Psychology and Aging
2009, Vol. 24, No. 4, 981–988

0882-7974/09/$12.00 DOI: 10.1037/a0017545

BRIEF REPORTS

Elevated False Recollection of Emotional Pictures in Young and Older Adults

David A. Gallo, Katherine T. Foster, and Elizabeth L. Johnson
University of Chicago

Current theories predict opposing effects of emotionally arousing information on false memory. If emotion enhances true recollection, then false recollection might be lower for emotional than for neutral pictures. However, if emotion enhances conceptual relatedness, then false recollection might increase for nonstudied but emotionally related pictures. We contrasted these 2 factors in young and older adults, using the International Affective Pictures Systems set (Lang, Bradley, & Cuthbert, 2005). Although both age groups used recollection in our task, false recollection was greatest for emotional pictures, supporting a conceptual relatedness account. Finally, even after accuracy differences were controlled, age was related to high-confidence false recollection of emotional pictures.

Keywords: aging, emotion, illusory recollection, retrieval monitoring

Emotionally arousing content can enhance memory for studied events (Cahill & McGaugh, 1998; Hamann, 2001; Kensinger, 2007; LaBar, 2007; Mather, 2007). In both young and older adults, specialized brain structures such as the amygdala respond to emotionally arousing stimuli (e.g., Mather et al., 2004). These neural responses may enhance attention to emotional details and facilitate encoding (Kensinger & Schacter, 2005a; 2005b). Thus, although normal aging can impair episodic memory, emotion can increase memory for studied events in both young and older adults (e.g., Kensinger, Brierly, Medford, Growdon, & Corkin, 2002; see also Comblain, D'Argembeau, Van der Linden, & Aldenhoff, 2004; Denburg, Buchanan, Tranel, & Adolphs, 2003; Emery & Hess, 2008; Kensinger, Garoff-Eaton, & Schacter, 2007).

In contrast to studied events, there are competing theories about the influence of emotionally arousing content on false memory of nonstudied events.[1] One theoretical position is that emotional content can suppress false memory (e.g., Kensinger & Schacter, 2005a, 2005b; Schacter, Gallo, & Kensinger, 2007). This position is based on the idea that emotionally arousing content can increase recollective distinctiveness (e.g., Ochsner, 2000; Schmidt, 2007), a factor that generally enhances retrieval monitoring accuracy (e.g., Johnson, Hashtroudi, & Lindsay, 1993). For example, research with emotionally neutral stimuli indicates that basing one's memory decision on perceptually distinctive recollections can reduce false recognition of nonstudied events (e.g., Gallo, Cotel, Moore, & Schacter, 2007; Schacter, Israel, & Racine, 1999). In these studies, both young and older adults were less susceptible to memory distortion when tested for pictures rather than words, because they expected more perceptually distinctive recollections for pictures (i.e., a *distinctiveness heuristic*). By extension, Schacter et al. (2007) argued that emotionally arousing content also might enhance a distinctiveness heuristic. To the extent that emotionally arousing content enhances recollective distinctiveness in both young and older adults, both age groups should be able to suppress false memory using a distinctiveness heuristic.

Evidence for an emotional distinctiveness heuristic on false memory is mixed. Kensinger and Corkin (2004) found that both young and older adults were less likely to falsely recognize emotionally negative words compared with neutral words (see also Pesta, Murphy, & Sanders, 2001; Piguet, Connally, Krendl, Huot, & Corkin, 2008) but argued that these effects may have been due either to emotional or conceptual distinctiveness. Using a different task, Fernandes, Ross, and Wiegand (2008) found fewer recall intrusions for emotional than for neutral pictures, but these effects may have reflected overly general recall (as opposed to false recall). Finally, a few studies have found greater source memory accuracy for emotional material in young and older adults (e.g., Kensinger, O'Brian, Swanberg, Garoff-Eaton, & Schacter, 2007; May, Rahhal, Berry, & Leighton, 2005; Rahhal, May, & Hasher, 2002), but such effects on typical source memory tests can be driven entirely by processes affecting studied items. Considering

David A. Gallo, Department of Psychology and Center for Cognitive and Social Neuroscience, University of Chicago; Katherine T. Foster and Elizabeth L. Johnson, Department of Psychology, University of Chicago.

This work was supported by National Institute of Aging Grant AG30345. We thank Mara Mather for comments on an earlier version of this article.

Correspondence concerning this article should be addressed to David A. Gallo, Department of Psychology, University of Chicago, 5848 S. University Ave., Chicago, IL 60637. E-mail: dgallo@uchicago.edu

[1] Our focus here is on false memories for nonstudied events (e.g., claiming a picture was studied when it was not). This phenomenon is different from the idea that the emotional aspects of studied events can impair memory for peripheral or nonemotional details of the same events (i.e., the Easterbrook or weapons-focus hypothesis; see Cook, Hicks, & Marsh, 2007; Kensinger, 2007; Mather, 2007).

only nonstudied items, we surveyed nine relevant studies and found little evidence that source misattributions were reduced under emotional conditions (Cook, Hicks, & Marsh, 2007; Doerksen & Shimamura, 2001; Kensinger & Corkin, 2004; Kensinger, O'Brian, et al., 2007; Kensinger & Schacter, 2005a, 2005b, 2006; May et al., 2005; Rahhal et al., 2002).

An alternative theoretical position is that emotional content can increase false memory. Emotion can serve as an organizational factor in memory, helping individuals to categorize or associate items, especially if the items have more preexisting conceptual overlap relative to neutral items (cf. Bower, 1981; Sison & Mather, 2007; Talmi, Luk, McGarry, & Moscovitch, 2007). This conceptual relatedness might increase confusability in memory, thereby increasing the risk of false recognition of nonstudied but conceptually related emotional items. Along these lines, Brainerd, Stein, Silveira, Rohenkohl, and Reyna (2008) found that there was more false recognition of emotional words than of neutral words in young adults. They argued that emotions serve as an organizational factor, enhancing false recognition of nonstudied items that are consistent with the emotional gist of studied items (also see Howe, 2007). This process might explain why emotional words can lead to more liberal recognition memory bias in both young and older adults (Budson et al., 2006; Kapucu, Rotello, Ready, & Seidl,2008), making subjects more prone to endorse studied and nonstudied items (also see D'Argembeau & Van der Linden, 2004; Dougal & Rotello, 2007; and Windmann & Kutas, 2001). In fact, the potential for emotion to elevate false recognition might be exacerbated by age, given that older adults can be more susceptible to false recognition of other types of conceptually related material such as semantically associated words (e.g., Tun, Wingfield, Rosen, & Blanchard, 1998; for a review, see Gallo, 2006).

It is important to note that in all of the aforementioned studies showing enhanced false recognition for emotional items, the materials studied were words, and not all of these studies used emotionally arousing materials (e.g., Brainerd et al., 2008). As discussed by Kapucu et al. (2008), it remains to be seen whether such effects could be found with more distinctive and emotionally arousing materials such as the International Affective Pictures System (IAPS; Lang, Bradley, & Cuthbert, 2005). The IAPS set contains detailed images of people and objects in various settings and is commonly used in emotional memory research. Use of these sorts of materials is especially important in studies of the influence of emotionally arousing content on memory distortion, because research on the distinctiveness heuristic indicates that pictures are likely to elicit recollection-based responding, whereas words are more susceptible to familiarity effects (e.g., Schacter et al., 1999). If emotionally arousing content can suppress false memory, then these effects should be most pronounced with distinctive pictures that elicit strong emotions. Similarly, if emotionally arousing content is more likely to contain conceptual overlap than neutral content, then this effect might be exaggerated with strongly emotional materials.

The current study provided a direct test of these contrasting theoretical positions with IAPS images and a memory task specifically designed to assess the influence of recollection on false recognition. Participants studied neutral nonarousing, positive arousing, and negative arousing pictures. Memory was then tested with a modified version of the criterial recollection task (Gallo, Weiss & Schacter, 2004). In this task, we selectively cued the recollection of different types of stimuli (e.g., emotional or neutral pictures) across different test blocks, holding the format of the retrieval cues constant (verbal labels corresponding to the studied pictures). Of critical interest was false recognition of test labels that were not associated with a picture during the study phase. According to the distinctiveness heuristic hypothesis, both young and older adults should expect more distinctive recollections for emotional pictures, and so false recognition should be reduced when participants are tested for these types of pictures as opposed to neutral pictures. In contrast, the conceptual relatedness hypothesis predicts that people should be more likely to confuse emotional items in memory. By this account, false recognition should be greater when participants are tested for emotional information than when they are tested for neutral information, and these effects may be exacerbated by age.

We also had two manipulation checks. First, we changed the font color of some of the emotionally neutral labels at study from black to red. Prior work with emotionally neutral materials has indicated that recollections of pictures are more distinctive than those of font color, leading to reduced false recognition in young and older adults (Gallo et al., 2007). By replicating this perceptual distinctiveness effect in the current study, we confirmed that our task was sensitive to recollection-based processes. Second, in a separate procedure, we measured the degree that the content portrayed in the various IAPS items was perceived as conceptually overlapping (see Materials section). This procedure confirmed that the emotionally arousing content in the IAPS was perceived as more conceptually related than the neutral content.

Method

Participants

Participants were 24 University of Chicago students (mean age = 21.3 years, $SD = 2.6$; mean education = 14.3 years, $SD = 1.9$) and 24 older adults from the surrounding community (mean age = 77.7 years, $SD = 6.9$; mean education = 16.7 years, $SD = 2.8$). Older adults were high functioning as measured on the Mini-Mental State Exam (MMSE; Folstein, Folstein, & McHugh, 1975; $M = 28$ of 30; $SD = 1.6$), reported no neurological diseases or medications for mental disorders, and scored low on the Geriatric Depression Scale (GDS; Brink et al., 1982; $M = 3.5$, $SD = 3.2$).[2] Before testing, older adults rated their mood as more positive than did young adults on a scale ranging from 1 to 7 (Ms = 5.54 and 4.62, respectively), $t(46) = 2.89$, $SEM = 0.317$, $p < .05$, and after testing (Ms = 5.65 and 4.75, respectively), $t(45) = 2.73$, $SEM = 0.330$, $p < .05$.

Materials

We drew 192 color pictures from the IAPS, which includes pictures of objects and people in naturalistic contexts. According

[2] MMSE and GDS data were missing for one older adult. Additionally, although they did not report mental illness or depression, one older adult reported taking drugs that are commonly prescribed for depression and/or anxiety, and one reported having experienced a transient ischemic attack several years earlier. Both participants performed well on the memory task, and excluding their data did not change the overall pattern of results or any of the conclusions based on statistical analysis.

to valence norms (1–9 scale), 96 pictures were considered neutral (M = 5.19, range = 4.03–5.99), 48 negative (M = 2.95, range = 1.91–3.91), and 48 positive (M = 7.17, range = 6.18–8.34). Negative and positive pictures were equated for arousal (Ms = 5.66 and 5.58, respectively; range = 5.00–6.99) and were more arousing than neutral pictures (M = 3.37; range = 1.72–4.96). Extremely taboo pictures were excluded (i.e., nudity and mutilations). Each picture was given a unique 2- to 3-word label (e.g., *tourist with book* and *clothes iron* for neutral images; *toxic waste cleanup* and *electric chair* for negative images; *astronaut in space* and *view of waterfalls* for positive images). Post hoc coding indicated that the pictures primarily differed in depicting people (73% negative, 56% positive, 29% neutral), movement/activities (50%, 50%, 18%, respectively), outdoors (54%, 67%, 34%, respectively), and stationary objects (8%, 23%, 58%, respectively).

We assessed the arousal and conceptual relatedness of our IAPS labels in a separate rating task given to new participants who had not participated in the memory experiment (n = 18 per age group). Pictures were not presented in this rating task, because our primary focus in the memory experiment was on false recognition of test labels that did not correspond to studied pictures. Participants received separate lists of positive, negative, and neutral labels and rated each list for (a) emotional arousal (ranging from −2 to +2) and (b) content overlap (ranging from 1–5; see Appendix). Participants assigned the emotional arousal ratings before the content overlap ratings to clarify the distinction. A 2 (group) × 3 (valence) analysis of variance (ANOVA) on ratings of emotional arousal confirmed an effect of valence, $F(2, 68) = 125.03$, $MSE = 0.531$, $p < .05$, $\eta_p^2 = .79$, but no effect of age or interaction, both Fs < 1. Positive items were more arousing than neutral items (1.21 vs. 0.05), $t(35) = 14.68$, $SEM = 0.079$, $p < .05$, $d = 2.46$, and negative items were more arousing than neutral items (−1.49 vs. 0.05), $t(35) = 7.26$, $SEM = 0.213$, $p < .05$, $d = 2.20$. Analysis of ratings of content overlap revealed an effect of valence, $F(2, 68) = 43.99$, $MSE = 0.263$, $p < .05$, $\eta_p^2 = .56$, but no effect of age or interaction, both Fs < 1. Positive and negative items did not differ in terms of content overlap (3.05 vs. 2.94), $t < 1$, but each had more content overlap than neutral items (2.02), $t(35) = 11.45$, $SEM = 0.09$, $p < .05$, $d = 1.71$, for positive; $t(35) = 6.25$, $SEM = 0.148$, $p < .05$, $d = 1.13$ for negative.[3] Overall, these ratings confirmed that the content of the emotional items was perceived as more arousing and more conceptually related than that of the neutral items in both age groups.

Procedure

Figure 1 outlines the design. For the study phase, all items were randomly intermixed in a single study list. Each study trial began with a label in black font on the computer screen (neutral, positive, or negative). Depending on the item type, the black label was followed either by its corresponding picture (neutral, positive, or negative) or by the same label in red font (neutral items only). Each label and picture were rated for arousal (1 = *low*, 2 = *medium*, 3 = *high*). Participants studied each red label three times (nonconsecutively) in order to minimize subsequent familiarity differences at test between items associated with red labels at study and those associated with more memorable pictures (cf. Gallo et al., 2004). Items were counterbalanced across the label with a picture, the label without a picture, and nonstudied conditions. In addition, neutral items were counterbalanced across the red font and the picture conditions. Study presentation was self-paced, and all participants were aware of the memory test. Immediately following the study phase was an unrelated filler task (25 min). Older adults then took the memory test, while young adults were tested 24 hr later. We used these retention intervals to avoid floor effects in false recognition. They also minimized accuracy differences across age groups, allowing us to more directly assess age effects on emotional memory (cf. Kensinger, Garoff-Eaton, et al., 2007; Kensinger, O'Brian, et al., 2007) and high-confidence false recollection (cf. Dodson, Bawa, & Krueger, 2007).

The self-paced test phase was divided into four separate test blocks (red word test, neutral picture test, negative picture test, positive picture test), with block order counterbalanced. On each test block, participants were presented with 48 verbal labels in black font. Thus, on the three picture test blocks, 16 labels had been studied with their picture, 16 had been studied without a picture, and 16 had not been studied. On the red font test block, 16 labels had been studied in black followed by red font (three times), 16 had been studied in black font only (one time), and 16 had not been studied. Participants were instructed to respond "yes" to test labels only if they recollected having studied a corresponding picture (picture tests) or red label (red font test) and "no" to labels that did not elicit such recollections (i.e., they had been studied only in black font or had not been studied). They then made a confidence judgment using a scale of 1–3. The ability to discriminate between labels that were studied with the criterial information (e.g., a picture) and labels that were studied without the criterial information provided an objective measure of the recollection of that criterial information, with "yes" responses to labels that had not been studied with the criterial information likely reflecting false recollection.[4]

Results and Discussion

Unless noted otherwise, results were significant at the conventional $p < .05$, and corresponding effect sizes were calculated with η_p^2 (F tests) or Cohen's d (t tests).

[3] We were primarily interested in the conceptual similarity of the verbal labels because we used verbal labels to assess false recollection at test. Nevertheless, we also had separate participants rate the conceptual similarity of the pictures along with their labels (n = 18 in both age groups). With this procedure, older adults gave higher similarity ratings than young adults, and negative items were rated more highly than positive items, potentially because the pictures exacerbated perceived similarity. More important, positive and negative items were rated as more conceptually similar than neutral items by both age groups, as in the other rating task (all ps < .01).

[4] We do not assume that this is a pure test of recollection, as familiarity of the test stimulus also may contribute to a participant's decisions. However, it has been shown that both young and older adults use recollection-based processes on this type of task independent from familiarity effects (Gallo et al., 2004, 2007). In the current task, the studied lures should have been familiar but should not have elicited the criterial recollection. Thus, to the degree that participants followed instructions and responded "yes" only when they believed they recollected the appropriate information, we consider "yes" responses to studied lures to have been based on some form of familiarity-induced false recollection (cf. Gallo & Roediger, 2003), especially when made with medium or high confidence.

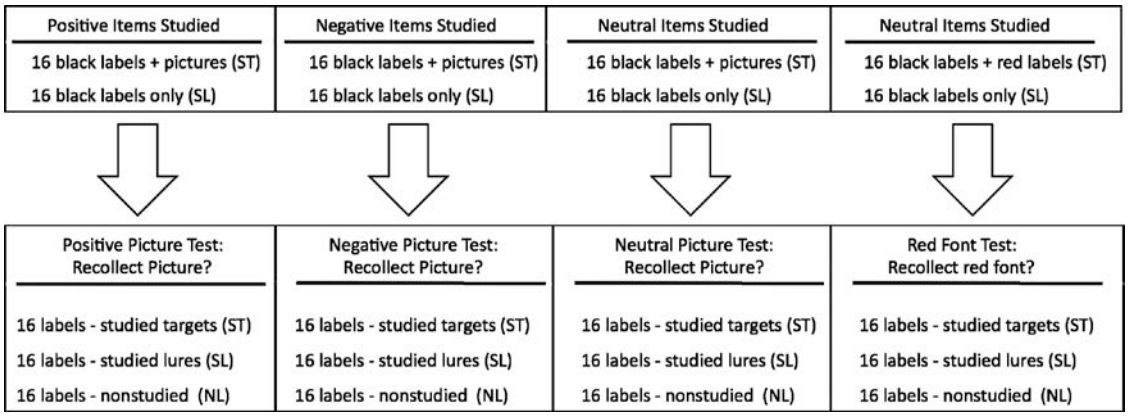

Figure 1. Experimental design. All items were intermixed during the study phase, but items from the different valence conditions were tested in separate test blocks. Comparing performance across the test blocks allowed us to compare the effects of different retrieval expectations for the different types of materials. ST = studied target; SL = studied lure; NL = nonstudied lure.

Study Ratings

In young adults, mean arousal ratings for pictures were 2.39 (positive), 2.51 (negative), and 1.67 (neutral), and corresponding ratings for labels were 1.72, 1.82, and 1.28. Consistent with our selection criteria, arousal for positive and negative items did not differ, $t(23) = 1.03$, $p = .32$, for pictures, and $t < 1$, for labels, and was greater than for neutral items (all ps $< .001$). These same means in older adults were 2.10, 2.56, and 1.49 (for pictures) and 1.89, 2.37, and 1.41 (for labels), respectively. (Due to experimental error, arousal ratings were not obtained for two older adults.) These ratings showed an identical pattern of statistical outcomes as in young adults, except positive items were rated as less arousing than negative items, $t(21) = 6.54$, $SEM = 0.07$, $d = 1.02$, for pictures; $t(21) = 7.35$, $SEM = 0.065$, $d = 0.96$, for labels. Overall, both types of emotional items were rated as more arousing than neutral items in both age groups.

Test Performance

The recollection of criterial information was quite high, as correct recognition of labels associated with pictures or red font at study (targets) was significantly greater than false recognition of studied labels that were not associated with pictures or red font (studied lures, all ps $< .001$; see Table 1). Recollection was not perfect, though, as studied lures were falsely recognized more than nonstudied lures on each test (collapsing across age groups, all ps $< .05$). False recognition of nonstudied lures was near floor and not analyzed further, although results for these lures tended to follow the same pattern as those for the studied lures described next.

To test the perceptual distinctiveness heuristic hypothesis, we compared performance across the neutral picture and red font tests. A 2 (item: targets, studied lures) × 2 (test: red font, pictures) × 2 (age) ANOVA revealed an interaction between item and test, $F(1, 46) = 17.18$, $MSE = 0.025$, $\eta_p^2 = .27$, and also between test and age, $F(1, 46) = 6.40$, $MSE = 0.038$, $\eta_p^2 = .12$. Analysis of studied lures revealed an effect of test, $F(1, 46) = 19.42$, $MSE = 0.024$, $\eta_p^2 = .30$, and no interaction with age. Consistent with the perceptual distinctiveness heuristic hypothesis, false recognition was reduced on the neutral picture test relative to the red font test in both age groups (e.g., Gallo et al., 2007). Analysis of targets revealed an interaction between test and group, $F(1, 46) = 5.25$, $MSE = 0.039$, $\eta_p^2 = .10$, as older adults recollected pictures (0.79) more than red labels (0.65), $t(23) = 2.44$, $SEM = 0.058$, $d = 0.57$, with no difference in young adults.[5]

To investigate the effects of emotion on false recollection, we compared performance for the three picture conditions. A 2 (item: targets, studied lures) × 3 (test: neutral, positive, negative) × 2 (age) ANOVA revealed an effect of item and an interaction between test and item, $F(1, 46) = 10.40$, $MSE = 0.011$, $\eta_p^2 = .18$, with no other effects or interactions. Analysis of studied lures revealed that false recognition was greater on the positive than on the neutral test (overall Ms = 0.17 and 0.08, respectively), $t(47) = 4.63$, $SEM = 0.019$, $d = 0.48$, and also was greater on the negative than on the neutral test (overall Ms = 0.14 and 0.08, respectively), $t(47) = 2.84$, $SEM = 0.019$, $d = 0.32$, with no difference between the positive and negative tests. These effects stand in contrast to the emotional distinctiveness heuristic hypothesis, which would predict lower false recognition of studied labels on the emotional tests compared with on the neutral test, and are more consistent with the conceptual overlap hypothesis. Analysis of targets revealed a difference between the negative and positive test (0.76 and 0.69, respectively), $t(47) = 2.86$, $SEM = 0.024$, $d = 0.37$, with

[5] Because the perceptual distinctiveness heuristic influences false alarms more than hits, one might expect greater discrimination (d′) and/or more conservative bias estimates (C) on the picture test compared with the word test (e.g., Schacter et al., 1999; but see Wixted & Stretch, 2000, and Gallo et al., 2004, for discussion of potential ambiguities in these estimates). Collapsing across age groups, comparison of the neutral picture and red word conditions confirmed that d′ (based on targets and studied lures) was greater on the picture test (2.11) than on the red word test (1.37), $t(47) = 5.23$, $SEM = 0.141$, $d = 0.86$, and C was more conservative, .36 and .14, respectively, $t(47) = 2.16$, $SEM = 0.103$, $d = 0.41$.

Table 1
Mean Criterial Recollection Judgments and Corresponding Confidence

	Young adults				Older adults			
	Recollection		Confidence		Recollection		Confidence	
Test/item	Mean	*SE*	Mean	*SE*	Mean	*SE*	Mean	*SE*
Red font test								
Red word targets	.72	.04	2.45	0.08	.65	.06	2.48	0.09
Studied lures	.28	.04	1.94	0.10	.17	.03	1.88	0.16
Nonstudied lures	.08	.03			.04	.03		
Neutral picture test								
Neutral targets	.67	.05	2.60	0.10	.79	.04	2.81	0.05
Studied lures	.09	.02	1.79	0.25	.08	.04	2.47	0.23
Nonstudied lures	.03	.01			.02	.01		
Positive picture test								
Positive targets	.66	.04	2.65	0.08	.71	.04	2.81	0.05
Studied lures	.18	.04	2.20	0.11	.16	.04	2.60	0.11
Nonstudied lures	.10	.03			.04	.02		
Negative picture test								
Negative targets	.75	.04	2.71	0.08	.76	.03	2.82	0.06
Studied lures	.14	.03	1.98	0.13	.13	.04	2.42	0.15
Nonstudied lures	.09	.03			.03	.01		

Note. Confidence judgments are for "yes" responses and hence are based on a variable number of responses across participants; confidence judgments range = from 1 (*low*) to 3 (*high*).

no other differences. This finding replicates two recent studies that showed more specific recollections for negative than for positive pictures in both young and older adults (Kensinger, Garoff-Eaton, et al., 2007; Kensinger, O'Brian, et al., 2007), although we found no consistent memory benefits for emotional over neutral items.[6]

Confidence Judgments

Participants were more confident when correctly recognizing targets than when falsely recognizing studied lures in all conditions (collapsing across age groups, all ps $< .01$). With respect to confidence judgments for picture hits, a 3 (test) $\times$ 2 (age) ANOVA revealed a marginal effect of age, $F(1, 46) = 3.11$, $MSE = 0.289$, $p = .09$, $\eta_p^2 = .06$, but no effect of test or interaction. The effect of age reflects somewhat higher confidence in older adults, a pattern that was even stronger in false recognition. Older adults were more confident than young adults when falsely recognizing lures on the positive test, $t(35) = 2.69$, $SEM = 0.149$, $d = 0.89$, and the negative test, $t(31) = 2.17$, $SEM = 0.204$, $d = 0.76$, and this effect was marginal on the neutral test, $t(18) = 1.97$, $SEM = 0.347$, $p = .07$, $d = 0.89$.[7] These confidence judgments generally ranged from 2 (*medium*) to 3 (*high*), which is consistent with the idea that these judgments were based, at least in part, on the illusory recollection of pictures.

General Discussion

False recollection was elevated for emotionally arousing items relative to neutral, nonarousing items in both age groups. These findings are inconsistent with the emotional distinctiveness heuristic hypothesis (e.g., Schacter et al., 2007), even though the current study provided a strong test of this hypothesis. We used complex and arousing images from the IAPS at study, we explicitly required the recollection of these images at test, and we even replicated the perceptually based distinctiveness heuristic that has been reported in both age groups. Overall, these results indicate that switching the to-be-remembered events from purely verbal information to more detailed pictures enhanced retrieval monitoring, but adding emotional content to the pictures did not. These results do not argue against previous findings of emotionally enhanced memory accuracy in young and older adults (Kensinger, O'Brian, et al., 2007), but they do suggest that those effects may have been driven solely by an emotional memory enhancement for studied events without a concurrent suppression effect on false recollection.

Our findings are more consistent with the hypothesis that emotion can increase the conceptual similarity across items (e.g., common meanings or associations), thereby increasing memory confusions (cf.

[6] We again analyzed d′ and C, on the basis of targets and studied lures and collapsing across age. In contrast to the emotional distinctiveness heuristic hypothesis, d′ was significantly greater on the neutral test (2.11) than on the positive test (1.56), $t(47) = 4.33$, $SEM = 0.126$, $d = 0.64$, and C was marginally more conservative, .36 and .26, respectively, $t(47) = 1.72$, $SEM = 0.061$, $p = .09$, $d = 0.21$. Similarly, d′ was marginally greater on the neutral test (2.11) than on the negative test (1.89), $t(47) = 1.72$, $SEM = 0.125$, $p = .09$, $d = 0.26$, and C was significantly more conservative, .36 and .24, respectively, $t(47) = 2.13$, $SEM = 0.059$, $d = 0.25$. Unlike in our analysis of the perceptual distinctiveness heuristic hypothesis, all of these differences were in the opposite direction as would be predicted by an emotional distinctiveness heuristic hypothesis. Finally, d′ was greater for negative items (1.89) than for positive items (1.56), $t(47) = 3.88$, $SEM = 0.085$, $d = 0.49$, reflecting the difference observed in hit rates.

[7] Many participants did not falsely recognize neutral studied lures, and so these comparisons were based on a subset of participants (reducing power and precluding ANOVA). Despite this limitation, we found a large aging effect on confidence judgments with all three comparisons.

Brainerd et al., 2008). Supporting this hypothesis, we obtained ratings of conceptual overlap for the content portrayed in the IAPS pictures and found that the emotional items were more related to each other than were neutral items. These similarities likely were due to the specific content that tends to elicit emotional reactions (people, activities, and so forth), but it also is possible that these ratings were influenced by the emotional reactions themselves (happy items, sad items, and so forth). Due to the complex nature of the IAPS, we did not control specific content at this level. Nevertheless, using words that were equated on semantic relatedness, Kapucu et al. (2008) also found elevated false recognition to emotional items, suggesting that general emotional reactions are more critical. More broadly, there is evidence that emotion can enhance retrieval of the general features of pictures (e.g., Denburg et al., 2003; Kensinger, Garoff-Eaton, et al., 2007). If these general features overlap with nonstudied but conceptually related emotional items, then confusion between the various features could enhance false recollection.

Our findings also are consistent with studies that have shown more liberal recognition memory bias for emotional than for neutral words, as we discussed in the Introduction. Response bias effects do not implicate a particular theoretical mechanism, but they are consistent with the idea that enhanced conceptual similarity for emotional items can increase the probability of endorsing any item that is perceived to be related to studied information. The similarity ratings of the current study provided additional evidence for this relation. Moreover, the current study extends these prior results to false recollection of more perceptually detailed and arousing events. While speculative, the idea that conceptual similarity between emotional events can increase memory distortion may be generalized even further, potentially explaining aspects of false emotional content in autobiographical memory (e.g., Laney & Loftus, 2008; Porter, Taylor, & ten Brinke, 2008).

In conclusion, our study provided two theoretically important age comparisons. First, we found that emotionally arousing content increased false recollection in both young and older adults. These results are consistent with aging studies on the effects of emotion on memory for words (Budson et al., 2006; Kapucu et al., 2008), and they extend these effects to false recollection of emotionally complex IAPS pictures. Also consistent with the results of these latter studies, our results did not show the memory benefit for positive information that is sometimes found in older adults (e.g., Carstensen, Mikels, & Mather, 2006). Although age may increase memory for positive information in some conditions, our results add to those of other studies showing that this effect is not always obtained (for a review, see Murphy & Isaacowitz, 2008).

Second, even though we minimized overall age differences in objective accuracy (via retention interval), older adults were more confident in false recollection of emotional pictures compared with young adults. This finding supports the idea that age enhances the subjective experience of illusory recollection (Dodson et al., 2007), and it extends this finding to memory for more detailed emotional events. This finding is of more general interest because many of the important everyday memories reported by older adults tend to contain emotionally arousing content (especially positive events; see Berntsen & Rubin, 2002). Our results suggest that older adults may be prone to overconfidence in the accuracy of their emotional memories.

References

Berntsen, D., & Rubin, D. C. (2002). Emotionally charged autobiographical memories across the life span: The recall of happy, sad, traumatic, and involuntary memories. *Psychology and Aging, 17,* 636–652.

Bower, G. H. (1981). Mood and memory. *American Psychologist, 36,* 129–148.

Brainerd, C. J., Stein, L. M., Silveira, R. A., Rohenkohl, G., & Reyna, V. F. (2008). How does negative emotion cause false memories? *Psychological Science, 19,* 919–925.

Brink, T. L., Yesavage, J. A., Lum, O., Heersema, P., Adey, M. B., & Rose, T. L. (1982). Screening tests for geriatric depression. *Clinical Gerontologist, 1,* 37–44.

Budson, A. E., Todman, R. W., Chong, H., Adams, E. H., Kensinger, E. A., Krangel, T. S., & Wright, C. I. (2006). False recognition of emotional word lists in aging and Alzheimer disease. *Cognitive Behavioral Neurology, 19,* 71–78.

Cahill, L., & McGaugh, J. L. (1998). Mechanisms of emotional arousal and lasting declarative memory. *Trends in Neurosciences, 21,* 294–299.

Carstensen, L. L., Mikels, J. A., & Mather, M. (2006). Aging and the intersection of cognition, motivation, and emotion. In J. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (6th ed., pp. 343–362). San Diego, CA: Academic Press.

Comblain, C., D'Argembeau, A., Van der Linden, M., & Aldenhoff, L. (2004). The effect of ageing on the recollection of emotional and neutral pictures. *Memory, 12,* 673–684.

Cook, G. I., Hicks, J. L., & Marsh, R. L. (2007). Source monitoring is not always enhanced for valenced material. *Memory & Cognition, 35,* 222–230.

D'Argembeau, A., & Van der Linden, M. (2004). Influence of affective meaning on memory for contextual information. *Emotion, 4,* 173–188.

Denburg, N. L., Buchanan, T. W., Tranel, D., & Adolphs, R. (2003). Evidence for preserved emotional memory in normal older persons. *Emotion, 3,* 239–253.

Dodson, C. S., Bawa, S., & Krueger, L. E. (2007). Aging, metamemory, and high-confidence errors: A misrecollection account. *Psychology and Aging, 22,* 122–133.

Doerksen, S., & Shimamura, A. P. (2001). Source memory enhancement for emotional words. *Emotion, 1,* 5–11.

Dougal, S., & Rotello, C. M. (2007). "Remembering" emotional words is based on response bias, not recollection. *Psychonomic Bulletin and Review, 14,* 423–429.

Emery, L., & Hess, T. M. (2008). Viewing instructions impact emotional memory differently in older and young adults. *Psychology and Aging, 23,* 2–12.

Fernandes, M., Ross, M., & Wiegand, M. (2008). Are the memories of older adults positively biased? *Psychology and Aging, 23,* 297–306.

Folstein, M F., Folstein, S. E., & McHugh, P. R. (1975). Mini-Mental State: A practical method for grading the mental state of patients for the clinician. *Journal of Psychiatric Research, 12,* 189–198.

Gallo, D. A. (2006). *Associative illusions of memory: False memory research in DRM and related tasks.* New York, NY: Psychology Press.

Gallo, D. A., Cotel, S. C., Moore, C. D., & Schacter, D. L. (2007). Aging can spare recollection-based retrieval monitoring: The importance of event distinctiveness. *Psychology and Aging, 22,* 209–213.

Gallo, D. A., & Roediger, H. L., III. (2003). The effects of associations and aging on illusory recollection. *Memory & Cognition, 31,* 1036–1044.

Gallo, D. A., Weiss, J. A., & Schacter, D. L. (2004). Reducing false recognition with criterial recollection tests: Distinctiveness heuristic versus criterion shifts. *Journal of Memory & Language, 51,* 473–493.

Hamann, S. (2001). Cognitive and neural mechanisms of emotional memory. *Trends in Cognitive Sciences, 5,* 394–400.

Howe, M. L. (2007). Children's emotional false memories. *Psychological Science, 18,* 856–860.

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

Kensinger, E. A. (2007). Negative emotion enhances memory accuracy: Behavioral and neuroimaging evidence. *Current Directions in Psychological Science, 16,* 213–218.

Kensinger, E. A., Brierley, B., Medford, N., Growdon, J. H., & Corkin, S. (2002). Effects of normal aging and Alzheimer's disease on emotional memory. *Emotion, 2,* 118–134.

Kensinger, E. A., & Corkin, S. (2004). The effects of emotional content and aging on false memories. *Cognitive, Affective, & Behavioral Neuroscience, 4,* 1–9.

Kensinger, E. A., Garoff-Eaton, R. J., & Schacter, D. L. (2007a). Effects of emotion on memory specificity in young and older adults. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 62,* 208–215.

Kensinger, E. A., O'Brian, J. L., Swanberg, K., Garoff-Eaton, R. J., & Schacter, D. L. (2007b). The effects of emotional content on reality-monitoring performance in young and older adults. *Psychology and Aging, 22,* 752–764.

Kensinger, E. A., & Schacter, D. L. (2005a). Emotional content and reality-monitoring ability: FMRI evidence for the influences of encoding processes. *Neuropsychologia, 43,* 1429–1443.

Kensinger, E. A., & Schacter, D. L. (2005b). Retrieving accurate and distorted memories: Neuroimaging evidence for effects of emotion. *NeuroImage, 27,* 167–177.

Kensinger, E. A., & Schacter, D. L. (2006). Reality monitoring and memory distortion: Effects of negative, arousing context. *Memory & Cognition, 34,* 251–260.

Kapucu, A., Rotello, C. M., Ready, R. E., & Seidl, K. N. (2008). Response bias in "remembering" emotional stimuli: A new perspective on age differences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 703–711.

LaBar, K. S. (2007). Beyond fear: Emotional memory mechanisms in the human brain. *Current Directions in Psychological Science, 16,* 173–177.

Laney, C., & Loftus, E. F. (2008). Emotional content of true and false memories. *Memory, 16,* 500–516.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2005). *International Affective Picture System (IAPS): Affective ratings of pictures and instruction manual* (Technical Report No. A-6). Gainesville: University of Florida.

Mather, M. (2007). Emotional arousal and memory binding: An object-based framework. *Perspectives on Psychological Science, 2,* 33–52.

Mather, M., Canli, T., English, T., Whitfield, S., Wais, P., Ochsner, K., . . . & Carstensen, L. L. (2004). Amygdala responses to emotionally valenced stimuli in older and younger adults. *Psychological Science, 15,* 259–263.

May, C. P., Rahhal, T., Berry, E. M., & Leighton, E. A. (2005). Aging, source memory, and emotion. *Psychology and Aging, 20,* 571–578.

Murphy, N. A., & Isaacowitz, D. M. (2008). Preferences for emotional information in older and younger adults: A meta-analysis of memory and attention tasks. *Psychology and Aging, 23,* 263–286.

Ochsner, K. N. (2000). Are affective events richly "remembered" or simply familiar? The experience and process of recognizing feelings past. *Journal of Experimental Psychology: General, 129,* 242–261.

Pesta, B. J., Murphy, M. D., & Sanders, R. E. (2001). Are emotionally charged lures immune to false memory? *Journal of Experimental Psychology: Learning, Memory, & Cognition, 27,* 328–338.

Piguet, O., Connally, E., Krendl, A. C., Huot, J. R., & Corkin, S. (2008). False memory in aging: Effects of emotional valence on word recognition accuracy. *Psychology and Aging, 23,* 307–314.

Porter, S., Taylor, K., & ten Brinke, L. (2008). Memory for media: Investigation of false memories for negatively and positively charged public events. *Memory, 16,* 658–666.

Rahhal, T. A., May, C. P., & Hasher, L. (2002). Truth and character: Sources that older adults can remember. *Psychological Science, 13,* 101–105.

Schacter, D. L., Gallo, D. A., & Kensinger, E. A. (2007). The cognitive neuroscience of implicit and false memories: Perspectives on processing specificity. In J. S. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger III* (pp. 353–377). New York, NY: Psychology Press.

Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language, 40,* 1–24.

Schmidt, S. R. (2007). Unscrambling the effects of emotion and distinctiveness on memory. In J. S. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger III* (pp. 141–158). New York, NY: Psychology Press.

Sison, J. A. G., & Mather, M. (2007). Does remembering emotional items impair recall of same-emotion items? *Psychonomic Bulletin & Review, 14,* 282–287.

Talmi, D., Luk, B. T. C., McGarry, L. M., & Moscovitch, M. (2007). The contribution of relatedness and distinctiveness to emotionally-enhanced memory. *Journal of Memory and Language, 56,* 555–574.

Tun, P. A., Wingfield, A., Rosen, M. J., & Blanchard, L. (1998). Response latencies for false memories: Gist-based processes in normal aging. *Psychology and Aging, 13,* 230–241.

Windmann, S., & Kutas, M. (2001). Electrophysiologial correlates of emotion-induced recognition bias. *Journal of Cognitive Neuroscience, 13,* 577–592.

Wixted, J. T., & Stretch, V. (2000). The case against a criterion-shift account of false memory. *Psychological Review, 107,* 368–376.

(*Appendix follows*)

Appendix

Conceptual Relatedness Norms

All of the labels were divided into 32 lists of six items each, with each list blocked by valence (positive, negative, or neutral). Each participant was presented with eight positive lists, eight negative lists, and eight of the 16 neutral lists (counterbalanced across participants), with list order randomized. Each list was printed on a sheet of paper, along with a rating scale for emotional arousal (ranging from −2 to +2) and content overlap (ranging from 1 to 5). Participants were instructed as follows:

> The attached pages contain several sets of phrases. Each set contains six phrases. These phrases vary in terms of the emotional arousal that they elicit (positive, neutral, or negative feelings) and also in terms of the specific content that they describe (e.g., objects, people, actions, locations, and so forth.). We would like you to make two ratings for each set of phrases.

Emotional Arousal

The first rating is for emotional arousal. After reading through the six phrases in the set, please rate the AVERAGE level of emotional arousal across all of the phrases within that set, ranging from −2 (very negative) to 0 (neutral) to +2 (very positive). By *emotional arousal*, we mean the strength or intensity of the feelings that the phrases elicit, with some phrases referring to very negative items or events, others referring to very positive items or events, and others being more neutral.

Note that the phrases within each set may vary in terms of the emotions they elicit. Nevertheless, your job is to consider all six of the phrases in the set and report a single "average" level of emotional arousal across the six phrases. You will find that some sets have many positive phrases, others have many negative phrases, and others have many neutral phrases. We want your ratings to reflect these average levels of emotion across all of the phrases within each set.

Content Overlap

The second rating is for content overlap. After reading through the six phrases in the set, please rate the AVERAGE degree of content overlap across all of the phrases within that set, ranging from 1 (very low) to 5 (very high). By *content overlap*, we mean the specific types or kinds of items and events depicted in the phrases, such as people, actions, objects, and locations. Do not let the emotional similarity of the phrases influence this rating, but instead focus only on the degree that they contain similar kinds of items and events.

Note that the phrases within each set may very in terms of the content they contain. Nevertheless, your job is to consider all six of the phrases in the set and report an "average" level of content overlap across the six phrases. You will find that some sets have many phrases with overlapping content (e.g., many phrases in one set may refer to objects, whereas many phrases in another set may refer to people) and that other sets have only a few phrases with overlapping content. We want your ratings to reflect these average levels of specific content overlap across all of the phrases within each set.

Note that there are no right or wrong answers in this task. Instead we want to separately measure the emotional arousal and content overlap for each set of phrases. Be sure to use the full scale when making your judgments (circling only one number), and do not take longer than a minute for each set. Do you have any questions?

Received August 20, 2008
Revision received June 22, 2009
Accepted September 4, 2009 ■