THE ROYAL SOCIETY

doi 10.1098/rstb.2001.0938

Misattribution, false recognition and the sins of memory

Daniel L. Schacter* **and Chad S. Dodson**

Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, MA, USA

Memory is sometimes a troublemaker. Schacter has classified memory's transgressions into seven fundamental 'sins': transience, absent-mindedness, blocking, misattribution, suggestibility, bias and persistence. This paper focuses on one memory sin, misattribution, that is implicated in false or illusory recognition of episodes that never occurred. We present data from cognitive, neuropsychological and neuroimaging studies that illuminate aspects of misattribution and false recognition. We first discuss cognitive research examining possible mechanisms of misattribution associated with false recognition. We also consider ways in which false recognition can be reduced or avoided, focusing in particular on the role of distinctive information. We next turn to neuropsychological research concerning patients with amnesia and Alzheimer's disease that reveals conditions under which such patients are less susceptible to false recognition than are healthy controls, thus providing clues about the brain mechanisms that drive false recognition. We then consider neuroimaging studies concerned with the neural correlates of true and false recognition, examining when the two forms of recognition can and cannot be distinguished on the basis of brain activity. Finally, we argue that even though misattribution and other memory sins are annoying and even dangerous, they can also be viewed as by-products of adaptive features of memory.

Keywords: misattribution; false recognition; neuroimaging; functional magnetic resonance imaging; amnesia; forgetting

1. INTRODUCTION

Memory plays an important role in numerous aspects of everyday life, allowing us to recollect past experiences and learn new facts, navigate our environments, and remember what we need to do in the future. But memory can also fail us and sometimes fool us: we may forget or distort past experiences, and even claim to remember events that never occurred. These imperfections can provide insights into how memory works, and are therefore studied intensively by many researchers. But they can also have serious consequences for many people in our society. Forgetfulness that turns into Alzheimer's disease (AD), for instance, has profound effects on patients and their families (Pierce 2000). During the 1990s, issues concerning the accuracy of memories of childhood trauma recovered during psychotherapy led to a bitter controversy that divided professionals and destroyed families (e.g. Conway 1997; Lindsay & Read 1994; Loftus 1993; Pendergrast 1995; Schacter 1996). Similar concerns arose in cases of preschool children who, as a result of suggestive questioning, 'remembered' abusive events at the hands of teachers and others—despite an absence of objective evidence that such events ever occurred (Bjorklund 2000; Ceci & Bruck 1995; Ceci & Friedman 2000). And cases in which DNA testing has documented wrongful imprisonment almost always involve some type of faulty eyewitness memory (e.g. Wells *et al.* 1998).

*Author for correspondence (dls@wjh.harvard.edu).

Schacter (1999, 2001) recently classified the misdeeds of memory into seven basic 'sins': transience, absent-mindedness, blocking, misattribution, suggestibility, bias and persistence. The first three sins involve types of forgetting. Transience involves decreasing accessibility of information over time; absent-mindedness entails inattentive or shallow processing that contributes to weak memories of ongoing events or forgetting to do things in the future; and blocking refers to the temporary inaccessibility of information that is stored in memory. The next three sins all involve distortion or inaccuracy. Misattribution involves attributing a recollection or idea to the wrong source; suggestibility refers to memories that are implanted at the time of retrieval; and bias involves retrospective distortions and unconscious influences that are related to current knowledge and beliefs. The seventh and final sin, persistence, refers to intrusive memories that we cannot forget, even though we wish that we could.

The seven sins of memory resemble in some respects the ancient seven deadly sins (pride, anger, envy, greed, gluttony, lust and sloth). Both types of sins occur often in our daily lives. Moreover, we can view the ancient sins as exaggerations of adaptive human traits that may be necessary for survival. Similarly, memory's sins can be conceptualized as by-products of adaptive features of memory, rather than as flaws in system design or errors made by Mother Nature during evolution (Schacter 1999, 2001; see also Anderson & Schooler 1991; Bjork & Bjork 1988).

Given the focus of this issue on episodic memory, it is perhaps worth noting that all of the sins described by

Schacter (1999, 2001) involve episodic memory, although they may involve other forms of memory as well. For instance, a prominent example of the sin of blocking is the 'tip-of-the tongue' state, where people cannot produce a word or name even though they feel that they are on the verge of recovering it (for reviews, see Brown 1991; Schwartz 1999). Although tip-of-the-tongue states primarily involve semantic memory, blocking can also occur in episodic memory (Roediger & Neely 1982; Schacter 1999, 2001).

In this paper we consider evidence and ideas pertaining to one of the major sins of episodic memory: the sin of misattribution. Misattribution of memories to the wrong source is a common form of memory distortion, both in the laboratory (Dodson & Schacter 2001*a*; Jacoby *et al.* 1989; Johnson *et al.* 1993; Schacter *et al.* 1998*a*), and in everyday life. For instance, after the tragic 1995 bombing of an office building in Oklahoma City, law enforcement officers quickly apprehended a suspect called 'John Doe 1', Timothy McVeigh, who was eventually convicted of the crime. But the officers also conducted a failed search for a second suspect called 'John Doe 2' who, they believed, had accompanied McVeigh when he rented a van two days before the bombing. An artist's sketch of John Doe 2 depicted a young square-faced man with dark hair and a stocky build wearing a blue-and-white cap. A witness who had seen McVeigh rent his van also recalled seeing John Doe 2 with him. But it was later discovered that the witness had actually seen a man who fit the description of John Doe 2 at the body shop the day after he saw McVeigh there. The witness misattributed his memory of John Doe 2 to the wrong episode, leading to needless confusion and wasted effort (for more details, see Schacter 2001).

We will discuss cognitive, neuropsychological and neuroimaging evidence that has bearing on a form of misattribution known as false recognition, where individuals mistakenly claim that a novel item or episode is familiar (e.g. Underwood 1965). During the past several years, false recognition has been studied extensively, and we are beginning to understand some of its cognitive and neural underpinnings. After summarizing relevant evidence from some of our own and others' research, we will argue that misattribution can be viewed as a by-product of otherwise adaptive memory processes.

2. FALSE RECOGNITION: COGNITIVE PERSPECTIVES

Cognitive psychologists have studied false recognition for decades, using a variety of experimental procedures and paradigms (for reviews, see Alba & Hasher 1983; Johnson *et al.* 1993; Roediger 1996; Schacter 1995). During the past few years, however, interest in the phenomenon has intensified, in large part because of demonstrations of conditions under which extremely high levels of false recognition can be obtained (Hintzman 1988; Roediger & McDermott 1995; Shiffrin *et al.* 1995). In particular, Roediger & McDermott (1995) demonstrated exceptionally high levels of false recognition in experiments using a modified version of a paradigm initially developed by Deese (1959). In what has come to be known as the Deese–Roediger–McDermott (DRM) paradigm, participants study lists of words (e.g. tired, bed, awake, rest, dream, night, blanket, doze, slumber, snore, pillow, peace, yawn and drowsy) that are related to a non-presented lure word (e.g. sleep). On a subsequent old–new recognition test containing studied words (e.g. tired, dream), new words that are unrelated to the study list items (e.g. butter) and new words that are related to the study list items (e.g. sleep), participants frequently claim that they previously studied the related lure words. In many instances, false recognition of the related lure words is indistinguishable from the true recognition rate of studied words (e.g. Dodson & Schacter 2000; Mather *et al.* 1997; Norman & Schacter 1997; Payne *et al.* 1996; Roediger & McDermott 1995; Schacter *et al.* 1996*a*; cf. Miller & Wolford 1999; Roediger & McDermott 1999).

Although the mechanisms underlying this robust false recognition effect are not yet fully understood, two main sources of the effect have been considered. First, false recognition could be the result of 'implicit associative responses' (Underwood 1965) that occur when participants are exposed to lists of semantic associates during the study phase of the experiment: studying associated words (e.g. bed, tired, doze, and so forth) might lead participants to generate on their own the non-presented lure word (i.e. sleep). On a later memory test, participants may experience source confusion (Johnson *et al.* 1993): they may mistakenly recollect that they heard or saw the related lure word on the study list when in fact they had generated it on their own (Roediger & McDermott 1995; Roediger *et al.* 2001). A second possibility is that studying numerous related words results in high levels of semantic overlap among the corresponding memory representations (Schacter *et al.* 1998*a*). This failure to keep representations of each item separate will produce robust memory for semantic similarities among the items, together with poor memory for the unique or distinctive aspects of each item. Because subjects have difficulty recollecting the distinctive characteristics of the specific studied items, they tend to make recognition responses on the basis of overall similarity of the lure item to the studied items—that is, participants will respond on the basis of semantic gist (Brainerd & Reyna 1998; Payne *et al.* 1996; Schacter *et al.* 1996*a*) rather than on the basis of specific recollections.

It seems likely that both implicit associative responses and memory for semantic gist play some role in the robust false recognition that occurs in the DRM paradigm (e.g. Roediger *et al.* 2001). Evidence that false recognition can be driven mainly by gist-based processes, rather than implicit associative responses, comes from a categorized pictures paradigm developed by Koutstaal & Schacter (1997). In their paradigm, younger and older adults studied large numbers of pictures from a variety of different categories (e.g. cars, shoes, and so forth). On a subsequent recognition test, participants were shown previously studied pictures, related new pictures from previously studied categories, and unrelated new pictures. Koutstaal & Schacter reasoned that it is highly unlikely that participants would generate the related new pictures during the study phase of the experiment, in the same way that they might generate the word 'sleep' as an implicit associative response when studying related words in the DRM procedure. Koutstaal & Schacter found that participants often falsely recognized related new pictures from the same categories as the studied pictures. Rates of false recognition for young adults were considerably lower than in the DRM procedure, whereas older adults showed high levels of false recognition that, in some conditions, approached the levels observed in the DRM procedure. Overall, these data indicate that similarity or gist-based false recognition can occur even when implicit associative responses are an unlikely source of misattribution.

Whether driven by implicit associative responses or gist-based similarity, false recognition of related lures reflects some form of memory for semantic or associative information acquired during list presentation that is misattributed to a specific past sensory encounter with the item.

One implication of this line of reasoning is that false recognition can be reduced by encouraging subjects to focus on distinctive characteristics of individual items, rather than on semantic or associative properties that are common to all items. Israel & Schacter (1997) and Schacter *et al.* (1999) tested this idea by examining whether false recognition of semantic associates is reduced when subjects encode DRM lists along with detailed pictures corresponding to each item. They reasoned that presenting study-list items as pictures should increase encoding of distinctive, specific details of each item (Hunt & McDaniel 1993), which in turn should make them easier to discriminate from non-studied semantic associates. The experiments involved two key study conditions. In the word-encoding condition, participants both heard and saw lists of DRM semantic associates, whereas in the picture-encoding condition, participants heard the same lists of semantic associates and also saw line-drawings corresponding to each word. After the word- or picture-encoding conditions, previously studied words, related lures and unrelated lures were tested on an old–new recognition task (test items presented as pictures or as words). Consistent with the reasoning outlined earlier, false recognition of related lures was significantly reduced after pictorial encoding compared with word encoding. Dodson & Schacter (2000) reported a similar pattern of results using a different manipulation of distinctive encoding (saying words aloud compared with hearing them).

We (Dodson & Schacter 2000; Israel & Schacter 1997; Schacter *et al.* 1999, experiment 1) have argued that reduced false recognition after encoding distinctive information depends on a shift in responding based on participants' metamemorial assessments of the kinds of information they believe they should remember. For instance, having encountered pictures with each of the presented words, participants in the picture-encoding condition may have used a rule of thumb where they demanded access to detailed pictorial information in order to support a positive recognition decision (cf. Rotello 1999; Strack & Bless 1994). Thus, reduced false recognition appears to result from a general expectation that a test item would elicit a vivid perceptual recollection if it had in fact been presented previously. Participants in standard word-encoding conditions, by contrast, did not expect to retrieve distinctive representations of previously studied items. By our account, they were much less likely to demand access to detailed recollections. Schacter *et al.* (1999) referred to the hypothesized 'rule of thumb' in the picture-encoding condition as a distinctiveness heuristic.

Related evidence indicates that the distinctiveness heuristic has some generality across subject populations and experimental paradigms. For instance, Schacter *et al.* (1999) examined false recognition in the DRM paradigm after word and picture encoding in a group of elderly adults. Replicating previous studies (Norman & Schacter 1997; Tun *et al.* 1998), in the word-encoding condition older adults showed relatively higher levels of false recognition than younger adults. But in the picture-encoding condition, older adults reduced their levels of false recognition just as much as the younger adults did, indicating an intact ability to use the distinctiveness heuristic.

We have also found that participants can use the distinctiveness heuristic to reduce false recognition in paradigms other than the DRM procedure. Dodson & Schacter (2001*b*,*c*) used a repetition lag procedure adapted from Jennings & Jacoby (1997). In this paradigm, subjects study a list of unrelated words, and later make old–new recognition judgements about previously studied words and new words. The critical manipulation is that some new words on the recognition tests are repeated after varying lags. Even though participants are specifically instructed to say 'old' only to words from the study list, and not to new words that are repeated, after sufficiently long lags participants make false alarms to some repeated new words. Participants misattribute their familiarity with the repeated new words to a prior appearance in the study list. Jennings & Jacoby (1997) demonstrated that the effect is especially pronounced in older adults. In recent experiments by Dodson & Schacter (2001*b*), younger and older adults studied words, pictures, or a mixture of both, and then completed a recognition test in which studied items appeared once and new items appeared twice. Several experiments demonstrated that, compared with the word-encoding condition, studying pictures produced a significant reduction in false recognition rates to repeated new words for both older and younger adults. The reduction in false recognition was particularly dramatic for older adults, who showed high false recognition rates in the word-encoding condition, replicating the earlier results of Jennings & Jacoby (1997). Dodson & Schacter (2001*b*) attributed the reductions in false recognition to the use of a distinctiveness heuristic.

3. NEUROPSYCHOLOGY OF FALSE RECOGNITION: EVIDENCE FROM AMNESIA AND ALZHEIMER'S DISEASE

Cognitive studies have provided important information about the properties of misattribution and false recognition, but they are mute about relevant underlying brain systems. Studies of brain-damaged patients have begun to provide some insights into the brain regions that are implicated in false recognition. For example, a number of studies have revealed that patients with damage to various sectors of the frontal lobes show increased levels of false recognition compared with age-matched control subjects (Curran *et al.* 1997; Delbecq-Derouesné *et al.* 1990; Parkin *et al.* 1996, 1999; Ward *et al.* 1999; Rapcsak *et al.* 1999; Schacter *et al.* 1996*b*). Explanations for this increased false recognition have varied across patients, depending on particular features of their performance. Some investigators have focused on inadequate encoding of item-specific details (e.g. Parkin *et al.* 1999), whereas others have emphasized defective retrieval monitoring processes (e.g. Delbecq-Derouesné *et al.* 1990; Rapcsak *et al.* 1999; Schacter *et al.* 1996*b*, 1998*a*; for more detailed discussion, see Dodson & Schacter 2001*a*).

Studies of false recognition in amnesic patients with memory disorders resulting from damage to the medial temporal lobes (MTLs) and related structures in the diencephalon have provided important additional insights. Such patients have great difficulty remembering recent experiences and acquiring new information, despite relatively intact perception, language and general intellectual function (e.g. Parkin 2001; Squire 1992).

Schacter *et al.* (1996*a*) used the DRM paradigm to investigate false recognition in amnesic patients. Amnesiacs and matched control subjects studied lists of semantic associates, and then made old–new recognition judgements about studied words, new related words (e.g. sleep after studying tired, bed, awake, rest, and so forth), and new unrelated words (e.g. point). They found that, as expected, amnesiacs recognized fewer studied items than did the matched controls, and also made more false alarms to unrelated new words. More importantly, amnesiacs showed reduced false recognition of the related lure words (for replication and extension, see Melo *et al.* 1999). Schacter *et al.* (1997*a*) demonstrated that amnesiacs' reduced false recognition of related lure items extends to perceptual materials. After studying perceptually related words (e.g. fade, fame, face, fake, mate, hate, late, date and rate), amnesiacs were less likely than controls both to correctly recognize studied words and to falsely recognize perceptually related lure words (e.g. fate). The results thus indicate that in amnesic patients, the same processes that support accurate recognition of studied words also contribute to the false recognition of semantically or perceptually related lures.

The same general pattern of results has been observed in experiments involving pictorial materials. Using Koutstaal & Schacter's (1997) categorized pictures paradigm described earlier, Koutstaal *et al.* (2001) found that amnesic patients were less likely than controls to falsely recognize new pictures after studying numerous similar items from the same category. Koutstaal *et al.* (1999) reported a similar pattern in an experiment that examined true and false recognition of pictures of abstract objects. Each picture belonged to a particular perceptual category and was similar to a prototype that defined the category. The categories included either one, three, six or nine pictures. All participants then completed a recognition test containing studied pictures and new pictures that were either related or unrelated to studied items. They (Koutstaal *et al.* 1999) found that, for control subjects, both true and false recognition of the abstract pictures increased with category size, whereas category size had only a slight effect on true and false recognition in amnesic patients. With larger categories, amnesic patients showed less true and false recognition than did control subjects.

Recent work by Budson *et al.* (2000) has revealed the same general pattern of results in studies of patients with memory disorders resulting from AD. AD patients showed reduced false recognition of semantic associates in the DRM word recognition paradigm. Further, Budson *et al.* (2001) have found that AD patients show reduced false recognition of pictures of abstract objects from large categories in the paradigm developed by Koutstaal *et al.* (1999).

Interestingly, studies of both amnesic patients and AD patients have identified conditions within the DRM paradigm in which patients can show as much or more false recognition of related lures compared with controls. Schacter *et al.* (1998*b*) presented amnesiacs and controls with DRM lists of semantic associates, and then gave a standard old–new recognition test. This study–test procedure was repeated five times. On the first trial, amnesiacs showed reduced true and false recognition of semantic associates compared with controls, replicating the previous results discussed (Schacter *et al.* 1996*a*, 1997*a*). Across the five study–test trials, control subjects showed increased true recognition of studied words together with reduced false recognition of semantic associates. As they remembered more detailed information about specific words that actually appeared on the study list, control subjects were able to use this information to reduce or suppress their tendency to make false recognition responses based on semantic similarity or gist information. Amnesic patients, in contrast, showed no reduction in false recognition across study–test trials. In fact, they showed an opposite tendency for increased false recognition across study trials. This latter result was mainly attributable to patients with amnesia that resulted from Korsakoff's disease, who are characterized by damage to diencephalic structures and often show signs of frontal lobe damage. Patients with amnesia that resulted from MTL damage tended to show a flat or fluctuating pattern of false recognition across trials. Budson *et al.* (2000) used the same paradigm, and found that AD patients who exhibited reduced false recognition of semantic associates on the first trial, showed steadily increasing levels of false recognition across trials. By the fifth and final trial, AD patients showed higher levels of false recognition than did controls.

Both Schacter *et al.* (1998*b*) and Budson *et al.* (2000) argued that with repeated study and testing of semantic associates, patients strengthened their initially degraded representation of the semantic features or gist of the studied items, leading them to false alarm more frequently to related lure words. Unlike control subjects, however, amnesic and AD patients do not develop increasingly detailed representations of the specific words they studied. Whereas controls can use such specific representations to counter or oppose the increasing influence of semantic gist, amnesic and AD patients do not. In this paradigm, then, misattribution occurs when participants rely on representations of semantic gist that are unchecked by specific episodic memories of the actual list words. For control subjects, this pattern of influences is maximal on the first trial and reduced by the final trial. For amnesic and AD patients, it takes several trials to build up the semantic gist information that is responsible for the misattribution error.

4. NEUROIMAGING OF TRUE AND FALSE RECOGNITION

Several neuroimaging studies using positron emission tomography (PET), functional magnetic resonance imaging (fMRI) and event-related potentials (ERPs) have begun to explore misattributions associated with false recognition. Two studies from our laboratory made use of the DRM false recognition paradigm.

Schacter *et al.* (1996*c*) investigated true and false recognition with PET, and Schacter *et al.* (1997*b*) carried out a similar study with fMRI. In both studies, participants heard lists of semantic associates prior to entering the scanner; scanning was performed later as participants made old–new judgements about previously studied words, critical lures that were semantically related to the studied items, and unrelated lure words. The main finding from the two studies is that patterns of brain activity were highly similar during true and false recognition. Differences in brain activity during true and false recognition were relatively small and depended on specific characteristics of recognition testing procedures (for discussion, see Schacter *et al.* 1997*b*; also see Johnson *et al.* 1997).

Frontal lobe activation was quite prominent in each of the PET and fMRI studies of false recognition. Consistent with other imaging evidence concerning episodic memory (e.g. Henson *et al.* 1999), both studies reported evidence that regions within the frontal lobes may be involved in strategic monitoring processes that are invoked as participants struggle to determine whether a related lure word was actually presented earlier in a study list (for further relevant evidence, see Johnson *et al.* 1997). These findings fit well with the neuropsychological data discussed earlier (§ 3) that damage to areas within the frontal lobe can be associated with heightened false recognition.

Consistent with the studies discussed earlier (§ 3), showing reduced false recognition of semantic associates in amnesic patients, both the PET and fMRI studies revealed some evidence of MTL activity during false and true recognition. However, there were no detectable differences in MTL activity during true and false recognition.

In a more recent fMRI experiment, Cabeza *et al.* (2001) attempted to create conditions that would increase the likelihood of finding differences in brain activity during true and false recognition. Recall the behavioural evidence discussed earlier (§ 2) that differences between true and false recognition can be increased when study lists are presented in a perceptually distinct manner (e.g. presenting pictures with words in the DRM paradigm; Israel & Schacter (1997) and Schacter *et al.* (1999)). In the previous neuroimaging studies that used the DRM procedure, words were presented auditorily to participants. In an attempt to increase the perceptual encoding of study list words, Cabeza *et al.* (2001) had a male and female speaker present lists of semantically related words. Participants were specifically instructed to try to remember which speaker said each word. Thus, studied words should be associated with a specific source or perceptual input in a way that related lure words would not.

The most relevant imaging findings from the Cabeza *et al.* (2001) experiment concern the MTL. They found that a region of posterior MTL, in the parahippocampal gyrus, showed increased activity (relative to an unrelated lure baseline) for previously studied words but not for semantically related lure words. By contrast, a more anterior MTL region within the hippocampus showed significantly increased activity for both studied words and related lures compared with unrelated lures. The posterior MTL region may have been sensitive to perceptual differences between studied words and non-studied semantic associates. Consistent with this suggestion, an earlier study by Schacter *et al.* (1997*c*) revealed that this same posterior MTL region responded more strongly to previously studied visual objects that were tested in the same perceptual format than in a different perceptual format (i.e. when the size or orientation of the object was altered between study and test). In contrast, the anterior MTL region appeared to be responding on the basis of semantic information common to both studied words and related lures. Cabeza *et al.* (2001) thus suggested that anterior MTL is involved primarily in the recovery of semantic information, whereas posterior MTL is involved primarily in the recovery of sensory information (for further evidence of different brain responses during true and false recognition, see Fabiani *et al.* 2000).

Studies concerned with the neural correlates of misattribution have also examined false recognition errors within the context of source confusions (e.g. Johnson *et al.* 1993, 1996; Nolde *et al.* 1998). As discussed earlier (§ 2), one explanation for the robust false recognition effect in the DRM paradigm involves source confusion: when individuals fail to distinguish between what they generate as an implicit associative response (e.g. sleep) and what was actually presented during the study phase (e.g. bed, tired, dream, etc., cf. Mather *et al.* 1997; Roediger *et al.* 2001). This source confusion error, also known as an error in reality monitoring, can also refer to instances in which an individual confuses test items that were earlier imagined with items that were earlier perceived (Johnson & Raye 1981). Recently, Gonsalves & Paller (2000) used ERPs to investigate brain potentials at encoding and at retrieval that are associated with reality monitoring errors (see also Johnson *et al.* 1996; Nolde *et al.* 1998). During the study phase of their experiment, participants saw words and imagined the corresponding object, such as seeing the word 'cat' and then imagining a picture of a cat. However, half of the words were followed by a corresponding picture (the 'word plus picture' trials), whereas no picture was presented for the remaining study words (the 'word only' trials). On a later memory test, participants indicated whether or not each test item had been seen earlier as a picture.

As expected on the basis of previous work (e.g. Durso & Johnson 1980), participants made reality monitoring errors and claimed that some items in the 'word only' study condition had been presented with a picture. Because ERPs were recorded during the study phase, it was possible to examine whether there were different patterns of brain activity at encoding that were later associated with true and false recognition, as has been done previously in ERP and fMRI studies of true recognition (for review, see Wagner *et al.* 1999). In other words, are there differences in brain potentials when participants initially see the study words that predict whether or not that word will be falsely recognized on the test as having been seen as a picture? Interestingly, Gonsalves & Paller (2000) found that ERPs recorded during the study phase in posterior cortical regions (i.e. occipital and parietal locations) were more positive for items in the 'word only' condition that were later falsely remembered as pictures than for those items that were correctly rejected as not accompanied with a picture. These positive ERPs may be providing an index of brain activity related to words that elicit strong visual imagery at encoding, which in turn enhances their likelihood of being falsely remembered as pictures on a later test.

5. MISATTRIBUTION AND FALSE RECOGNITION: VICES OR VIRTUES?

Misattribution and false recognition are sins of memory in the sense that they are associated with inaccurate reports about the past: people make claims about previous experiences that are, at least in some respects, incorrect. Similar sentiments apply to the other memory sins, which involve errors of omission or commission. These observations could easily lead one to conclude that evolution burdened us with an extremely inefficient, even defective, memory system—a system so prone to error that it may even jeopardize our well-being.

In contrast to this rather dark view of memory, Schacter (1999, 2001) argued that each of the sins is a by-product of adaptive features of memory, and in that sense they do not represent flaws or defects in the system (cf. Anderson & Schooler 1991; Bjork & Bjork 1988). Instead, the memory sins can be thought of as costs associated with benefits that make memory work as well as it does most of the time. Here we consider the case for the adaptive value of misattribution.

Consider that many instances of misattribution reflect poor memory for the source of an experience, such as the precise details of who said a particular fact, when and where a familiar face was encountered, or whether an object was perceived or only imagined. When source details are not initially well encoded, or become inaccessible with the passing of time, conditions are ripe for the kinds of misattributions associated with false recognition. But consider the consequences and costs of retaining each and every contextual detail that defines our numerous daily experiences. Anderson & Schooler (1991) (see also Schooler & Anderson 1997) have argued persuasively that memory is adapted to retain information that is most likely to be needed in the environment in which it operates. Because we do not frequently need to remember all the precise, source-specifying details of our experiences, an adapted system should not routinely record all such details as a default option. Instead, the system would record such details only when circumstances warn that they will be needed—and this is what human memory tends to do.

A second and related factor that contributes to misattributions involving false recognition concerns the distinction between memory for gist and specific information discussed earlier (§ 2) (Brainerd & Reyna 1998; Reyna & Brainerd 1995). False recognition often occurs when people remember the semantic or perceptual gist of an experience but do not recollect specific details. However, memory for gist may also be fundamental to such abilities as categorization and comprehension, and may facilitate the development of transfer and generalization across tasks. McClelland (1995, p. 84) has argued along these lines and noted that generalization often results from gist-like, accumulated effects of prior experiences. While depicting such generalization as central to intelligent behaviour and cognitive development, McClelland also noted that 'such generalization gives rise to distortions as an inherent by-product'.

We have already considered evidence that fits well with such a position, namely the consistent finding that amnesic patients show reduced false recognition of semantically and perceptually related words (Schacter *et al.* 1996*a*, 1997*a*, 1998*b*). Further, amnesiacs also show reduced false recognition of categorized pictures and novel objects that depend on retaining gist information (Koutstaal *et al.* 1999). These observations suggest that false recognition and misattribution based on gist reflect the workings of a normally functioning, healthy memory system. A damaged system such as that in amnesic patients does not either encode, retain or retrieve the information that leads healthy people to show high levels of gist-based false recognition.

Beversdorf *et al.* (2000) have recently provided striking evidence that supports this point. They studied adults with a condition known as autistic spectrum disorder. Such individuals are in many respects similar to autistic children, except that they tend to function at a higher level. Autistic children rely on a relatively literal style of processing information, and do not take advantage of semantic context to the same extent as normal children. Beversdorf *et al.* (2000) used the DRM semantic associates paradigm to assess whether adults with autistic spectrum disorder also show impaired sensitivity to semantic context. After studying lists of semantic associates, the autistic spectrum disorder group showed significantly less false recognition of semantically related lure words than did the healthy controls. However, the autistic spectrum disorder group showed normal hit rates for previously studied words, resulting in greater discrimination between true and false memories than in the control group. The autistic spectrum disorder group appeared to be relying on a highly literal form of memory, missing out on the semantic gist that supports false recognition responses. Once again, misattributions resulting in false recognition appear to reflect the operation of a healthy memory system.

These findings support the idea that misattribution can reflect adaptive aspects of memory function. Buss *et al.* (1998) (see Schacter 1999, 2001) have distinguished between two senses of the term 'adaptive' in evolutionary discussions. One is based in evolutionary theory: it rests on a technical definition of an adaptation as a feature of a species that came into existence through the operation of natural selection because it in some way increased reproductive fitness. The other is a non-technical, everyday sense of the term that refers to features of organisms that have positive functional consequences, regardless of whether they arose directly from natural selection. Some generally useful or 'adaptive' features of organisms are not adaptations in the strict sense. Sometimes termed 'exaptations' (Gould & Vrba 1982), these useful functions arise as a consequence of other related features that are adaptations in the technical sense. Such adaptations are sometimes coopted to perform functions other than the one for which they were originally selected.

With respect to the memory sins, it is difficult to determine definitively which (if any) are genuine adaptations and which are exaptations (for detailed discussion, see Schacter 2001). One strong candidate for an adaptation is the sin of persistence—intrusive, unwanted recollection of disturbing experiences—which usually results from the encoding and retrieval of enduring memories for highly emotional or arousing events. If persistence arose in response to life-threatening situations that endangered an organism's survival, animals and people who could remember those experiences persistently would probably be favoured by natural selection. If persistence is indeed an adaptation, many species should have neural mechanisms dedicated to preserving life-threatening experiences across lengthy time-periods. And, indeed, LeDoux (1996) has argued forcefully that the amygdala and related structures are implicated in long-lasting fear learning across diverse species, including humans, monkeys, cats and rats.

In contrast to this view of persistence, misattributions involved in false recognition are unlikely to be adaptations: it is difficult to see how or why remembering an experience inaccurately would result in increased survival and hence reflect the operation of natural selection. Instead, misattribution is more probably a by-product of adaptations and exaptations that have yielded a memory system that does not routinely preserve all the source-specifying details of an experience. Misattribution could also be a by-product of gist-based memory processes, which could have arisen as either adaptations or exaptations.

This sort of by-product fits what Gould & Lewontin (1979) called a 'spandrel': a type of exaptation that is a side consequence of a particular function. The term spandrel is used in architecture to refer to left-over spaces between structural elements in a building. Gould & Lewontin described the example of the four spandrels in the central dome of Venice's Cathedral of San Marco, which are left-over spaces between arches and walls that were later decorated with four evangelists and four Biblical rivers. The spandrels were not built in order to house these paintings, but they do the job quite well (for further discussion of spandrels, see Buss *et al.* 1998; Gould 1991).

If we think of misattribution and false recognition as psychological spandrels, they differ in at least one important way from the architectural spandrels discussed by Gould & Lewontin (1979). The latter generally have benign consequences, whereas misattribution and false recognition can cause serious problems, as illustrated by the consequences of faulty eyewitness identifications (Wells *et al.* 1998) and problems created by false memories recovered in psychotherapy (e.g. Bjorklund 2000; Loftus 1993; Lindsay & Read 1994; Pendergrast 1995; Schacter 1996). We thus might think of misattribution and false recognition as spandrels gone awry (Schacter 2001): consequences of adaptive features of memory sometimes turn the system's virtues into vices.

The authors are supported by grants from the US National Institute on Aging (AGO8441) and the National Institute of Mental Health (MH60941). We thank Steven Prince (Department of Psychology, Duke University, Durham, NC, USA) for assistance in manuscript preparation.

REFERENCES

Alba, J. W. & Hasher, L. 1983 Is memory schematic? *Psychol. Bull.* **93**, 203–231.

Anderson, J. R. & Schooler, L. J. 1991 Reflections of the environment in memory. *Psychol. Sci.* **2**, 396–408.

Beversdorf, D. Q. (and 10 others) 2000 Increased discrimination of 'false memories' in autistic spectrum disorder. *Proc. Natl Acad. Sci. USA* **97**, 8734–8737.

Bjork, R. A. & Bjork, E. L. 1988 On the adaptive aspects of retrieval failure in autobiographical memory. In *Practical aspects of memory: current research and issues* (ed. M. M. Gruneberg, P. E. Morris & R. N. Sykes), pp. 283–288. Chichester: Wiley.

Bjorklund, D. F. (ed.) 2000 *False-memory creation in children and adults: theory, research, and implications.* Mahwah, NJ: Lawrence Erlbaum.

Brainerd, C. J. & Reyna, C. F. 1998 When things that were never experienced are easier to 'remember' than things that were. *Psychol. Sci.* **9**, 484–489.

Brown, A. S. 1991 A review of the tip-of-the-tongue experience. *Psychol. Bull.* **109**, 204–223.

Budson, A. E., Daffner, K. R., Desikan, R. & Schacter, D. L. 2000 When false recognition is unopposed by true recognition: gist-based memory distortion in Alzheimer's disease. *Neuropsychology* **14**, 277–287.

Budson, A. E., Desikan, R., Daffner, K. R. & Schacter, D. L. 2001 Perceptual false recognition in Alzheimer's disease. *Neuropsychology* **15**, 230–245.

Buss, D. M., Haselton, M. G., Shackelford, T. K., Bleske, A. L. & Wakefield, J. C. 1998 Adaptations, exapations, and spandrels. *Am. Psychol.* **53**, 533–548.

Cabeza, R., Rao, S., Wagner, A. D., Mayer, A. & Schacter, D. L. 2001 Can the hippocampal memory system distinguish true from false? (Submitted.)

Ceci, S. J. & Bruck, M. 1995 *Jeopardy in the courtroom.* Washington DC: APA Books.

Ceci, S. J. & Friedman, R. D. 2000 The suggestibility of children: scientific research and legal implications. *Cornell Law Rev.* **86**, 33–108.

Conway, M. A. (ed.) 1997 *Recovered memories and false memories.* Oxford University Press.

Curran, T., Schacter, D. L., Norman, K. A. & Galluccio, L. 1997 False recognition after a right frontal lobe infarction: memory for general and specific information. *Neuropsychologia* **35**, 1035–1049.

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

Delbecq-Derouesné, J., Beauvois, M. F. & Shallice, T. 1990 Preserved recall versus impaired recognition. *Brain* **113**, 1045–1074.

Dodson, C. S. & Schacter, D. L. 2000 'If I had said it I would have remembered it': reducing false memories with a distinctiveness heuristic. *Psychonomic Bull. Rev.* **8**, 155–161.

Dodson, C. S. & Schacter, D. L. 2001*a* Memory distortion. In *The handbook of cognitive neuropsychology: what deficits reveal about the human mind* (ed. B. Rapp), pp. 445–463. Philadelphia, PA: Psychology Press.

Dodson, C. S. & Schacter, D. L. 2001*b* Aging, strategic retrieval processes and false recognition: reducing false memories with a distinctiveness heuristic. (Submitted.)

Dodson, C. S. & Schacter, D. L. 2001*c* When false recognition meets metacognition: the distinctiveness heuristic. *J. Memory Language.* (In the press.)

Durso, F. T. & Johnson, M. K. 1980 The effects of orienting tasks on recognition, recall, and modality confusion of pictures and words. *J. Verb. Learn. Verb. Behav.* **19**, 416–429.

Fabiani, M., Stadler, M. A. & Wessels, P. M. 2000 True but not false memories produce a sensory signature in human lateralized brain potentials. *J. Cogn. Neurosci.* **12**, 941–949.

Gonsalves, B. & Paller, K. A. 2000 Neural events that underlie remembering something that never happened. *Nature Neurosci.* **3**, 1316–1321.

Gould, S. J. 1991 Exaptation: a crucial tool for evolutionary psychology. *J. Social Issues* **47**, 43–65.

Gould, S. J. & Lewontin, R. C. 1979 The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc. R. Soc. Lond.* B **205**, 581–598.

Gould, S. J. & Vrba, E. S. 1982 Exaptation: a missing term in the science of form. *Paleobiology* **8**, 4–15.

Henson, R. N. A., Rugg, M. D., Shallice, T., Josephs, O. & Dolan, R. J. 1999 Recollection and familiarity in recognition memory: an event-related functional magnetic resonance imaging study. *J. Neurosci.* **19**, 3962–3972.

Hintzman, D. L. 1988 Judgements of frequency and recognition memory in a multiple-trace memory model. *Psychol. Rev.* **95**, 528–551.

Hunt, R. R. & McDaniel, M. A. 1993 The enigma of organization and distinctiveness. *J. Mem. Lang.* **32**, 421–445.

Israel, L. & Schacter, D. L. 1997 Pictorial encoding reduces false recognition of semantic associates. *Psychonomic Bull. Rev.* **4**, 577–581.

Jacoby, L. L., Kelley, C. M. & Dywan, J. 1989 Memory attributions. In *Varieties of memory and consciousness: essays in honor of Endel Tulving* (ed. H. L. Roediger III & F. I. M. Craik), pp. 391–422. Hillsdale, NJ: Lawrence Erlbaum Associates.

Jennings, J. M. & Jacoby, L. L. 1997 An opposition procedure for detecting age-related deficits in recollection: telling effects of repetition. *Psychol. Aging* **12**, 352–361.

Johnson, M. K. & Raye, C. L. 1981 Reality monitoring. *Psychol. Rev.* **88**, 67–85.

Johnson, M. K., Hashtroudi, S. & Lindsay, D. S. 1993 Source monitoring. *Psychol. Bull.* **114**, 3–28.

Johnson, M. K., Kounios, J. & Nolde, S. F. 1996 Electrophysiological brain activity and memory source monitoring. *NeuroReport* **8**, 1317–1320.

Johnson, M. K., Nolde, S. F., Mather, M., Kounios, J., Schacter, D. L. & Curran, T. 1997 The similarity of brain activity associated with true and false recognition memory depends on test format. *Psychol. Sci.* **8**, 250–257.

Koutstaal, W. & Schacter, D. L. 1997 Gist-based false recognition of pictures in older and younger adults. *J. Mem. Lang.* **37**, 555–583.

Koutstaal, W., Schacter, D. L., Verfaellie, M., Brenner, C. J. & Jackson, E. M. 1999 Perceptually based false recognition of novel objects in amnesia: effects of category size and similarity to category prototypes. *Cogn. Neuropsychol.* **16**, 317–341.

Koutstaal, W., Verfaellie, M. & Schacter, D. L. 2001 Recognizing identical vs. similar categorically related common objects: further evidence for degraded gist-representations in amnesia. *Neuropsychology* **15**, 269–290.

LeDoux, J. E. 1996 *The emotional brain.* New York: Simon & Schuster.

Lindsay, D. S. & Read, J. D. 1994 Psychotherapy and memories of childhood sexual abuse: a cognitive perspective. *Appl. Cogn. Psychol.* **8**, 281–338.

Loftus, E. F. 1993 The reality of repressed memories. *Am. Psychol.* **48**, 518–537.

McClelland, J. L. 1995 Constructive memory and memory distortions: a parallel-distributed processing approach. In *Memory distortion: how minds, brains and societies reconstruct the past* (ed. D. L. Schacter), pp. 69–90. Cambridge, MA: Harvard University Press.

Mather, M., Henkel, L. A. & Johnson, M. K. 1997 Evaluating characteristics of false memories: remember/know judgments and memory characteristics questionnaire compared. *Mem. Cogn.* **25**, 826–837.

Melo, B., Winocur, G. & Moscovitch, M. 1999 False recall and false recognition: an examination of the effects of selective and combined lesions to the medial temporal lobe/diencephalon and frontal lobe structures. *Cogn. Neuropsychol.* **16**, 343–359.

Miller, M. B. & Wolford, G. L. 1999 Theoretical commentary: the role of criterion shift in false memory. *Psychol. Rev.* **106**, 398–405.

Nolde, S. F., Johnson, M. K. & D'Esposito, M. 1998 Left prefrontal activation during episodic remembering: an event-related fMRI study. *NeuroReport* **9**, 3509–3514.

Norman, K. A. & Schacter, D. L. 1997 False recognition in young and older adults: exploring the characteristics of illusory memories. *Mem. Cogn.* **25**, 838–848.

Parkin, A. J. 2001 The structure and mechanisms of memory. In *The handbook of cognitive neuropsychology: what deficits reveal about the human mind* (ed. B. Rapp), pp. 399–422. Philadelphia, PA: Psychology Press.

Parkin, A. J., Bindschaedler, C., Harsent, L. & Metzler, C. 1996 Pathological false alarm rates following damage to the left frontal cortex. *Brain Cogn.* **32**, 14–27.

Parkin, A. J., Ward, J., Bindschaedler, C., Squires, E. J. & Powell, G. 1999 False recognition following frontal lobe damage: the role of encoding factors. *Cogn. Neuropsychol.* **16**, 243–265.

Payne, D. G., Elie, C. J., Blackwell, J. M. & Neuschatz, J. S. 1996 Memory illusions: recalling, recognizing, and recollecting events that never occurred. *J. Mem. Lang.* **35**, 261–285.

Pendergrast, M. 1995 *Victims of memory: incest accusations and shattered lives.* Hinesburg, VT: Upper Access.

Pierce, C. P. 2000 *Hard to forget: an Alzheimer's story.* New York: Random House.

Rapcsak, S. Z., Reminger, S. L., Glisky, E. L., Kasniak, A. W. & Comer, J. F. 1999 Neuropsychological mechanisms of false facial recognition following frontal lobe damage. *Cogn. Neuropsychol.* **16**, 267–292.

Reyna, V. F. & Brainerd, C. J. 1995 Fuzzy-trace theory: an interim synthesis. *Learn. Individ. Differences* **7**, 1–75.

Roediger III, H. L. 1996 Memory illusions. *J. Mem. Lang.* **35**, 76–100.

Roediger III, H. L. & McDermott, K. B. 1995 Creating false memories: remembering words not presented in lists. *J. Exp. Psychol. Learn. Mem. Cogn.* **21**, 803–814.

Roediger III, H. L. & McDermott, K. B. 1999 False alarms about false memories. *Psychol. Rev.* **106**, 406–410.

Roediger III, H. L. & Neely, J. H. 1982 Retrieval blocks in episodic and semantic memory. *Can. J. Psychol.* **36**, 213–242.

Roediger III, H. L., Balota, D. A. & Watson, J. M. 2001 Spreading activation and the arousal of false memories. In *The nature of remembering: essays in honor of Robert G. Crowder* (ed. H. L. Roediger III, J. S. Nairne, I. Neath & A. M. Surprenant). Washington, DC: American Psychological Association.

Rotello, C. M. 1999 Metacognition and memory for nonoccurrence. *Memory* **7**, 43–63.

Schacter, D. L. 1995 Memory distortion: history and current status. In *Memory distortion: how minds, brains and societies reconstruct the past* (ed. D. L. Schacter), pp. 1–43. Cambridge, MA: Harvard University Press.

Schacter, D. L. 1996 *Searching for memory: the brain, the mind, and the past.* New York: Basic Books.

Schacter, D. L. 1999 The seven sins of memory: insights from psychology and cognitive neuroscience. *Am. Psychol.* **54**, 182–203.

Schacter, D. L. 2001 *The seven sins of memory: how the mind forgets and remembers.* Boston, MA: Houghton Mifflin.

Schacter, D. L., Verfaellie, M. & Pradere, D. 1996*a* The neuropsychology of memory illusions: false recall and recognition in amnesic patients. *J. Mem. Lang.* **35**, 319–334.

Schacter, D. L., Curran, T., Galluccio, L., Milberg, W. & Bates, J. 1996*b* False recognition and the right frontal lobe: a case study. *Neuropsychologia* **34**, 793–808.

Schacter, D. L., Reiman, E., Curran, T., Yun, L. S., Bandy, D., McDermott, K. B. & Roediger III, H. L. 1996*c* Neuroanatomical correlates of veridical and illusory recognition memory: evidence from positron emission tomography. *Neuron* **17**, 267–274.

Schacter, D. L., Verfaellie, M. & Anes, M. D. 1997*a* Illusory memories in amnesic patients: conceptual and perceptual false recognition. *Neuropsychology* **11**, 331–342.

Schacter, D. L., Buckner, R. L., Koutstaal, W., Dale, A. M. & Rosen, B. R. 1997*b* Late onset of anterior prefrontal activity during retrieval of veridical and illusory memories: an event-related fMRI study. *NeuroImage* **6**, 259–269.

Schacter, D. L., Uecker, A., Reiman, E., Yun, L. S., Bandy, D., Chen, K., Cooper, L. A. & Curran, T. 1997*c* Effects of size and orientation change on hippocampal activation during episodic recognition: a PET study. *NeuroReport* **8**, 3993–3998.

Schacter, D. L., Norman, K. A. & Koutstaal, W. 1998*a* The cognitive neuroscience of constructive memory. *A. Rev. Psychol.* **49**, 289–318.

Schacter, D. L., Verfaellie, M., Anes, M. D. & Racine, C. 1998*b* When true recognition suppresses false recognition: evidence from amnesic patients. *J. Cogn. Neurosci.* **10**, 668–679.

Schacter, D. L., Israel, L. & Racine, C. A. 1999 Suppressing false recognition in younger and older adults: the distinctiveness heuristic. *J. Mem. Lang.* **40**, 1–24.

Schooler, L. J. & Anderson, J. R. 1997 The role of process in the rational analysis of memory. *Cogn. Psychol.* **32**, 219–250.

Schwartz, B. L. 1999 Sparkling at the end of the tongue: the etiology of tip-of-the-tongue phenomenology. *Psychonomic Bull. Rev.* **6**, 379–393.

Shiffrin, R. M., Huber, D. E. & Marinelli, K. 1995 Effects of category length and strength on familiarity in recognition. *J. Exp. Psychol. Learn. Mem. Cogn.* **21**, 267–287.

Squire, L. R. 1992 Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychol. Rev.* **99**, 195–231.

Strack, F. & Bless, H. 1994 Memory for nonoccurrences: metacognitive and presuppositional strategies. *J. Mem. Lang.* **33**, 203–217.

Tun, P. A., Wingfield, A., Rosen, M. J. & Blanchard, L. 1998 Response latencies for false memories—gist-based processes in normal aging. *Psychol. Aging* **13**, 230–241.

Underwood, B. J. 1965 False recognition produced by implicit verbal responses. *J. Exp. Psychol.* **70**, 122–129.

Wagner, A. D., Koutstaal, W. & Schacter, D. L. 1999 When encoding yields remembering: insights from event-related neuroimaging. *Proc. R. Soc. Lond.* B **354**, 1307–1324.

Ward, J., Parkin, A. J., Powell, G., Squires, E. J., Townshend, J. & Bradley, V. 1999 False recognition of unfamiliar people: 'seeing film stars everywhere'. *Cogn. Neuropsychol.* **16**, 293–315.

Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M. & Brimacombe, C. A. E. 1998 Eyewitness identification procedures: recommendations for lineups and photospreads. *Law Hum. Behav.* **22**, 603–647.