



Feels like the real thing: Imagery is both more realistic and emotional than verbal thought

Andrew Mathews^{1,2}, Valerie Ridgeway, and Emily A. Holmes³

¹Department of Psychology, University of California, Davis, CA, USA

²Psychology Department, King's College, London, UK

³Department of Psychiatry, University of Oxford, Oxford, UK

The production of mental images involves processes that overlap with perception and the extent of this overlap may contribute to reality monitoring errors (i.e., images misremembered as actual events). We hypothesised that mental images would be more confused with having actually seen a pictured object than would alternative representations, such as verbal descriptions. We also investigated whether affective reactions to images were greater than to verbal descriptions, and whether emotionality was associated with more or less reality monitoring confusion. In two experiments signal detection analysis revealed that mental images were more likely to be confused with viewed pictures than were verbal descriptions. There was a general response bias to endorse all emotionally negative items, but accuracy of discrimination between imagery and viewed pictures was not significantly influenced by emotional valence. In a third experiment we found that accuracy of reality monitoring depended on encoding: images were more accurately discriminated from viewed pictures when rated for affect than for size. We conclude that mental images are both more emotionally arousing and more likely to be confused with real events than are verbal descriptions, although source accuracy for images varies according to how they are encoded.

Keywords: Mental imagery; Emotional arousal; Reality monitoring; Verbal thought.

Episodic memories for past events, or events that could happen in the future, can be experienced as mental images or as verbally mediated thoughts (or sometimes as a mixture of both). On your way home from work, for example, you might say to yourself, “I wonder if I left the office door open” (a verbal thought) or you might see in your mind’s eye the office door wide open (a mental image). Content analysis of reported mental images and

verbal thoughts suggests that they differ in a number of ways; for example, image descriptions are more likely to contain references to sensory characteristics (Holmes, Mathews, Mackintosh, & Dalgleish, 2008).

Consistent with such descriptive reports, converging evidence suggests that mental imagery involves some of the same processes as are employed when perceiving real objects. First,

Correspondence should be addressed to: Andrew Mathews, Department of Psychology, 1 Shields Avenue, Davis, CA 95616-8686, USA. E-mail: andrew.mathews@sbcglobal.net

Emily A. Holmes was supported by a Wellcome Trust Clinical Fellowship (WT088217).

We would like to thank Paula Hertel and Marcia Johnson for helpful comments on an earlier version of this paper.

there is evidence of competition between mental imagery and perceptual processing when they share the same sensory modality. Holding a visual image selectively interferes with the detection of a faint visual signal, and, likewise, auditory images interfere with the detection of auditory stimuli (Segal & Fusella, 1969). The reverse relationship also holds: judged vividness of visual images is reduced by simultaneous performance of a visuospatial task, and auditory image vividness is decreased by counting aloud (Baddeley & Andrade, 2000). This mutual interference strongly suggests that mental images and perceptual processes draw on overlapping cognitive resources.

Second, neuroimaging studies have revealed that visual mental imagery activates areas in early visual cortex (Kosslyn & Thompson, 2003) when making comparative judgements of imagined shapes. Visual cortex is not the only brain area revealing overlap between the activation associated with imagery and perception; rather, the areas activated depend on the type of imagery involved. In a whole brain activation study, Ganis, Thompson, and Kosslyn (2004) concluded that visual imagery and perception draw on similar neural machinery, with considerable overlap in frontal and parietal areas, and some, albeit less complete, overlap in temporal and occipital areas. Strikingly, when the perception of different types of object activates different processing areas, imagination of those objects does too. For example, activation of the fusiform face area is greater than of the parahippocampal place area when perceiving faces, relative to the pattern seen when perceiving places. The same selective activation pattern emerges when people simply imagine familiar faces or places, albeit at lower levels of intensity (O'Craven & Kanwisher, 2000). Thus, imagery selectively activates the same areas as are involved in processing perceived objects.

Although less extensively documented, similar conclusions apply to the perception and imagination of emotional scenes. Looking at faces with negative emotional versus neutral expressions activates several different brain areas, but particularly the amygdala. This pattern is also seen when facial expressions are simply imagined (Kim et al.,

2007). The imagination of future emotional events, as well as the recall of past emotional episodes, similarly activates the amygdala (Cabeza & St Jacques, 2007; Sharot, Riccardi, Raio, & Phelps, 2007). In sum, mental imagery activates many of the brain systems involved in equivalent forms of perception, and—when the imagery is emotional in content—brain systems involved in processing emotional information, in much the same way as with perceived events.

Reality monitoring errors, that is, confusions between imagined and actual events, provide additional evidence that rather than there being entirely separate mechanisms underlying memories for real versus imagined events, judgements about the source of memories depend on features such as the amount of perceptual and emotional detail they include, and on their consistency with other knowledge (Johnson, 2006; Johnson & Raye, 1981). People reporting more vivid images tend to make more reality monitoring errors (Dobson & Markham, 1993), presumably because vividness ratings reflect the extent of similarity between activation due to images and perceptual experiences. Research by Johnson and colleagues also suggests that focusing on personal feelings can increase source confusion. When participants listened to emotional or neutral statements with instructions to focus either on the speaker's or their own feelings, the latter instruction resulted in better memory for what was said, but poorer memory for who said it (Johnson, Nolde, & De Leonardi, 1996). In reviewing this work, Johnson (2006) has suggested that a focus on one's own affective state may reduce source-monitoring accuracy by decreasing attention to other features that could otherwise help to distinguish the source.

In apparent contrast, Kensinger and colleagues have reported that reality monitoring may be *more* accurate for emotional than neutral imagery. Participants saw a series of word captions, half negative and half neutral, and decided whether the object described was bigger or smaller than a shoe box (Kensinger & Schacter, 2005). Half the captions were followed by a matching picture (e.g., the word *frog*, followed by a picture of a frog) and half by a blank square, so that

participants presumably had to imagine the object to make their size judgement. One or two days later, participants heard the old captions, mixed with new ones, and decided whether each had been followed by a picture. Incorrect endorsements of a previously imagined caption as having been viewed were more frequent for neutral than negative captions, and both were endorsed more than new captions. These results were replicated and extended in other experiments (Kensinger, O'Brien, Swanberg, Garoff-Eaton & Schacter, 2007; Kensinger & Schacter, 2006). For example, in one experiment participants heard words and then either saw or imagined them while comparing the first and last letter size. In later source memory judgements many of the imagined words were incorrectly judged as previously seen, but with emotional words having lower misattribution rates.

Before concluding that images, whether emotional or neutral, are especially likely to be confused with having perceived an event, it should be noted that existing reality monitoring studies have not usually contrasted imagery with any alternative representational form. In a rare exception, Hyman and Pentland (1996) contrasted the effects of instructions to either imagine a (false) childhood event or just think about it, and found that those instructed to imagine the event were more likely to report later that the event was real. Although these data are consistent with the hypothesis that imagery is more confusable with perceived events than are alternative representational forms, the absence of precise information about what "thinking about the event" actually involved leaves room for uncertainty about the meaning of this finding. Similarly, later elaboration on the conceptual or perceptual properties of misleading information about a previously seen video can increase erroneous reports that this information had actually been viewed, in contrast to a control condition involving non-elaborative verbal manipulations (Zaragoza, Mitchell, Payment, & Drivdahl, 2011). Thus, subsequent elaboration can increase the extent to which (false)

memories are confused with having actually experienced an event. Despite these suggestive findings, and perhaps surprisingly, the widely held assumption that imagination is especially prone to being confused with actually perceiving an object has not yet been adequately investigated in studies that experimentally manipulated the form of encoded representation by comparing instructions to use mental images with alternatives such as verbal description. The primary aim of the present experiments was to investigate the assumption that mental imagery is more likely to be confused with having actually seen a picture than is an alternative (verbal) form of representation. We also investigated whether imagery would be associated with a greater emotional response than verbal processing of the same event (cf. Holmes & Mathews, 2005). Finally, we asked whether the emotional content of imagery would interact with reality monitoring accuracy, although—in the light of the mixed evidence described above—without making a specific prediction about the direction of reality monitoring differences due to emotion.

EXPERIMENT 1

Method

The method used here was adapted from that described by Gonsalves and Paller (2000) and Kensinger and Schacter (2005, 2006). Participants saw 216 word captions, half negative and half benign,¹ followed in 72 trials by a corresponding picture (cued by the word "look"), in another 72 trials by generation of a mental image (cued by "imagine"), or construction of a descriptive sentence in the remaining 72 (cued by "sentence"). The captions were divided into three matched sets of 72 each (36 negative and 36 benign), with similar content across sets (e.g., the same number of animals, humans or inanimate objects in each). Assignment of sets to conditions (look, imagine, or sentence) was counterbalanced across participants.

¹We generally use the term "benign" rather than "neutral" because captions judged to be neutral when presented alone sometimes elicited images that were rated as being mildly positive.

Each trial ended with a pleasantness/unpleasantness rating of the picture, image or sentence on a 1–5 scale. A day later, participants completed a source memory questionnaire that listed all the captions seen the previous day, together with instructions to report for each one whether or not they thought that the caption had previously been followed by a picture.

Reality monitoring accuracy is often assessed using raw false alarm rates: that is, the number of occasions on which imagined items are later falsely identified as having been seen previously. However, a potential problem with relying on false alarm rates as the sole index of source monitoring accuracy is that false alarms can also be influenced by variations in the response criteria used (e.g., a greater willingness to endorse certain types of item). For example, if a more lax response criterion is used for endorsing emotional images then the resulting higher false alarm rates may be taken as evidence of less accurate reality monitoring, even if the same response bias applies to emotional items that were actually perceived. If so, then false alarm rates alone provide a misleading index of accuracy in discriminating between imagined and perceived items. Signal detection analysis (as used here), provides an index of sensitivity (d' ; the difference between standardised hit and false alarm rates) that allows assessment of source monitoring accuracy independent of response criterion (c ; the mean of standardised hit and false alarm rates; see Macmillan & Creelman, 2005). Advantages of the signal-detection approach to source memory have been discussed previously by Brown, Kosslyn, Breiter, Baer, and Jenike (1994) and by Slotnick, Klein, Dodson, and Shimamura (2000).

Participants and procedure. Forty-two undergraduates (14 male) took part, and received course credit for their participation. Instructions were presented by computer, followed by three practice and 216 experimental trials. Participants were instructed that (depending on the cue presented) they should either mentally imagine the object (or

event) described in the caption, or construct a sentence in their head that described that object, or just look at the displayed picture. Trials began with a central “Ready?” display that remained until participants pressed the space bar. This initiated central presentation of a cue word for 1,500 ms prompting the action to be performed on that trial (i.e., “look”, “imagine”, or “sentence”), followed by a caption for 1,500 ms specifying what object would be displayed, imagined or described. In “look” trials, the caption was replaced by a corresponding picture displayed centrally for four seconds (captions and pictures were taken from Kensinger & Schacter, 2005).² In “imagine” trials the screen was darkened for four seconds during mental imagery; in “sentence” trials, the screen was illuminated but blank for four seconds while participants generated a descriptive verbal sentence. In half of each trial type captions were emotionally negative (e.g., snake), and in half they were benign (e.g., sheep). After four seconds, participants were prompted to rate the picture, or their mental image, or their sentence, using a 1–5 scale (1 = *Very pleasant*; 2 = *Pleasant*; 3 = *Neutral*; 4 = *Unpleasant*; and 5 = *Very unpleasant*).

After all trials were complete participants were instructed that the second part of the experiment would involve completing a questionnaire the following day, but they were not informed (in this or in subsequent experiments) about its content nor that the questionnaire tested memory. The next day participants were sent and completed the source memory questionnaire (via e-mail) that listed all the captions seen previously in random order, and responded either “yes” or “no” according to whether or not they thought they had seen a corresponding picture following each caption.

Results

Affective ratings. The rating data were used to test the second hypothesis, that emotional response to imagery would be greater than to verbal descriptions. Mean affective ratings were first entered into

²Caption and picture sets were kindly supplied by Elizabeth Kensinger.

Table 1. Mean (and standard deviations) of the number of “yes” responses (hits in the case of pictures or false alarms otherwise), signal detection sensitivity (d') scores, response bias (c) and affective ratings, for each trial type (hit rates for Experiment 3 are based on 18 cases rather than 36)

		Hits/FAs	d'	c	Rating
<i>Experiment 1</i>					
Picture					
	Negative	27.50 (5.06)	—	—	3.72 (0.59)
	Benign	23.81 (5.54)	—	—	2.57 (0.33)
Image					
	Negative	5.49 (4.35)	1.99 (0.82)	0.20 (0.38)	3.70 (0.60)
	Benign	4.21 (4.39)	1.88 (0.88)	0.47 (0.40)	2.46 (0.30)
Sentence					
	Negative	4.93 (3.95)	2.06 (0.80)	0.24 (0.38)	3.55 (0.55)
	Benign	2.83 (3.00)	2.08 (0.76)	0.58 (0.35)	2.55 (0.39)
<i>Experiment 2</i>					
Picture					
	Negative	27.84 (4.88)	—	—	4.10 (0.29)
	Benign	25.90 (5.82)	—	—	2.60 (0.30)
Image					
	Negative	5.75 (3.92)	1.95 (0.69)	0.15 (0.37)	3.97 (0.28)
	Benign	5.21 (4.06)	1.91 (0.95)	0.27 (0.37)	2.46 (0.29)
Sentence					
	Negative	5.39 (4.06)	2.10 (0.95)	0.22 (0.42)	3.82 (0.28)
	Benign	4.51 (3.81)	2.07 (0.99)	0.36 (0.45)	2.56 (0.28)
None					
	Negative	3.29 (2.91)	2.42 (0.90)	0.38 (0.41)	—
	Benign	2.51 (2.74)	2.49 (1.06)	0.57 (0.46)	—
<i>Experiment 3</i>					
Picture (affect)					
	Negative	14.14 (2.63)	—	—	—
	Benign	12.52 (3.70)	—	—	—
Picture (size)					
	Negative	13.07 (2.39)	—	—	—
	Benign	10.50 (3.45)	—	—	—
Image (affect)					
	Negative	5.36 (4.19)	2.03 (0.74)	0.18 (0.37)	—
	Benign	5.24 (4.36)	1.82 (0.80)	0.34 (0.41)	—
Image (size)					
	Negative	4.67 (3.65)	1.96 (0.74)	0.31 (0.33)	—
	Benign	5.07 (4.58)	1.40 (0.85)	0.46 (0.47)	—

a repeated-measure analysis of variance (ANOVA) having within-participant factors of Source (images, sentences or pictures) and Emotional Valence (negative or benign). In this and subsequent analyses the assumption of sphericity for comparisons involving more than two conditions was confirmed (unless otherwise stated, when the Greenhouse–Geisser correction was used). ANOVA revealed the expected main effect due to Valence, with negative items being rated as more unpleasant than benign items, $F(1, 41) = 214.03$,

$p < .001$, partial eta-square (η_p^2) = .84, qualified by an interaction with source, $F(2, 82) = 11.04$, $p < .001$, $\eta_p^2 = .21$. This interaction remained significant in a planned contrast of image with sentence trials, $F(1, 41) = 28.27$, $p < .001$, $\eta_p^2 = .41$. For negative items, images were rated as being more unpleasant than sentences, $t(41) = 4.87$, $p < .001$, $d = 0.75$, whereas for benign items, images were rated as more pleasant than sentences, $t(41) = 2.53$, $p < .05$, $d = 0.39$ (see Table 1 for means).

Source memory. In an initial exploration of the first hypothesis (concerning the effects of trial type on the accuracy of source memory) within-participant analyses were performed using either hit and false alarm rates, derived from responses on the source memory questionnaire. The mean number of hits (correct “yes” responses) for captions preceding pictures was higher for negative than for benign pictures, $t(41) = 5.38$, $p < .01$, $d = 0.83$. For false alarm rates (incorrect “yes” responses for captions preceding images or sentences), a repeated-measure ANOVA with factors of Source (image or sentence) and Valence (negative or benign), revealed significant main effects of Source, with more false alarms for image than sentence trials, $F(1, 41) = 10.61$, $p < .01$, $\eta_p^2 = .21$, and of Valence, with more false alarms for negative than benign items, $F(1, 41) = 14.22$, $p < .01$, $\eta_p^2 = .26$. The interaction of Source with Valence was not significant. The analysis of false alarm data thus seemed consistent with the hypothesis of less accurate reality monitoring for images than verbal descriptions, but also suggested that source monitoring was generally less accurate for emotional items.

For reasons noted earlier, the main analysis of source memory accuracy was conducted using a signal detection measure of sensitivity (d'). Sensitivity scores were computed using hit rates for pictures and false alarm rates for images or sentences (with zero values converted to $1/72$; Macmillan & Creelman, 2005) entered into a repeated-measure ANOVA having within-participant factors of Source (image or verbal description) and Valence (negative or benign). There was a main effect of Source, with lower sensitivity (less accurate discrimination) for images than sentences, $F(1, 41) = 5.51$, $p < .05$, $\eta_p^2 = .12$. Neither the main effect of emotional Valence nor the interaction of Valence with Source approached significance, $F_s < 1$. The only significant effect in a similar analysis of response bias scores (c) was for Valence, $F(1, 41) = 42.46$, $p < .001$, $\eta_p^2 = .51$, with a more lax criterion for emotionally negative items.

In summary, initial analysis of false alarms alone suggested that reality monitoring was less

accurate for both images and emotionally negative items. However, analysis of signal detection sensitivity scores (d') indicated that although participants were indeed less accurate in distinguishing previously viewed pictures from mental images than from descriptive sentences, there was no significant difference due to emotional valence. In contrast, the analysis of response criterion scores (c) indicated a more lax criterion was used for emotionally negative items, with corresponding captions being more likely to be endorsed than benign items. This last finding suggests that the higher false alarm rate for emotional items can be attributed to a more lax response criterion, rather than to reduced source monitoring accuracy.

EXPERIMENT 2

Method

Experiment 2 was designed as a replication of Experiment 1, but with a new set of 72 captions that were not presented at all during the main part of the experiment, and that appeared for the first time in the source memory questionnaire. This addition was intended to provide a baseline measure of false alarm rates for items that had not previously been presented and thus had not been either imagined or verbally described. This allowed us to test not only whether images were less accurately discriminated from pictures than were verbal descriptions, but also whether verbal descriptions led to more reality monitoring errors than did new captions.

The three previously used sets of captions were reassigned to image trials, sentence trials, or appeared for the first time in the questionnaire, with set assignment counterbalanced across participants. A new set of 72 matched captions and corresponding pictures was selected and used only in “look” trials.

Participants and procedure. Fifty-one undergraduates (4 male) took part and received course credit for their participation (two sets of affective ratings were lost). Procedure was the same as Experiment 1

with the exception of the addition of non-exposed captions in the memory questionnaire.

Results

Affective ratings. These data were analysed as before using a repeated-measure ANOVA having within-participant factors of Source (images, sentences or pictures) and Valence (negative or benign). There was a main effect of Valence due to negative items being rated as more unpleasant than benign items, $F(1, 48) = 832.85$, $p < .001$, $\eta_p^2 = .95$, that was qualified by an interaction with Source, $F(1.42, 96) = 10.06$, $p < .001$, $\eta_p^2 = .17$ (Mauchly tests of sphericity revealed significant differences in variance so the degrees of freedom used to test the source by valence interaction were reduced according to the Greenhouse–Geisser correction). This interaction remained significant in a planned comparison of image versus sentence captions, $F(1, 48) = 12.51$, $p < .001$, $\eta_p^2 = .21$. For negative items, images were rated as being more unpleasant than sentences, $t(48) = 3.19$, $p < .002$, $d = 0.46$, whereas for benign items, images were rated as being more pleasant than sentences, $t(48) = 2.54$, $p < .02$, $d = 0.36$.

Source memory. Initial analyses were conducted using a within-participant comparison of hit rates for captions from picture trials, and of false alarm rates from image or sentence trials, or that were new. Comparison of mean hit rates for captions that had preceded negative or benign pictures again showed a significant effect of Valence, with higher hit rates for captions preceding negative rather than benign pictures, $t(50) = 4.07$, $p < .001$, $d = 0.57$. A repeated-measure ANOVA of false alarm rates, having within-participant factors of Source (images, sentences or new) and Valence (negative or benign) revealed significant main effects of both Source, $F(2, 100) = 35.19$, $p < .001$, $\eta_p^2 = .41$; and Valence, $F(1, 50) = 5.53$, $p < .03$, $\eta_p^2 = .10$. The interaction of Source with Valence was not significant. As expected, false alarms were lowest for new captions (see Table 1 for means). The main effect of Valence remained significant in a planned contrast of responses to

captions from image and sentence trials, $F(1, 50) = 5.53$, $p < .03$, $\eta_p^2 = .10$, with more false alarms to negative than benign items; while the effect of Source fell short of significance, $F(1, 50) = 3.56$, $p < .06$, $\eta_p^2 = .07$.

The main analysis of reality monitoring accuracy employed sensitivity (d') scores, derived as before and submitted to a repeated-measure ANOVA having two within-participant factors, Source (image, sentence or new) and Valence (negative or benign). This revealed only one significant effect, due to Source, $F(2, 100) = 24.74$, $p < .001$, $\eta_p^2 = .33$. As expected, new captions were more accurately rejected than were those seen previously. In a planned comparison of image and sentence trials, images were more likely to be misattributed than sentences, $F(1, 50) = 5.65$, $p < .03$, $\eta_p^2 = .10$. However, sentence captions were less accurately discriminated from pictures than were new captions, $F(1, 50) = 23.23$, $p < .01$, $\eta_p^2 = .32$. Again, neither the main effect of emotional Valence nor the interaction with Source was significant, $F_s < 1$.

A similar analysis of response bias scores (c) revealed main effects of Source, $F(2, 100) = 24.74$, $p < .001$, $\eta_p^2 = .33$, and emotional Valence, $F(1, 50) = 14.25$, $p < .001$, $\eta_p^2 = .22$, but no significant interaction between them. Both main effects remained significant on analysis of image and sentence trials, with a more lax response criterion for negative items, $F(1, 50) = 10.86$, $p < .001$, $\eta_p^2 = .18$, and for images than for sentences, $F(1, 50) = 5.65$, $p < .05$, $\eta_p^2 = .10$.

In exploratory analyses, we looked for evidence that the two effects of imagery found here (reduced source accuracy and enhanced emotion) were related, but found none. The correlation between mean affective ratings given to negative images and the corresponding value of d' (computed across both Experiments 1 and 2) was far from significant, $r(91) = .10$, $p = .33$ (the same correlation for benign images was $r(91) = -.08$, $p = .46$). This suggests that the reduced reality monitoring accuracy for images is largely independent of the emotion elicited by those images.

Discussion

In the second experiment, generating descriptive sentences led to significant source memory confusion beyond that arising from new captions not previously seen. However, in both Experiments 1 and 2, imagery was associated with significantly less accurate reality monitoring accuracy than sentences describing the same object, whether negative or benign. Thus, both forms of representation (imagery or verbal description) led to less accurate reality monitoring, in comparison with a baseline level for completely new items, but mental images were consistently more likely to be confused with actually seeing pictures than were verbal descriptions. The present findings thus provide support for the relatively untested assumption that mental imagery is more likely to be confused with actually having perceived an event than are alternative (e.g., verbal) forms of representation.

In both Experiments 1 and 2, negative emotional items were rated as being more unpleasant than benign items, although, as predicted, this difference was consistently greater for images than for verbal descriptions. Negative emotional items were also associated with elevated false alarm rates, but no such differences due to emotional content were found in the signal detection analysis of sensitivity, nor was there a significant correlation between ratings of emotional reaction and sensitivity. In contrast, negative emotional content *was* found to be associated with the use of a more lax response criterion, leading us to suggest that this might underlie the greater number of false alarms. The reason for this emotion-related elevation in false alarms is not clear, although it could be associated with the greater potential importance of mistaking real threats as being imaginary (“better safe than sorry”). Whatever the explanation, signal detection analysis in both Experiments 1 and 2 indicated that the higher false alarm rates associated with emotional content do not reflect reduced discrimination accuracy.

Our finding of an emotion-related elevation in false alarms differs from that of Kensinger and Schacter (2005, 2006), who reported fewer false

alarms for negative than neutral images. This difference is unlikely to be due to variations in the material used, because captions and pictures in both studies were taken from the same set. Affective ratings confirmed the expected differences between the present negative and benign sets; although, as noted earlier, the degree of associated affect depended on how the caption was processed, with images amplifying the experienced emotion. One major procedural difference was that Kensinger and Schacter (2005, 2006) used size judgements to unobtrusively elicit imagery, whereas in the present experiments participants were explicitly instructed to produce an image (or sentence), and then to rate it for pleasantness–unpleasantness.

An explanation offered by Kensinger and Schacter (2005, 2006) for their finding of emotional–neutral differences was that the better memory for emotional content might serve to enhance reality monitoring accuracy. However, it remains unclear why generally better memory for emotional content should necessarily lead to more accurate discrimination between imagery and perception, unless the characteristics differentiating images from actual percepts are also enhanced by emotion. Alternatively, it is possible that our use of affective rather than size judgements resulted in more attention being paid to the emotional content of pictures (and images) that could be used later to discriminate between mental images and percepts, and so increase reality monitoring accuracy. We therefore carried out a final experiment designed to examine whether emotion-related differences in discrimination vary according to the type of rating made.

EXPERIMENT 3

Method

Experiment 3 was designed to investigate the possibility that our assessment of reality monitoring accuracy may have been influenced by the type of encoding task used. Specifically, we wondered whether our use of affective judgements might account for why we found little difference in

reality-monitoring errors due to emotionality of mental images, whereas Kensinger and Schacter (2005) concluded from their false alarm data that emotionally negative images were distinguished from pictures more accurately than were benign images.

Experiment 3 followed a similar method to that of Experiment 2, with one set of 72 captions and corresponding pictures (36 negative and 36 benign) being seen in “look” trials, but now with half of these trials followed by an affective judgement and half by a size judgement. In the remaining 144 trials participants were asked to produce a mental image prompted by a caption, followed by affective judgements in 72 trials (36 negative and 36 benign), and size judgements in the other 72 (again 36 of each valence).

Participants and procedure. Forty-two undergraduates (10 male) took part, and received course credit for their participation. Initial instructions emphasised that participants would be asked to focus on either the size of the object pictured or imagined (as it would appear in real life), or on how pleasant or unpleasant the picture or image was. Size was judged using a 1–5 scale, from much smaller than the computer monitor to much larger than the computer monitor (approximately 33 × 38 cm) and affect using another 5-point scale (from *Very unpleasant* to *Very pleasant*). Participants practised making these ratings in six trials, and when it was clear that the instructions were understood, they continued on to 216 experimental trials.

Each trial was initiated by participants pressing the space bar, and began with a central “SIZE?” or “FEEL?” prompt according to the type of judgement required for that trial, followed by the caption and then either a corresponding picture or a dark screen indicating that a mental image should be produced. The appropriate 5-point scale was then presented and participants rated the picture or their mental image by pressing a number key. The source memory questionnaire sent and completed the following day included all the 216 captions seen previously and participants responded

with yes/no answers according to their memory of whether or not they had seen a picture after each caption.

Results

As in previous experiments, an initial exploration of effects due to emotional valence and type of rating on source memory was carried out using hit rates (frequency of “yes” responses for captions that had preceded pictures) and false alarms (“yes” responses to captions that had been followed by imagery). Hit rates were entered into a repeated-measure ANOVA having within-participant factors of Valence (negative vs. benign) and Rating (affect vs. size). This revealed main effects due to both emotional Valence, with higher hit rates for negative than benign items, $F(1, 41) = 38.16$, $p < .001$, $\eta_p^2 = .48$; and Rating, with higher hit rates for items that had been rated for affect rather than size, $F(1, 41) = 18.93$, $p < .001$, $\eta_p^2 = .32$, but no significant interaction between them. A similar repeated-measure analysis of false alarm rates for captions that had been followed by images failed to show significant effects due to valence, type of rating or their interaction.

The main test of whether reality monitoring accuracy was influenced by type of rating was again carried out using signal detection sensitivity (d') scores, computed as in previous experiments. A repeated-measure ANOVA having within-participant factors of Emotional Valence and Type of Rating revealed main effects of Valence, $F(1, 41) = 24.04$, $p < .001$, $\eta_p^2 = .37$, with higher sensitivity scores for emotionally negative items, and of rating, $F(1, 41) = 8.44$, $p < .006$, $\eta_p^2 = .17$, with higher sensitivity for items that had been rated for affect. Importantly for the present hypothesis, there was also a significant interaction between Valence and Rating, $F(1, 41) = 4.38$, $p < .05$, $\eta_p^2 = .10$ (see Table 1). The difference in accuracy due to emotionally negative versus benign content was greater after making size ratings (a difference of 0.56 in d') than for affective ratings (a difference of 0.19). Inspection of cell means in Table 1 shows that the larger difference after size ratings mainly reflects particularly poor

discrimination of benign images from pictures that had been rated for size (mean $d' = 1.40$, relative to the other three sensitivity means of 1.96, 2.03 and 1.82).

As a further test of the emotional effect on response bias found in previous experiments, response criterion scores (c) were entered into a repeated-measure ANOVA examining effects due to within-participant factors of Emotional Valence (negative vs. benign) and Rating (size vs. affect). This confirmed the effect of Valence, with a more lax response criterion for endorsing negative items, $F(1, 41) = 11.48$, $p < .002$, $\eta_p^2 = .22$. There was also a significant effect of Rating, with a more lax response criterion for items rated for affect, $F(1, 41) = 8.52$, $p < .006$, $\eta_p^2 = .17$, but there was no significant interaction between Valence and Rating. Thus both emotional content and affective encoding led to participants being more willing to endorse items as having been seen previously as pictures, irrespective of accuracy.

Results of Experiment 3 are thus consistent with the earlier findings of Kensinger and Schacter (2005), to the effect that emotionally negative images were more accurately rejected as not having been seen previously as pictures than were benign images, when both had been rated for size. At the same time the results of Experiment 3 are also broadly consistent with the findings from Experiments 1 and 2 reported here: when rated for affect there was much less difference in reality monitoring accuracy between emotionally negative and benign images. It appears that either emotionally negative image content or the use of affective ratings can enhance source monitoring accuracy, relative to benign images rated for size, with images in the last condition being particularly prone to being confused with having seen a picture. Negative emotional content may prompt incidental encoding of the type of affective perceptual detail that helps to discriminate images from percepts, whereas benign content is encoded in a similar way only when an affective rating is required. This would account for the similar accuracy levels for emotional and benign images rated for affect, as well

as the particularly poor discrimination between imagery and perception when benign items were rated for size.

GENERAL DISCUSSION

To our knowledge, the current data are the first to provide direct evidence supporting the widely held assumption that mental images are more likely to be confused with perceived events than are alternative forms of representation (e.g., verbal description). We found that generation of either verbal descriptions or mental images led to significant source monitoring errors, in which generated representations were sometimes confused with actually having seen a picture, relative to a baseline level for new captions that had not been seen previously (in Experiment 2). More critically for present purposes, in both Experiments 1 and 2, signal detection sensitivity measures indicated that images were less accurately discriminated from actually viewed pictures than were sentences, consistent with expectations based on the previously documented overlap between the neural processes involved in mental imagery and perception. The largely untested assumption that images are especially prone to being confused with having perceived an event was thus supported, at least in comparison with the main alternative form of representation—verbal description.

Affective ratings for pictures, images and sentences confirmed that the items selected here as emotionally negative were indeed experienced as being more unpleasant than were those designated as neutral (or benign). More importantly, and consistent with earlier findings (e.g., Holmes & Mathews, 2005), negative images led to higher unpleasantness ratings than did verbal descriptions. Conversely, benign images were rated as slightly pleasant on average (mean ratings were midway between “neutral” and “pleasant”) and these ratings were lower (i.e., more pleasant) than for verbal descriptions of the same object. Images thus seem to amplify emotion in either a negative

or positive direction, according to the valence of their content.

In each experiment we found consistent differences in response criterion indicating greater willingness to endorse emotionally negative items—whether images or verbal descriptions—as having been seen before as pictures. Importantly, this suggests that people tend to report remembering more emotionally negative than benign events as having occurred in reality, regardless of whether they were actually perceived or originated as images or verbal descriptions. In contrast to this emotion-related difference in response bias, in Experiments 1 and 2 we found little difference in a signal detection measure of sensitivity between emotionally negative and benign items. That is, the generally greater willingness to report having seen emotionally negative events was not accompanied by any less accurate discrimination of whether they had been imagined or perceived.

In Experiment 3 we investigated the apparent discrepancy between this finding and previous results (Kensinger & Schacter, 2005) that were interpreted as showing greater reality monitoring accuracy for emotional than for neutral images. By manipulating whether images were rated for size or affect, we found that the type of encoding significantly influenced sensitivity, with affect rating being associated with greater reality monitoring accuracy than ratings of size. Images of neutral (or benign) objects were particularly less well discriminated from actually having seen a picture when they had been rated for size rather than affect.

Why might ratings of size reduce reality monitoring accuracy in comparison to making affective ratings? We suggest that the requirement to make size ratings is likely to discourage attention to the type of perceptual detail that could help to differentiate images from actually viewed objects, by directing attention instead to global spatial attributes (such as the peripheral outline of the object or the volume of space occupied). Attention to such global size attributes is unlikely to be helpful when later trying to discriminate pictures from images (and this would be particularly true if participants seeing pictures

also imagined the external dimensions of the pictured objects to help estimate their size in real life). In contrast, we suggest that rating the affect associated with a picture or image requires attention to the critical perceptual features that serve to elicit emotion and recall of these details is likely to be helpful in distinguishing between mental images and actual percepts.

At first glance, this account does not provide an obvious explanation of why, when rated for size, the source of benign (or neutral) items was less accurately identified than was that of emotionally negative items. However, emotional content of pictures typically captures attention more readily than neutral content (e.g., Calvo & Lang, 2005; Kensinger & Schacter, 2007), so that, even when rating size, it is likely that attention was more often captured by emotionally negative perceptual details that could help in distinguishing between memories for images and actually viewed pictures. Such involuntary attentional capture effects would be much less likely to occur when rating the size of neutral or benign objects. This account provides an explanation of why both emotional content and the requirement to encode for affect improved reality monitoring accuracy. More source monitoring errors should occur when neither factor was present—as when benign objects were rated for size.

We have argued that encoding emotional content results in more accurate distinctions being made between mental images and viewed pictures. As was noted in the introduction, however, some earlier data had suggested that attending to one's own feelings can *reduce* source monitoring accuracy (Johnson et al., 1996). In the latter study, participants listened to statements made by others and rated either how they felt about the content of each statement, or rated how they thought the speaker felt. Results indicated that rating one's own feelings led to *less* accurate source monitoring (that is, who had made each statement) than did rating the speaker's feelings. This finding can be understood by noting that rating how you feel about another person's statement is likely to direct attention to the relation between the statement's meaning and one's own attitudes and beliefs, and

thus *away* from the critical perceptual features (e.g., voice characteristics) that could help in later identifying the source.

In contrast, attending to how one feels about a picture requires that attention is focused on the critical perceptual features that evoke emotion, and which may help in distinguishing between memory for an image or a viewed picture. For example, rating feelings about emotional pictures (such as a bloody wound, or a striking snake) depends on attention to the perceptual aspects that elicit emotional reactions (e.g., visual details such as torn flesh, exposed fangs, etc.). Encoding the perceptual features that give rise to emotion can thus help to distinguish between memories of imaged and perceived events and enhance reality monitoring accuracy. This contrasts with the situation when rating one's feelings about the content of verbal statements, which is likely to direct attention away from (irrelevant) perceptual features, such as the speaker's voice characteristics. If so, then, rather than concluding that emotional encoding always leads to more (or less) accurate reality monitoring, source monitoring effects will vary according to whether or not attention is directed to distinguishing information, such as emotion-provoking perceptual detail (that helps reality monitoring), or to emotional associations in semantic memory (that does not).

In conclusion, our results confirm the hypothesis that mental images of emotional events or objects typically evoke higher levels of affect than do verbal descriptions of the same event. Mental images are also more likely to be confused with actual percepts than are verbal representations, consistent with the overlap in the processes involved in imagery and perception. However, reality monitoring accuracy can be significantly influenced by the type of information that is encoded. Benign images were less accurately distinguished from viewed pictures than were negative emotional images, when instructions prompted attention to unhelpful information such as size. Importantly, however, this disadvantage was much reduced when affective encoding was encouraged, indicating that reality monitoring accuracy is enhanced by attention to distinguish-

ing information such as perceptual content linked with affect. Rather than the source of a memory for emotional events always being more (or less) accurately recognised than for neutral events, reality monitoring accuracy depends on whether or not the type of affective information encoded helps to distinguish imagery from perception.

Manuscript received 27 December 2011

Revised manuscript received 21 May 2012

Manuscript accepted 22 May 2012

First published online 10 July 2012

REFERENCES

- Baddeley, A., & Andrade, J. (2000). Working memory and the vividness of imagery. *Journal of Experimental Psychology: General*, *129*, 126–146.
- Brown, H. D., Kosslyn, S. M., Breiter, H. C., Baer, L., & Jenike, M. A. (1994). Can patients with obsessive-compulsive disorder discriminate between percepts and mental images? A signal detection analysis. *Journal of Abnormal Psychology*, *103*, 445–454.
- Cabeza, R., & St Jacques, P. (2007). Functional neuroimaging of autobiographical memory. *Trends in Cognitive Science*, *11*, 219–227.
- Calvo, M. G., & Lang, P. J. (2005). Parafoveal semantic processing of emotional visual scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 502–519.
- Dobson, M., & Markham, R. (1993). Imagery ability and source monitoring: Implications for eye witness memory. *British Journal of Psychology*, *84*, 111–118.
- Ganis, G., Thompson, W. L., & Kosslyn, S. M. (2004). Brain areas underlying visual mental imagery and visual perception: An fMRI study. *Cognitive Brain Research*, *20*, 226–241.
- Gonsalves, B., & Paller, K. A. (2000). Neural events that underlie remembering something that never happened. *Nature Neuroscience*, *3*, 1316–1321.
- Holmes, E. A., & Mathews, A. (2005). Mental imagery and emotion: A special relationship? *Emotion*, *5*, 489–497.
- Holmes, E. A., Mathews, A., Mackintosh, B., & Dalgleish, T. (2008). The causal effect of mental imagery on emotion assessed using picture-word cues. *Emotion*, *8*, 395–409.

- Hyman, I. E., & Pentland, J. (1996). The role of mental imagery in the creation of false childhood memories. *Journal of Memory and Language*, *35*, 101–117.
- Johnson, M. K. (2006). Reality and memory. *American Psychologist*, *61*, 760–771.
- Johnson, M. K., Nolde, S. F., & De Leonardis, D. M. (1996). Emotional focus and source monitoring. *Journal of Memory and Language*, *35*, 135–156.
- Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, *88*, 67–85.
- Kensinger, E. A., O'Brien, J. L., Swanberg, K., Garoff-Eaton, R. J., & Schacter, D. L. (2007). 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. (2005). Emotional content and reality-monitoring ability: fMRI evidence for the influences of encoding processes. *Neuropsychologia*, *43*, 1429–1443.
- Kensinger, E. A., & Schacter, D. L. (2006). Reality monitoring and memory distortion: Effects of negative, arousing content. *Memory & Cognition*, *34*, 251–260.
- Kensinger, E. A., & Schacter, D. L. (2007). Remembering the specific visual details of presented objects: Neuroimaging evidence for effects of emotion. *Neuropsychologia*, *45*, 2951–2962.
- Kim, S. E., Kim, J. W., Kim, J. J., Jeong, B. S., Choi, E. A., Jeong, Y., G., et al. (2007). The neural mechanism of imagining facial affective expressions. *Brain Research*, *1145*, 128–137.
- Kosslyn, S. M., & Thompson, W. L. (2003). When is early visual cortex activated during visual mental imagery? *Psychological Bulletin*, *129*, 723–746.
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide*. New York, NY: Psychology Press.
- O'Craven, K. M., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *Journal of Cognitive Neuroscience*, *12*, 1013–1023.
- Segal, S., & Fusella, V. (1969). Effects of imagining on signal-to-noise ratio, with varying signal conditions. *British Journal of Psychology*, *60*, 459–464.
- Sharot, T., Riccardi, A. M., Raio, C. M., & Phelps, E. A. (2007). Neural mechanisms mediating optimism bias. *Nature*, *450*, 102–105.
- Slotnick, S. D., Klein, S. A., Dodson, C. S., & Shimamura, A. P. (2000). An analysis of signal detection and threshold models of source memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *26*, 1499–1517.
- Zaragoza, M. S., Mitchell, K. J., Payment, K., & Drivdahl, S. (2011). False memories for suggestions: The impact of conceptual elaboration. *Journal of Memory and Language*, *64*, 18–31.