experiment 1a = 22%; Experiment 1b = 16%), resulting in a noisy estimate in this category. Furthermore, performance on the face cued-recall task was notably lower (e.g., proportion correct = 28% in Experiment 1a) than the overall performance for verbal memoranda used in Grybinas et al. (2019; proportion correct = 49%), and this may have resulted in similar levels of retrieval effort for memory searches that result in successful recall versus intrusions in the current report. Nonetheless, the key prediction of the COS model—namely, that rated attractiveness would be higher for confirmed than for disconfirmed memory searches was supported. However,

future research of the COS model may benefit from the collection of retrieval confidence ratings, which may help gauge the relative effort and/or uncertainty of the different retrieval outcomes.

Although we focused on attractiveness ratings, the COS model assumes that many evaluative ratings should likewise be influenced by the outcomes of memory search. For example, analogous effects would be predicted for judgments of trustworthiness or conscientiousness. In contrast, evaluations that are more negative in nature (e.g., "threateningness") might show an inverse effect in which confirmation decreases ratings and disconfirmation increases them. However, predictions are less clear for attributes that are not as universally positive or negative among raters. For example, questions of whether the face cue reflected a wealthy person might yield a mix of effects depending upon political or social attitudes. Nonetheless, since observers are likely unaware of the relation between their retrieval outcomes and these subtle increases and decreases in affective responding, such ambiguous (from a normative perspective) judgments may effectively capture individual differences in attitudes implicitly. For instance, negative attitudes towards wealth might be implicitly expressed through decreases in ratings following successful retrieval, whereas positive wealth attitudes might be implicitly expressed via increased judgments of wealth following retrieval. Thus, the COS model may lead to interesting insights in implicit attitudes and biases, and may be less open to demand characteristics than other implicit attitude techniques which have recently been critiqued in this regard (e.g., Bar-Anan & Nosek, 2012).

A second interesting avenue for the COS model and paradigm is research on special populations. For example, those suffering the early stages of impaired episodic memory, such as mild cognitive impairment patients, may be prone to misattributing the frustration of retrieval failure onto the focus of those retrieval attempts, such as close others, adding an unfortunate secondary consequence of their declining memory abilities.

Finally, from a basic research perspective the operating characteristics of the COS framework require more study. For instance, it is unknown whether COS affective responses outlast the immediate test period, or whether they in turn promote or impair further encoding of the cues. The latter is an interesting possibility, given evidence that value-based responses to the provision of choice during encoding promote retention, and that this affect is even more pronounced at delayed (~24 hours) versus immediate testing (DuBrow, Eberts, & Murty, 2019).

In conclusion, the current study demonstrates that affective responses to face memoranda are moderated by the retrieval status of associated biographical information. Thus, COS effects established during verbal explicit memory generalize to social stimuli and the kinds of biographical retrieval attempts

that often accompany social interaction. Additional potential consequences of these judgment-linked affective misattributions remain to be studied, but the fact that evaluative judgments of memoranda are not judged independently of episodic retrieval outcomes demonstrates an important novel interaction between cognitive and emotional processing.

References

Atkinson, R. C., & Raugh, M. R. (1975). An application of the mnemonic keyword method to the acquisition of a Russian vocabulary. *Journal of Experimental Psychology: Human Learning and Memory*, *1*(2), 126. https://doi.org/10.1037/0278-7393.1.2.126

Bar-Anan, Y., & Nosek, B. A. (2012). Reporting intentional rating of the primes predicts priming effects in the affective misattribution procedure. *Personality and Social Psychology Bulletin*, *38*(9), 1194–1208. https://doi.org/10.1177/0146167212446835

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). *Fitting linear mixed-effects models using lme4*. ArXiv Preprints. Article 1406.5823.

Bornstein, R. F., & D'Agostino, P. R. (1992). Stimulus recognition and the mere exposure effect. *Journal of Personality and Social Psychology*, *63*(4), 545–552. https://doi.org/10.1037/0022-3514.63.4.545

Bornstein, R. F., & D'Agostino, P. R. (1994). The attribution and discounting of perceptual fluency: Preliminary tests of a perceptual fluency/attributional model of the mere exposure effect. *Social Cognition*, *12*(2), 103–128. https://doi.org/10.1521/soco.1994.12.2.103

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*(3), 305–327. https://doi.org/10.1111/j.2044-8295.1986.tb02199.x

Chetverikov, A. (2014). Warmth of familiarity and chill of error: Affective consequences of recognition decisions. *Cognition and Emotion*, *28*(3), 385–415. https://doi.org/10.1080/02699931.2013.833085

Coleshill, S. G., Binnie, C. D., Morris, R. G., Alarcón, G., van Emde Boas, W., Velis, D. N., … van Rijen, P. C. (2004). Material-specific recognition memory deficits elicited by unilateral hippocampal electrical stimulation. *Journal of Neuroscience*, *24*(7), 1612–1616. https://doi.org/10.1523/JNEUROSCI.4352-03.2004

Cooper, P. A., Geldart, S. S., Mondloch, C. J., & Maurer, D. (2006). Developmental changes in perceptions of attractiveness: A role of experience? *Developmental Science*, *9*(5), 530–543. https://doi.org/10.1111/j.1467-7687.2006.00520.x

DuBrow, S., Eberts, E. A., & Murty, V. P. (2019). A common mechanism underlying choice's influence on preference and memory. *Psychonomic Bulletin & Review*, *26*(6), 1958–1966. https://doi.org/10.3758/s13423-019-01650-5

Foo, Y. Z., Simmons, L. W., & Rhodes, G. (2017). Predictors of facial attractiveness and health in humans. *Scientific Reports*, *7*(1), 39731. https://doi.org/10.1038/srep39731

Grybinas, D., Kantner, J., & Dobbins, I. G. (2019). Mere exposure effect(s) in the context of explicit memory search. *Memory & Cognition*, *47*(7), 1314–1327. https://doi.org/10.3758/s13421-019-00935-3

Herlitz, A., & Lovén, J. (2013). Sex differences and the own-gender bias in face recognition: A meta-analytic review. *Visual Cognition*,

21(9/10), 1306–1336. https://doi.org/10.1080/13506285.2013.823140

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*(1), 3–28. https://doi.org/10.1037/0033-2909.114.1.3

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package 'lmertest' (R Package Version 2.0) [Computer software]. https://cran.r-project.org/. Accessed March-May 2020.

Lawrence, M. A. (2016). ez: Easy analysis and visualization of factorial experiments (R Package Version 4.4-0) [Computer software]. https://CRAN.R-project.org/package=ez. Accessed March-May 2020.

Lee, A. Y. (2001). The mere exposure effect: An uncertainty reduction explanation revisited. *Personality and Social Psychology Bulletin*, *27*(10), 1255–1266.

Lüdecke, D. (2018). sjPlot: Data visualization for statistics in social science (R Package Version, 2.1) [Computer software]. https://rdrr.io/cran/sjPlot/. Accessed March-May 2020.

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods, 47*(4),1122–1135. https://doi.org/10.3758/s13428-014-0532-5

McCarty, D. L. (1980). Investigation of a visual imagery mnemonic device for acquiring face–name associations. *Journal of Experimental Psychology: Human Learning and Memory, 6*(2), 145.

Newell, B. R., & Shanks, D. R. (2007). Recognising what you like: Examining the relation between the mere-exposure effect and recognition. *European Journal of Cognitive Psychology*, *19*(1), 103–118. https://doi.org/10.1080/09541440500487454

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175–220. https://doi.org/10.1037/1089-2680.2.2.175

Norman, D. A., & Bobrow, D. G. (1979). Descriptions: An intermediate stage in memory retrieval. *Cognitive Psychology*, *11*(1), 107–123. https://doi.org/10.1016/0010-0285(79)90006-9

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. Journal of Neuroscience Methods, 162(1/2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Rajaram, S., Srinivas, K., & Roediger III, H. L. (1998). A transfer-appropriate processing account of context effects in word-fragment completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(4), 993.

Rhodes, G., Proffitt, F., Grady, J. M., & Sumich, A. (1998). Facial symmetry and the perception of beauty. *Psychonomic Bulletin & Review*, *5*(4), 659–669. https://doi.org/10.3758/BF03208842

Robinson, D. (2017). broom: Convert statistical analysis objects into tidy data frames (R Package Version 0.4, 2) [Computer software]. https://cran.r-project.org/web/packages/broom/index.html. Accessed March-May 2020.

Robotham, R. J., & Starrfelt, R. (2017). Face and word recognition can be selectively affected by brain injury or developmental disorders. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01547

Thornhill, R., & Gangestad, S. W. (1999). Facial attractiveness. *Trends in Cognitive Sciences*, *3*(12), 452–460. https://doi.org/10.1016/S1364-6613(99)01403-5

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*(5), 352–373. https://doi.org/10.1037/h0020071

Wickham, H. (2017). tidyverse: Easily install and load the 'tidyverse' (R Package Version 1.2.1, 1(1)) [Computer software]. https://CRAN.R-project.org/package=tidyverse. Accessed March-May 2020.

Woodhead, M. M., & Baddeley, A. D. (1981). Individual differences and memory for faces, pictures, and words. *Memory & Cognition*, *9*(4), 368–370. https://doi.org/10.3758/BF03197561

Yonelinas, A. P., Kroll, N. E. A., Dobbins, I. G., & Soltani, M. (1999). Recognition memory for faces: When familiarity supports associative recognition judgments. *Psychonomic Bulletin & Review*, *6*(4), 654–661. https://doi.org/10.3758/BF03212975

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*(2, Pt. 2), 1–27. https://doi.org/10.1037/h0025848