

Mistaken Memories: Remembering Events That Never Happened

BRIAN GONSALVES and KEN A. PALLER

*Department of Psychology and Institute for Neuroscience
Northwestern University
Evanston, Illinois*

Our memories can be accurate, but they are not always accurate. Eyewitness testimony, for example, is notoriously unreliable. Insights into both veridical and false remembering have come from recent investigations of memory distortion. Behavioral measures have been used to demonstrate false memory phenomena in the laboratory, and neuroimaging measures have been used to provide clues about the relevant events in the brain that support remembering versus misremembering. A central category of misremembering results from confusion between memories for perceived and imagined events, which may result from overlap between particular features of the stored information composing memories for perceived and imagined events. *NEUROSCIENTIST* 8(5):000–000, 2002. DOI: 10.1177/107385802236964

KEY WORDS *Memory, False memory, Neuroimaging, Event-related potentials, ERP's*

Each one of us has a life story, a collection of personally experienced events, that we can bring to mind on command. These episodic memories are stored in the brain such that we can recall events experienced either a few minutes earlier or many years ago, and in doing so re-experience multiple aspects of that event, including a variety of perceptual details, thoughts, observations, and emotions. The brain networks that enable this memory retrieval to happen, however, do not include exact records of those past events. Instead, episodic remembering is a reconstructive process, whereby one must make use of stored information that is often limited and incomplete. Because of this reconstructive nature of retrieval, memories are subject to distortion (Bartlett 1932; Schacter 1995; Schacter and others 1998). The study of memory distortion and its basis in the brain, as reviewed below,

can provide many insights into this process of reconstruction, thus helping to clarify the inner workings of memory in general, both when it works well and when it fails.

False memories take many forms, and false remembering can be produced in a laboratory setting using several different methods. One time-honored method for studying human memory involves requiring people to learn a list of words. When recognition is subsequently tested for those words, people occasionally make “false alarms,” indicating that they remember a word that was not, in fact, present on the original list. The number of false alarms can be substantially boosted using a method pioneered by Deese (1959) and extended by Roediger and McDermott (1995), in which the learning list includes groups of words that are each closely associated with one theme word. For example, people will be asked to remember a list of words that are all associated with the theme word *sleep*—but importantly, the theme word is never

presented (Fig. 1). People frequently claim to remember the theme word from the learning list, even though it was not presented, and their level of confidence in these false memories can be as high as for their true memories. Apparently, these false memories do not reflect a vague sense of familiarity but rather resemble the full-blown experience of accurate remembering.

In line with findings that people can mistakenly claim that a prior event occurred while remaining highly confident, results from brain imaging have demonstrated similar patterns of brain activity for true and false memories, using positron emission topography (PET), functional magnetic resonance imaging (fMRI), and event-related potentials (ERP) methods for monitoring the brain in action (Schacter and others 1996; Düzel and others 1997; Schacter and others 1997). Other evidence, however, indicates that the brain treats true and false memories differently. In two studies, veridical memories tended to be rated as including more perceptual detail than false memories (Mather and others 1997; Norman and Schacter 1997). These experiential differences suggest that brain activity associated with true and false memories should also differ. Indeed, differential activation of prefrontal cortex was found when extended periods of false remembering and of accurate remembering were compared, perhaps reflecting post-retrieval monitoring processes (Schacter and others 1996; Johnson and others 1997; Schacter and others 1997). However, no definitive connection has been established between those prefrontal differences and the experiential differences observed in behavioral studies. In contrast, greater activity in the auditory cortex was found during memory retrieval for true than false memories for heard words, which may reflect the retrieval of auditory information about words that were actually heard (Schacter and others 1996). In an fMRI study, both true and false memories were associated with hip-

Address correspondence to: Brian Gonsalves, 2029 Sheridan Rd., Evanston, IL 60208-2710 (e-mail: gonsalves@northwestern.edu).

Attempt to learn a list of words



Recognition test yielding false memories

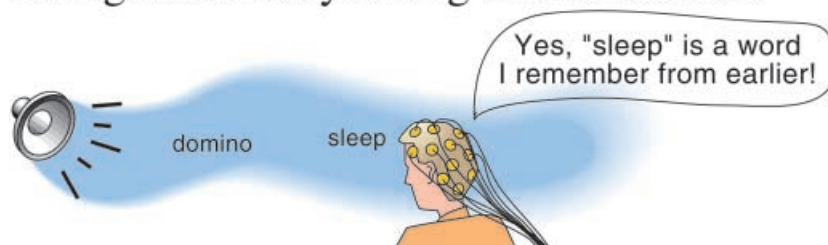


Fig. 1. An outline of the Deese-Roediger-McDermott procedure for inducing false memories, in which participants hear a series of words associated with a nonpresented theme word. Later, participants often claim to have heard the nonpresented theme word.

pocampal activity, whereas accurate retrieval elicited relatively more activity in the parahippocampal cortex, perhaps reflecting the greater reactivation or re-experiencing of perceptual details (Cabeza and others 2001). Differences in neural activity associated with true and false memory were also observed in several ERP studies (Walla and others 2000; Curran and others 2001; Miller and others 2001; Nessler and others 2001). In one particularly elegant ERP study, word learning was accomplished using lateralized visual presentations such that contralateral visual areas were preferentially engaged (Fabiani and others 2000). When recognition was subsequently tested using central presentations, true memories of presented words showed appropriately lateralized ERP activity, whereas falsely recognized theme words did not. Only words actually presented during the learning elicited brain activity indicative of a sensory memory trace.

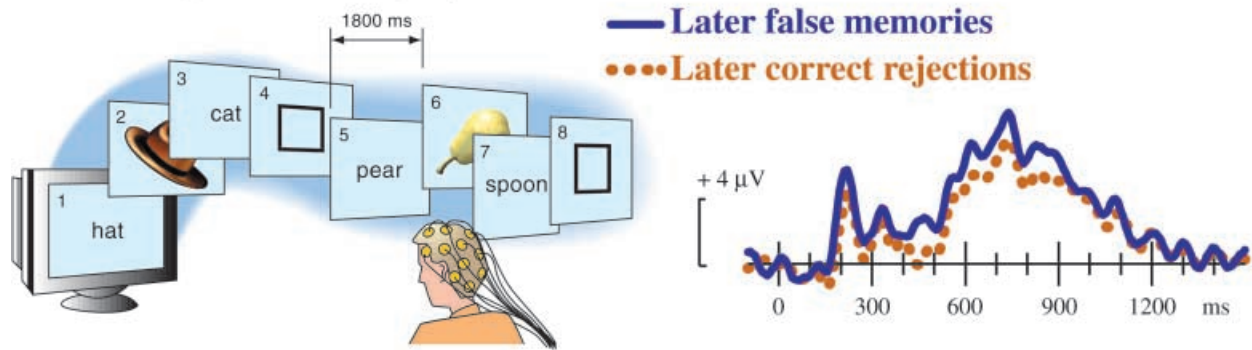
False memories do not occur only when, as in the aforementioned studies, extensive semantic associations between learned words and a theme word engender false alarms on memory tests. Another arguably more

ecological way in which false remembering can be induced is by leading people to believe that an event that they imagined had actually occurred. This is known as an error in “reality monitoring” because the individual confuses memories for imagined and perceived events. These errors may arise as a consequence of similarities between how imagined and perceived events are stored in the brain, and between the event features that are reactivated during retrieval (Johnson and Raye 1981; Johnson and others 1993). People are generally quite competent at this form of source discrimination, given that several cues can be used to distinguish between perceived and imagined events. For example, memories for perceived events tend to include more perceptual detail than do memories for imagined events (Johnson and others 1988; Mather and others 1997; Norman and Shacter 1997). Memories for imagined events, on the other hand, often include more records of the cognitive operations required for internally generating the information in question. Although subjective characteristics of mnemonic traces associated with real and imagined events tend to

differ, there are similarities. Critically, when memory records for imaginations and perceptions extensively overlap in features, the two tend to be confused, and we may then mistake imagination for reality.

We recently developed a new procedure for generating reality-monitoring errors, which allowed us to obtain neural correlates of true and false remembering (Gonsalves and Paller 2000b). As shown in Figure 2, subjects in our experiment viewed a series of object names and generated a visual image of each corresponding object. For half of the words, a picture of the object was also presented 2 seconds after the word, but the pictures required no response. Subsequently, in a surprise memory test, subjects heard a series of spoken words. Some of these words were seen earlier, some were seen with the corresponding picture, and some were not seen at all. For each word, subjects decided whether or not they saw a picture, and some were not seen at all. For each word, subjects decided whether or not they saw a picture of the named object earlier. This paradigm thus creates substantial overlap in memory attributes for imagined objects and for perceived objects, thereby eliciting frequent reality-monitoring errors. Subjects claimed to have seen pictures of 30% of the items that they had only seen as a word during encoding, whereas they only claimed to have seen 9% of the new items. Brain activity recorded during the retrieval phase of this experiment differed between true and false memories. Specifically, from 900 to 1200 ms after word onset, ERP responses at parietal and occipital scalp locations were more positive for true memories than for false memories. In interpreting this effect, we were able to build on our previous observation of an ERP correlate of visual imagery during memory retrieval (Gonsalves and Paller 2000a). Participants in this previous experiment studied sets of spoken words either with or without instructions to generate visual images, and they later recognized those words encoded with visual imagery better. ERPs from occipital scalp locations differed for these two types of words,

Generating visual imagery for each word



Memory test for the pictures



Fig. 2. Procedure used by Gonsalves and Paller (2000b) for generating false memories, as outlined in the text. Brain potentials observed in the two parts of this experiment are shown on the right. (From Paller 2002, adapted with permission requested; Figure 1 and the *left side* of this figure from de Schipper S. 2000, December 2. *Sterke verbeelding: valse herinnering blijkt uithersengolven* [Strong imagination: false memory is evident from brainwaves]. *NRC Handelsblad*, p. 47. Copyright 2000 by *NRC Handelsblad*. Adapted by permission.)

and these differences were interpreted as reflecting the enhanced retrieval of visual object representations for words encoded using visual imagery. Similarly, we interpreted the ERP difference between true and false memories as reflecting the enhanced retrieval of visual perceptual information about object pictures for true compared to false memories. This interpretation accords well with the reality-monitoring framework, which predicts that memories for perceptions should include, on average, more perceptual detail than memories for imaginations (Johnson and Raye 1981; Johnson and others 1988; Johnson and others 1993).

In addition to these differences between true and false memories, ERPs recorded during the study phase were computed according to whether or not items were later recognized. Neural differences between later remembered and later forgotten items, so called subsequent memory effects, have been observed using both ERP and event-related fMRI

methods (for review, see Paller and Wagner, in press). Typically, study-phase items that were subsequently remembered elicited a more positive ERP than did items later forgotten. In our experiment, we sorted study-phase ERP responses to pictures according to whether they were later remembered or forgotten, and we observed a subsequent memory ERP effect widely distributed across the scalp and likely reflecting the activity of multiple brain regions. We also recorded ERP responses to study words not followed by a picture and sorted them according to whether the person later falsely claimed to have seen the picture. ERPs associated with later false remembering were more positive than those associated with later correct rejections, most reliably from 600 to 900 ms at occipital and parietal scalp locations. Given our prediction that vivid visual imagery can promote subsequent reality-monitoring errors, such that an item presented only as a word could later be falsely remembered as a picture, we interpreted these effects

as a reflection of strong visual imagery that promoted later false remembering. Preliminary results from an event-related-fMRI study of false memory using this same design suggest that imagery-associated activity in the precuneus, anterior cingulate cortex, and inferior occipital gyrus likewise promotes later false remembering (Gonsalves and others 2001). Precuneus activation has been associated previously with memory-related imagery (Fletcher and others 1995, 1996), and precuneus, anterior cingulate, and inferior occipital gyrus are areas engaged during both visual imagery and visual perception (Kosslyn and Thompson 2000). These results are in line with our earlier interpretations that more elaborate or vivid visual imagery during the imagery task precipitates a later false memory in this paradigm.

We interpreted these neurophysiological harbingers of false memories as reflections of the association between vivid visual imagery and later reality-monitoring errors, an

account consonant with prior work on the effects of imagery. For example, when people actively imagine a concocted childhood event, the likelihood that subjects later believe the event had actually happened increases (Hyman and Pentland 1996). Repeatedly imagining an event increases subjective ratings of the likelihood that the event occurred, a phenomenon called *imagination inflation* (Garry and others 1996). Encoding words with an emphasis on imagery increases the number of errors when people later attempt to remember whether particular items were seen as words or pictures (Durso and Johnson 1980; Lane and Zaragoza 1995). The constructive memory framework of Schacter and colleagues (Schacter and others 1998) suggests a general neural mechanism for these reality-monitoring errors. This framework takes as a starting point the idea that event memory traces consist of distributed networks localized in cortical regions involved in initially encoding an event and that these networks are reactivated when the memory is retrieved (Squire 1987; McClelland and others 1995; Paller 2002). Two highly similar events will have a large amount of overlap in their distributed features, such that events may be difficult to distinguish from one another during retrieval. This general mechanism can be applied to the occurrence of reality-monitoring errors, and in particular the confusion between visual perceptions and visual imaginations. It is well established that visual imagery activates many of the same brain areas active during visual perception (Kosslyn and Thompson 2000), which suggests that images and perceptions can have overlapping memory representations in the cerebral cortex. Particularly vivid visual imagery of an event might lead to the formation of a memory trace that is indistinguishable from the trace that would have been formed had the event actually been perceived. These general mechanisms might be applied to false memory phenomena such as the effects of misleading questions on eyewitness testimony, whereby post-event suggestions about what

happened during an event lead to distortion in people's account of events (Loftus and others 1995). This reality-monitoring mechanism may also provide a model for the creation of false childhood memories through suggestive questioning (i.e., Hyman and Pentland 1996). In this case, when people actively imagine a fabricated childhood event, they sometimes come to believe that the event actually happened. Continuing research on memory distortion in the laboratory, using the methods of cognitive neuroscience, can be expected to shed additional light on the brain mechanisms whereby imagined events may later become confused with actual events, leading to false memories outside the laboratory.

References

Bartlett FC. 1932. Remembering. Cambridge, UK: Cambridge University Press.

Cabeza R, Rao SM, Wagner AD, Mayer AR, Schacter DL. 2001. Can medial temporal lobe regions distinguish true from false? An event-related fMRI study of veridical and illusory recognition memory. *Proc Natl Acad Sci U S A* 98:4805–10.

Curran T, Schacter DL, Johnson MK, Spinks R. 2001. Brain potentials reflect behavioral differences in true and false recognition. *J Cogn Neurosci* 13:201–16.

Deese J. 1959. On the prediction of occurrence of particular verbal intrusions in immediate recall. *J Exp Psychol* 58:17–22.

Durso FT, Johnson MK. 1980. The effects of orienting tasks on recognition, recall, and modality confusion of pictures and words. *J Verbal Learn Verbal Behav* 19:416–29.

Düzel E, Yonelinas AP, Mangun GR, Heinze HJ, Tulving E. 1997. Event-related brain potential correlates of two states of conscious awareness in memory. *Proc Natl Acad Sci U S A* 94:5973–78.

Fabiani M, Stadler MA, Wessels PM. 2000. True but not false memories produce a sensory signature in human lateralized brain potentials. *J Cogn Neurosci* 12:941–9.

Fletcher PC, Frith CD, Baker SC, Shallice T, Frackowiak RSJ, Dolan RJ. 1995. The mind's eye—precuneus activation in memory-related imagery. *Neuroimage* 2:195–200.

Fletcher PC, Shallice T, Frith CD, Frackowiak RSJ, Dolan RJ. 1996. Brain activity during memory retrieval: the influence of imagery and semantic cueing. *Brain* 119:1587–96.

Garry M, Manning CG, Loftus EF, Sherman SJ. 1996. Imagination inflation: imagining a childhood event inflates confidence that it occurred. *Psychon Bull Rev* 3:208–14.

Gonsalves B, Paller KA. 2000a. Brain potentials associated with recollective processing of spoken words. *Mem Cogn* 28:321–30.

Gonsalves B, Paller KA. 2000b. Neural events that underlie remembering something that never happened. *Nat Neurosci* 3:1316–21.

Gonsalves B, Reber PJ, Gitelman DR, Parrish TB, Mesulam M-M, Paller KA. 2001. Event-related fMRI reveals brain activity at encoding that predicts true and false memories for visual objects. *Soc Neurosci Abstr* 27: Program No. 236.9.

Hyman IE, Pentland J. 1996. The role of mental imagery in the creation of childhood memories. *J Mem Language* 35:101–17.

Johnson MK, Hashtroudi S, Lindsay SD. 1993. Source monitoring. *Psychol Bull* 114:3–28.

Johnson MK, Nolde SF, Mather M, Kounios J, Schacter DL. 1997. The similarity of brain activity associated with true and false recognition memory depends on test format. *Psychol Sci* 8:250–57.

Johnson MK, Raye CL. 1981. Reality monitoring. *Psychol Rev* 88:67–85.

Johnson MK, Suengas AG, Foley MA, Raye CL. 1988. Phenomenal characteristics of memories for perceived and imagined autobiographical events. *J Exp Psychol Gen* 117:371–6.

Kosslyn SM, Thompson WL. 2000. Shared mechanisms in visual imagery and visual perception: insights from cognitive neuroscience. In: Gazzaniga MS, editor. *The new cognitive sciences*. Cambridge (MA): MIT Press. p 975–85.

Lane SM, Zaragoza MS. 1995. The recollective experience of cross-modality confusion errors. *Mem Cogn* 23:607–10.

Loftus EF, Feldman J, Dashiell R. 1995. The reality of illusory memories. In: Schacter DL, editor. *Memory distortion*. Cambridge (MA): Harvard University Press. p 47–68.

Mather M, Henkel LA, Johnson MK. 1997. Evaluating characteristics of false memories: remember/know judgements and memory characteristics questionnaire compared. *Mem Cogn* 25:826–37.

McClelland JL, McNaughton BL, O'Reilly RC. 1995. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychol Rev* 102:419–57.

Miller AR, Baratta C, Wynveen C, Rosenfeld JP. 2001. P300 latency, but not amplitude or topography, distinguishes between true and false recognition. *J Exp Psychol Learn Mem Cogn* 27:354–61.

Nessler D, Mecklinger A, Penney TB. 2001. Event related brain potentials and illusory memories: the effects of differential encoding. *Cogn Brain Res* 10:283–301.

Norman KA, Schacter DL. 1997. False recognition in young and older adults: exploring the characteristics of illusory memories. *Mem Cogn* 25:838–48.

Paller KA. 2002. Cross-cortical consolidation as the core defect in amnesia: prospects for hypothesis-testing with neuropsychology and neuroimaging. In: Squire LR, Schacter DL, editors. *The neuropsychology of memory*. 3rd ed. New York: Guilford, in press.

Paller KA, Wagner Ad. In press. Monitoring the transformation of experience into memory. *Trends Cogn Sci*.

- Roediger HL, McDermott KB. 1995. Creating false memories: remembering words not presented in lists. *J Exp Psychol Learn Mem Cogn* 21:803–14.
- Schacter DL, editor. 1995. *Memory distortion*. Cambridge (MA): Harvard University Press.
- Schacter DL, Buckner RL, Koutstaal W, Dale AM, Rosen BR. 1997. Late onset of anterior prefrontal activity during true and false recognition: an event-related fMRI study. *Neuroimage* 6:259–69.
- Schacter DL, Norman KA, Koutstaal W. 1998. The cognitive neuroscience of constructive memory. *Annu Rev Psychol* 49:289–318.
- Schacter DL, Reiman E, Curran T, Yun LS, Bandy D, McDermott KB, and others. 1996. Neuroanatomical correlates of veridical and illusory recognition memory: evidence from positron emission tomography. *Neuron* 17:267–74.
- Squire LR. 1987. *Memory and brain*. New York: Oxford University Press.
- Walla P, Endl W, Lindinger G, Deecke L, Lang WF. 2000. False recognition in a verbal memory task: an event-related potential study. *Cogn Brain Res* 9:41–4.