NEUROBIOLOGY OF RECONSTRUCTED MEMORY

W. Jake Jacobs and Lynn Nadel
University of Arizona

The authors present evidence that normal autobiographical memories and "recovered" autobiographical memories of long-forgotten traumatic events are produced by the same mechanisms. The basic process involves the parallel storage of information in a set of independent modules, the selective retrieval and reaggregation of this dispersed information within an appropriate spatiotemporal context, and the organization of this aggregate by a narrative. The result is a seamless blend of retrieved information (that which is recalled) and knowledge (that which is inferred) experienced as an autobiographical memory. The critical difference between normal and recovered memories, by this account, is the impact of trauma on the storage process: The physiological consequences of trauma can include a disabling of the neural module responsible for encoding the appropriate spatiotemporal context. Recovered memory involves retrieval of memory fragments, confabulation (innocent or not) driven by inference, and the fitting of a context to this incomplete aggregate. This too is experienced as an autobiographical memory. The implications of this view for estimating the veridicality of recovered memories are discussed.

Background

Remembering seems simple—experience something, form a memory of it, and some time later retrieve and report that memory. Evidence from cognitive and neural science, however, shows that what seems simple is actually quite complex and that the common-sense notion of remembering as a faithful report of prior experience is misguided. Although few would dispute that experience leaves traces behind in the brain, the nature of these traces, and how they contribute to the ability to remember the past, remain open questions. In this article, we seek to unravel memory so that we can ask meaningful questions about the veracity of specific recollections. We consider how information about prior experience is acquired, how it is stored in the brain, and how this stored information is used in remembering. We focus on the brain systems involved, on differences between infants and adults, and on the impact of traumatic stress. All of these considerations are important in assessing whether a specific memory from childhood or from a traumatic episode is veridical.

We start by posing some questions to help frame the discussion: (a) How is information of various kinds stored and subsequently used to remember the past? (b) What happens to stored information over time, and in what ways is it, or can it be, affected by subsequent experience? and (c) What effect do stress and trauma have on the storage and retrieval of various kinds of information? Central to each
of these questions is a broader one: Of what are memories made? This question can be posed at many levels. One level addresses the neurobiological substrates of memory, the cellular mechanisms and molecules responsible for the storage of information. A second level addresses the neuropsychological substrates of memory. Is there a unitary trace in the brain that contains all the stored information pertinent to any particular memory, or is this information fragmented and dispersed? A third level addresses the format in which information is stored. Is information stored much like audio or video recordings, as replicas of past experience retrievable in their entirety? Or are there other ways to conceive of memory that may be more consistent with what we know about how the brain processes, organizes, and stores information? Yet another level addresses the various ways in which memories are expressed. Can they be expressed indirectly through biases exerted on how we perceive, interpret, and integrate new experiences, directly through our actions, or only subjectively through verbal descriptions that we give to events in our life, or things we have learned in the past? Are all of these forms of expression reflections of the same memory process?

In this article, we sketch the broad outlines of how memory might work, how it is instantiated in the brain, and how stress and trauma affect it. Putting these ideas together with some understanding of the special conditions of early life or during trauma will allow us to at least begin to answer some of the questions raised above.

One Memory System or Many?

A key insight about the organization of memory that has emerged in recent decades is the notion that memory is not a single thing, but instead is subserved by multiple systems. Early psychologists such as Tolman (1948) stressed that there are several kinds of learning, but the idea that memory might depend on multiple systems surfaced as a consequence of the inability of early researchers to replicate in animal models the amnesic deficit reported in the case of H. M. (Scoville & Milner, 1957) and other patients with similar damage in the hippocampal complex and neighboring structures. Several investigators (Gaffan, 1974; Hirsh, 1974, Nadel & O'Keefe, 1974) suggested that the hippocampus might have a selective role in memory, readily tapped by the kinds of tasks used to assess deficits in H. M., but not tapped by the tasks used in the early animal studies (e.g., Orbach, Milner, & Rasmussen, 1960). Milner (1996) and Warrington and Weiskrantz (1968) had hinted at this insight, which was subsequently confirmed in studies with animals (e.g., Nadel & MacDonald, 1980; O'Keefe, Nadel, Keightley, & Kill, 1975) and humans (e.g., Cohen & Squire, 1980; Kinsbourne & Wood, 1975). The exact way to capture this selective role remains a hotly debated topic within cognitive neuroscience (see, e.g., D. L. Schacter & Tulving, 1994, for a range of views).

We take as the starting point in our discussion the idea that there are multiple forms of stored information, that these forms are subserved by separate neural systems, and that autobiographical memory entails accessing and integrating fragments of information from these separate systems. This assertion has two parts: (a) there are multiple systems underlying memory, a view that is now widely accepted, and (b) there is no such thing as a self-contained memory trace or engram for a specific episode that one could readily point to in the brain (if only one knew what to look for). As a corollary, it implies that there is no brain system
specialized just for the purpose of storing complete autobiographical memories. This perspective, long championed by Lashley (1950) among others, is more controversial, because it contradicts common-sense views of memory. The notion that experience is disaggregated for purposes of storage in the brain, and re-aggregated for purposes of memory retrieval, has serious consequences for thinking about memory for traumatic events and the veracity of recollections about such events.

Figure 1 provides an overview of the way we think about memory. It shows that experiences of the external world are transduced and disaggregated and then represented in a set of storage modules. It also shows that information from these modules can be used in perception or behavior or as the basis of an act of autobiographical remembering. In some of these cases an integration or re-aggregation process is necessary.

In discussing the selective function of the hippocampus in memory, O'Keefe and Nadel (1978) argued that this brain structure is concerned with the formation and storage of representations of environments that an organism experiences (in Tolman's, 1948, terminology cognitive maps), and with providing the spatial context essential to autobiographical episode memory in humans. Other neural systems represent the features and characteristics of objects such as food, predators, spiders, and the like. O'Keefe and Nadel referred to hippocampal and nonhippocampal systems as locale and taxon systems, respectively. Memory for autobiographical episodes involves the re-aggregation of featural information.
from the taxon systems within a contextual framework provided by the hippocampal locale system.

We have extended this modular view, arguing that there are at least five kinds of information-storing systems in the brain: feature systems, semantic systems, an evaluation system, procedural systems, and episodic context systems (see Nadel & Jacobs, 1996, for descriptions of these systems and relevant citations). These modules have distinct neural underpinnings and differ in the kind of information they process, represent, and store. They also contribute differentially to various forms of remembering, as we discuss in detail below. Their unique developmental trajectories and differential sensitivity to traumatic stress are central to the issue of recovered memory.

Memory Development

In most mammals, including humans, the hippocampus matures after birth (e.g., Seress, 1992), and its full contribution to memory is not available until this maturation has been completed (Nadel & Willner, 1989). Nadel and Zola-Morgan (1984) argued that the neurobiological immaturity of this critical system might account for at least a portion of "infantile amnesia." Many of the other modules involved in memory are, in contrast to the hippocampus, functional at birth, allowing a variety of forms of learning to occur (e.g., Bayer, 1980; Humphrey, 1968). Thus, much can be learned early in life, but we do not store (and hence cannot retrieve) information about the time and place in which this learning occurred. The information that is acquired exerts its effects in subtle and important ways, but does so in such a way that it is impossible for the individual to recall when and where it was acquired. These facts about memory function early in life indicate that various aspects of experience can be dissociated and can separately exert their influence over subsequent behavior. Following maturation of the neural systems underlying all forms of information storage, such dissociations are less easily observed. However, experimental manipulations or experiential situations that differentially influence the functional status of the separate neural memory modules can bring them out. Brain damage (or deterioration) is a dramatic example. Stress, and its neurobiological consequences, is a more common example and is central to this article.

Stress and Memory

In adults, stress triggers a related set of neurobiological activities leading to the release of three different hormones: corticosterone (cortisol in primates), epinephrine, and norepinephrine. Corticosterone has considerable impact on the hippocampus. Increased levels of corticosterone initially enhance, but eventually disrupt, normal activity of the hippocampus, with concomitant effects on its functions in encoding memories within a spatiotemporal frame of reference (de Kloet, Oitzl, & Joëls, 1993; Oitzl & de Kloet, 1992). Within a fairly wide range, increased concentrations of epinephrine and norepinephrine potentiate the function of the amygdala and hence facilitate the storage of information about the emotional aspects of experience (McGaugh, 1989, 1990). Thus, the storage of information in various memory modules can be differentially affected by increasing levels of stress, producing a continuum of memory retrieval characteristics. Every-day experience permits "normal" autobiographical episode memories within which
context, events, emotions, and the relations among them are seamlessly blended. At intermediate levels of stress, enhanced encoding of spatiotemporal attributes (as a function of the facilitative effect of corticosterone on hippocampus function) and enhanced encoding of emotional attributes (as a function of the facilitative effects of epinephrine and norepinephrine on amygdala function) can produce seamless and exceptionally vivid autobiographical memories. Traumatic levels of stress, however, simultaneously shut down the hippocampus-based system and facilitate the amygdala-based system, leading to the storage of emotional information unconnected to the contexts within which the relevant events occurred. Retrieval of this disconnected information, as we discuss in detail below, can produce disorganized, fragmented, and perhaps incomprehensible sense impressions and emotional memories (see, e.g., Thomas, Laurance, Jacobs, & Nadel, 1995; van der Kolk & Fisler, 1995).

Development, Stress, and Memory Conjoined

These notions about differential development and sensitivity to stress among various memory modules were at the root of our attempt to provide a neurobiological account of "irrational" fears and phobias (Jacobs & Nadel, 1985). The model addressed interactions between the neural systems separately concerned with fear and spatiotemporal context. This analysis was couched within the view, sketched above, that the brain contains a collection of semi-autonomous memory modules, each of which processes and represents information of a certain type; that these modules mature at different times in early life; and that they are differentially influenced by neurobiological consequences of stress. We emphasized the critical role of the late-maturing hippocampal system, necessary for autobiographical episode memory, and densely populated with receptors sensitive to the glucocorticoid hormones released during stress (McEwen, de Kloet, & Rostene, 1986). As noted above, there is an inverted U-shaped relationship between hippocampus function and these stress-related hormones. When the receptors for these hormones are either too sparsely or too extensively occupied, the behavior of the neurons containing them, and hence of the hippocampus, is severely disrupted (de Kloet et al., 1993).

This suppression of the normal function of the hippocampal system has several consequences. It produces a return to functional states appropriate to infancy or early childhood, because stress-induced disruption of the hippocampal system leaves intact nonhippocampal systems that are functional early in life. Jacobs and Nadel (1985) argued that this condition sets the stage for the return or reinstatement of emotional memories encoded in early-maturing systems under the veil of infantile amnesia. Under the right conditions, a specific phobia may appear, with the individual unaware of where, when, and how the phobia got started. This notion of an emotional memory, disconnected from its originating event, which we first developed in thinking about "irrational" fears, is central to our analysis of the characteristics of memory for traumatic events.

1The right conditions, according to the model, involve a conjunction of stress, which disables the hippocampal system and opens up the underlying emotional memory system, and the occurrence of an event capable of reinstating the buried emotional memory (e.g., a stimulus like the one involved in the early fear-provoking event).
One direct implication of our model is that an adult, under severe or traumatic stress, may acquire information of which he or she is not consciously aware. We discussed this implication some years ago (Tataryn, Nadel, & Jacobs, 1989), arguing that selective suppression of the hippocampal system by stress in the adult sets the stage for acquisition of emotional information devoid of autobiographical referents. We showed how the differential impact of stress on particular memory modules provides a neurobiological foundation for understanding the emergence of dissociative phenomena following trauma. Indeed, we argued that a range of syndromes (e.g., "blindsight"), characterized by the presence and efficacy of knowledge of which the bearer is not consciously aware, can be understood within this framework. In what follows, we further explore the ways in which modular information systems and stress can interact to yield a range of memory phenomena.

**Remembering and Memory Modules**

We have argued that life events and episodes are represented as disaggregated sets of traces of entities (objects, people, etc.); their interactions in space and time; their good, bad, or neutral characteristics; and the actions triggered by, or directed at or away from, them. From such life events we learn about the properties of these entities (how they look, sound, feel, smell, etc.), whether they lead to good or bad consequences, and how they were connected in space and time. We also learn something about how to react to entities and events and the consequences of our own actions.

By this analysis, the retrieval of memory for an event from the past does not involve simply pulling it out in its entirety from a single storage location. Rather, independent representations must be re-aggregated within a contextual framework identifying the fragments as part of the same event. Thus, an instance of producing a normal autobiographical memory is fundamentally a reconstructive activity. It follows that such an autobiographical memory is veridical if (a) the spatiotemporal context frame has been properly represented; (b) this frame has been linked to the entities that comprised the event; and (c) the frame, the entities, and the linkage between them are all available at the time the act of remembering is required. Disruption of the spatiotemporal frame or its linkage to the entities comprising the event can have profound consequences, leading to partial or even complete amnesia for personal experience.

This point is exemplified by the phenomenon of retrograde amnesia, in which some, but not all, forms of information acquired in the personal past may be lost because of some injury or brain damage. Recent research (cf. Nadel & Moscovitch, 1997, for a review) shows that when brain trauma or damage is relatively focal, quite specific, but intuitively puzzling, forms of retrograde amnesia can result. For example, the loss of autobiographical memory can occur with no loss in the recall of the faces and names of famous people from the past. Loss of episodic and personal semantic memory can co-exist with retention of general semantic and procedural forms of learning.

Traumatic stress provides a somewhat different example. It can cause amnesia for the autobiographical context of the stressful event and hypermnesia for the emotional memories produced by them. Because the structure critical to forming and storing the contextual frame (the hippocampus) is disrupted by the hormonal
cascade initiated by stress, this aspect of memory is lost. Because the structure critical to forming emotional memories (the amygdala) is enhanced by this stress-induced hormonal cascade, emotional hypermnesia results. Putting these two together would yield strong emotional memories divorced from their appropriate frame of reference.

In summary, because memory depends on dispersed information storage modules, the central questions become these: (a) How are representations that have been stored in a disaggregated fashion put together again in the service of memory retrieval? (b) Is this re-aggregation process influenced by the occurrence of traumatic stress either at the time of the event itself or at the time of retrieval? and (c) If re-aggregation is influenced by stress, how can we characterize the effects?

Re-Aggregation and Use

Figures 2–5 illustrate several ways in which dispersed representations might exert their influence during remembering. One possibility is a wholly re-aggregated autobiographical memory, for example recollecting where and what one ate for dinner last night (Figure 2). Another possibility is a partially re-aggregated procedural memory—knowing how to react on the squash court to a shot along the wall without recollecting where or when one learned it (Figure 3). Yet another possibility is a disaggregated memory (Figure 4), such as recalling the sound of a grandparent coughing at night many years ago or feeling the fear associated with an automobile accident some weeks ago. In addition to these behaviorally potent effects of stored information, there are cases where information can lie outside of awareness entirely, with no apparent behavioral impact (Figure 5). To understand both normal memory and the phenomenon of recovered memory we must explore each of these cases, beginning with the minimal case of behaviorally silent representations.

Figure 2. A normal autobiographical memory is retrieved when information from all relevant storage modules is integrated (step 3) and then used to generate an output (step 4).
Retrieval of a Partially Re-Aggregated Memory

Figure 3. A partially re-aggregated memory is retrieved when information from only some of the storage modules is integrated (step 3) and subsequently used for output purposes (step 4). In this case, information in the habit module (how to hit the ball) combines with information in a semantic system (what is the best strategy in this situation) to generate the output.

Retrieval of a Dis-Aggregated Memory

Figure 4. A dis-aggregated memory is retrieved when information from only one module is retrieved (3) and used for creating an output (step 4). In this case, the fear associated with an automobile accident is retrieved, alone, from the value module to generate an output.
Figure 5. Stored information can lie outside of awareness, with no obvious behavioral impact. As explained in the text, however, such information can nonetheless have an impact on how new experience is transduced, interpreted, and assimilated into existing stores of information (step 4). In this case information from any of the modules could have such an impact.

Behaviorally Silent Representations

One of the most clinically relevant examples of behaviorally silent representations is that of an “extinguished” emotional memory. Jacobs and Nadel (1985) summarized the data showing that when conditional emotional learning is followed by extinction (e.g., tone conditioned stimulus [CS]–shock unconditioned stimulus [UCS] pairings followed by CS alone presentations), the emotional memory for the pairing of tone and shock is not actually eliminated. Rather, new learning only inhibits but does not dissolve the old CS–UCS bond, rendering the surviving link behaviorally impotent (see LeDoux, Romanski, & Xagoraris, 1991, and Morgan, Romanski, & LeDoux, 1993, for empirical support). If this inhibition is disrupted (e.g., by disinhibition), conditioned emotional responses re-surface. Although such silent representations appear without behavioral consequence, they are not entirely without effect (see Figure 5). One indirect effect appears to be that these silent representations can bias the ease with which new emotional representations are formed. This assertion is consistent with a substantial body of experimental evidence ranging from Konorski and Swejkowska’s demonstrations of the primacy of first learning (Konorski & Swejkowska, 1950, 1952a, 1952b, 1956), through the several procedurally different demonstrations of savings effects following extinction procedures (e.g., Frey & Butler, 1977; Reberg, 1972; Rescorla & Heth, 1975; see Jacobs & Blackburn, 1995, for a theoretical review and Mackintosh, 1974, for an empirical review). Another indication of the survival of previously established emotional memories that have been “lost” through extinction or counterconditioning procedures is that they somehow remain available for renewal, or “reactivation,” even with highly degraded inputs.

This phenomenon is not limited to emotional memories. Experiences that yield
no explicit effects on behavior nonetheless partially determine what an adult might subsequently see (Blakemore, 1974; Hubel & Wiesel, 1965), hear (Werker & Tees, 1992), prefer to taste (Galef & Sherry, 1973), and do (Harlow & Mears, 1979). That is, the nature of information processing is shaped by circuits “prepared” through earlier experience (Bandura, 1977; Randich & LoLordo, 1979; Testa & Ternes, 1977). Thus at the very least behaviorally silent representations based on prior experience can implicitly bias the extraction of information from any subsequent experience.

From Behaviorally Silent to Disaggregated Memories

Some behaviorally silent memories, established during infancy or at times of traumatic stress and subsequently “forgotten,” may be reinstated by hormonal or environmental manipulations (see, e.g., Delamater & LoLordo, 1991, for a review). Such emotional information, however, does not necessarily reappear as conscious autobiographical memories. To the extent it enters awareness, it does so as a constellation of “body memories” (van der Kolk, 1994), highly elevated autonomic responses to reminders of traumatic incidents that may have happened years or even decades before testing (e.g., Blanchard, Kolb, Pallmeyer, & Gerardi, 1986; Dobbs & Wilson, 1960; Kolb & Multipassi, 1982; Malloy, Fairbank, & Keane, 1983; Pitman, Orr, Forgue, de Jong, & Clairborn, 1987). These body memories appear to be dependent upon previous experience and may have been established through associative conditioning, sensitization, or pseudoconditioning (see Mackintosh, 1974, for a review).

That such emotional hypermnesia may result from traumatic stress is also consistent with early reports (e.g., Charcot, 1887; Janet, 1889, 1893, 1909). Further evidence supporting this view comes from self-reports that, even in the presence of extensive autobiographical amnesia, intrusive emotions or images presumably associated with the trauma may appear. Intrusions appear during simple grief, anxiety disorders, mood disorders, or dissociative disorders (see, e.g., American Psychiatric Association, 1994; Brewin, 1996; Gibbs, 1996; Horowitz, 1986; and Howe, Courage, & Peterson, 1995, for selective reviews). Such intrusions may also be elicited in the laboratory. Van der Kolk (1994), for example, showed that about 20% of his participants (all of whom were diagnosed with posttraumatic stress disorder) experienced intrusions when presented with auditory stimuli that produce strong startle responses. Rainey et al. (1987) and Southwick et al. (1993) showed that injection of drugs such as lactate or yohimbe, which stimulate the autonomic nervous system, can trigger visual images associated with previous traumatic experiences in people with documented histories of trauma. Controls reported no such intrusions. Van der Kolk and Fisler (1995) presented evidence that intrusive memories for traumatic events initially are fragmentary (or disaggregated) and organized nonverbally, as intense waves of feelings, visual images, olfactory, or auditory or behavioral sensations. Participants in this questionnaire-based study universally reported they initially remembered the trauma as uncorrelated affective, behavioral, or sensory flashbacks without a coherent narrative memory for the traumatic events.

Elsewhere we have discussed the sequelae of such body memories (Jacobs, Nadel, & Hayden, 1992), arguing that they consist of four interlocking but imperfectly correlated constellations of responses: physiological, behavioral,
affective, and cognitive. Although the causal interactions among these constellations, if any, have not been specified, we proposed that close systemic connections among them exist. For example, there appears to be a close relationship between the physiological constellation (body memory) and behavioral constellation (species-specific defense reactions such as fight, flight, or freezing; Bolles, 1970). Similarly, it appears that people, in the presence of physiological or behavioral manifestations of such memory, experience emotional responses that at least partially depend on particular interpretations belonging to the cognitive constellation (e.g., S. Schacter & Singer, 1962). From this point of view, repeated uncontrollable manifestations of an emotional memory in the physiological or behavioral constellation interacts with affective and cognitive constellations to produce various anxiety disorders (Jacobs et al., 1992). The processes responsible for such interactions, we argued, remain largely outside of awareness (Tataryn et al., 1989).

Central to our present argument is the view that the acquisition and storage of information in the memory modules associated with these various constellations may be independent. Information re-aggregated from physiological, affective, or behavioral constellations without accompanying spatiotemporal contextual information will be experienced as intrusive (e.g., occurring without volition or autobiographical context). Such experiences bear the clinical labels body memories, emotional memories, and re-enactments (Jacobs, Laurance, Thomas, Luzcak, & Nadel, 1996).

Memory Retrieval: Where the Truth Lies—A Case Study

One of us (W. J.) was involved in a minor traffic accident on the 110 freeway in Los Angeles. His car (a Honda) was rear-ended by a large Buick. The Honda was pushed into a Ford Ranger, which was pushed into a Cadillac. The damage was minor, so the drivers exchanged information and went their way. About a year later, two of the passengers in the Cadillac initiated a lawsuit against the driver of the Ranger. W. J. was contacted by the defense attorney and, under penalty of perjury, was asked to describe what happened during the accident.

Initially, his memory of the accident was visual, one car hitting another. On reflection, he realized that he could not have seen what his memory said he saw. What he could actually recall was the sensation of his car lurching forward, stopping suddenly, lurching forward, and then coming to rest. His initial reaction was, "it must have been an earthquake—and I don’t want to be stuck in this tunnel if it's happening." There was no more movement and his car was sandwiched between two other vehicles. Each of four drivers then left their vehicles, met on the freeway, decided the damage was minor so they should move to a safe spot that did not impede traffic, exchange information, and be on their way. The rest of his event memory, that of seeing the Buick striking the rear of his car, his car striking the Ranger immediately ahead, and the Ranger truck striking the Cadillac, was composed of (a) physical sensations, (b) the information in his immediate visual field (the back end of the Ranger truck, sitting still), (c) the physical evidence he observed after the event, and (d) the narratives supplied by the other drivers.

---

2The behavioral constellation may induce re-enactments (e.g., Janet, 1893; van der Kolk, 1989; van der Kolk & Kadish, 1987). This effect, however, is less secure empirically than are the species-specific defense reactions.
In the deposition, one of the questions was “Did you see the Ranger strike the Cadillac and if so was it forced into the Cadillac as a result of your car being forced into the Ranger?” Although he can still see an image of the collision between the Ranger and the Cadillac, that image—as a veridical memory—is impossible. His view of the Ranger–Cadillac collision was hidden by the back end of a large truck (the Ranger) with a canopy attached.

To answer the question without perjuring himself, he was forced to deconstruct the event memory—and to acknowledge that, at best, he saw his car strike the back of the Ranger, forcing the Ranger forward, but the image of the Ranger striking the Cadillac must have been produced by some process that filled in a gap in the sequence of perceived events. Even now, much of the narrative he uses to describe the event fills gaps between what he remembers about the event (what he actually perceived) with what he knows about the event (what reasonably must have happened).

It appears this experience is not unusual. Many if not all of our autobiographical memories are composed of re-aggregated representations cobbled together by a narrative based on inference. Because these representations are retrieved within and organized by an appropriate spatiotemporal representation, and because the way events in the universe unfold in space and time permits our inferences to be relatively accurate, many if not most of our narratives reflect reality relatively well.

When this sort of process is observed in people with frontal lobe dysfunction, it is termed confabulation. The term originally referred to the fabricated stories produced by individuals with various forms of brain damage to questions about situations or events they could not recall (from Sidran Foundation Psych Trauma Glossary [retrieved from http://www.sidran.org/glossary.html]). More clearly, it has been defined as “a falsification of memory occurring in clear consciousness in association with an organic amnesia” (Berlyne, 1972, p. 33). These well-accepted definitions, which focused on the bizarre or patently false aspects of confabulated narratives, miss two characteristics that unite them with other forms of inferred narratives: (a) confabulations are innocent, done without intent; and (b) spontaneous confabulations in the brain injured do not represent a falsification of memory but rather an incoherent retrieval of memories “patched together” in a narrative form (Baddeley & Wilson, 1986; Kopelman, 1987a). Indeed, Korsakoff (e.g., 1889/1955) emphasized that his patients recalled unrelated but genuine events in a jumbled sequence rather than simply fabricating them (Huppert & Piercy, 1976, 1978; Meudell, Mayes, Ostergaard, & Pickering, 1985; Victor, Adams, & Collins, 1971). By focusing on the grandiose and bizarre juxtaposition of facts contained in confabulated narratives, some authors have missed the inferential consistencies and sound semantic structure found in almost all such narratives (Dalla Barba, 1993). Independent of content, confabulated narratives “make sense” as a story, with a beginning, an end, and appropriate thematic development.

A number of studies conducted in the 1980s attempted to distinguish between confabulation as seen in patients with organic problems and memory as seen in healthy people. In an important study, Kopelman (1987b) proposed that provoked confabulation (as opposed to spontaneous confabulation) is a nonpathological phenomenon that is relatively common among amnesic patients, and because it strongly resembles the memory performance of healthy people at long retention intervals, it may represent a normal response to faulty memory. Case reports
suggest that the subjective experience of a veridical memory is indistinguishable from that of a confabulated memory (e.g., Dalla Barba, 1993; Dalla Barba, Cipolotti, & Denes, 1990). Indeed, Kassin and Kiechel (1996) have shown that, under mild duress, people may falsely report their own actions. Once convinced, these individuals confabulate details in memory consistent with the false report. These data, which are consistent with a number of naturalistic reports (Ofshe, 1992; Pratkanis & Aronson, 1991; Wright, 1994), suggest that innocent confabulation is a normal state of autobiographical memory.

This case study, with its accompanying analysis of confabulation, suggests that normal autobiographical memory is produced through the re-aggregation of dispersed information and a “filling in” or “smoothing out” accomplished through inferential processes. This suggests that normal, autobiographical memory, as reported through a narrative, is a seamless blend of retrieved information (that which is recalled) and knowledge (that which is known through inferential processes), that is organized by a narrative—taking a story form with a beginning, end, and thematic development.

It is interesting to note that “filling in” is something the central nervous system does at a variety of levels. Examples are seen at the retina (filling in the blind spot), in a variety of perceptual phenomena (e.g., producing illusions through Gestalt principles such as “closure”), during learning (e.g., excitatory conditioning produced by random relations between the CS and the UCS), cognition (e.g., illusory correlations), and awareness (e.g., during sleep or hypnosis). Given the ubiquity of such inferential processes, it is perhaps not surprising that a similar phenomenon is observed in the memory domain.

From Disaggregated to Re-Aggregated (Recovered) Memories

As we have discussed, autobiographical episode memory depends on the integration of information from various modules. When re-aggregation occurs properly, a contextual representation in the hippocampal system acts to bind together the various featural representations stored in other memory modules (see Metcalfe & Jacobs, 1998; Nadel & Moscovitch, 1997), and an appropriate autobiographical memory is retrieved. Somewhat more speculatively, we suggested that autobiographical memories are enhanced, or supplemented, by innocent confabulation—a process that through inference “fills in the holes” and somehow produces a narrative based on that which is recalled and that which is inferred. Even when examining normal autobiographical memory, it requires some skill to distinguish what is inferred from what is recalled.

Substantial bodies of data suggest that the critical role played in this process by the hippocampus can fail under two distinct conditions: (a) early in life the hippocampus may not store the spatiotemporal context information essential to episode memory because of biological immaturity, and (b) during traumatic stress, the hippocampus may not store the spatiotemporal context essential to episode memory because of hormones, such as the glucocorticoids, that selectively target neurons in the hippocampus, changing their anatomical structure and excitability. Under both of these conditions, information critical to normal autobiographical memory is missing because it is unavailable at the time of retrieval.

Putting this material together, we suggested that anatomical immaturity or traumatic levels of stress may produce “scotoma” in memory; infantile amnesia in
the first case (see Nadel & Zola-Morgan, 1984) and trauma-induced amnesia in the second (Jacobs et al., 1996; Nadel & Jacobs, 1996). These scotoma are experienced as a selective loss of autobiographical information for periods immediately before, during, and after maturation or traumatic stress. It is critical to note that in such cases there is not a general failure of memory. Representations are formed and stored in modules that are already mature or unaffected by stress, but they lie outside of awareness (Tataryn et al., 1989).

We have elsewhere reviewed evidence supporting the notion that hippocampal immaturity contributes importantly to the subjective experience of infantile amnesia (Nadel & Zola-Morgan, 1984). The assertion that stress can cause amnesia for autobiographical content is consistent with reports of trauma-induced amnesia dating back at least to the time of James (1884), Janet (1889, 1893, 1909), and Charcot (1887). Partial or complete amnesia has been reported by a substantial number of people following combat exposure, torture, kidnapping, experience in concentration camps, murder, and physical or sexual abuse (Archibald & Tuddenham, 1956; Briere & Conte, 1993; Burkett & Bruno, 1993; Elliot & Briere, 1995; Goldfield, Mollica, Pesavento, & Faraone, 1988; Grinker & Spiegel, 1945; Hendin, Haas, & Singer, 1984; Herman & Schatzow, 1987; Kardiner, 1941; Kinzie, 1993; Kubic, 1943; Loftus, Polensky, & Fulfilove, 1994; Madakasira & O’Brian, 1987; Myers, 1940; Niederland, 1968; Sargeant & Slater, 1941; D. L. Schacter, 1986; Sonnenberg, Blank, & Talbot, 1985; Southard, 1919; Thom & Fenton, 1920; van der Kolk & Kadish, 1987; Wilkinson, 1983; Williams 1994, 1995). Trauma-induced amnesia appears related to age and to the intensity and duration of the trauma (see, e.g., van der Kolk & Fisler, 1995; and van der Kolk, McFarlane, & Weisaeth, 1996, for reviews).

When complete re-aggregation fails, so too does autobiographical memory. This, however, is not the end of the story, for the human mind apparently abhors the discontinuity of an episodelike memory retrieved without some link to the when and where of its occurrence—and without a coherent narrative to describe events as they unfold in space and time. Our model suggests that discontinuous, disaggregated memories, memories without a spatiotemporal frame to bind them together, should obtain on initial retrieval of traumatic memories. This, of course, is exactly what is observed in the clinic and what has been reported from the laboratory. As described above, van der Kolk and Fisler (1995) presented evidence that memories of traumatic events are initially experienced as uncorrelated waves of feelings, images, or sensations without a coherent narrative memory for the traumatic events. Only subsequently do these memories take on an autobiographical character. These data suggest that an autobiographical memory of a traumatic event can eventually emerge in narrative form, but only after memory manifestations from physiological, behavioral, and emotional constellations “came into consciousness with greater intensity [and] more sensory modalities were activated” (van der Kolk, 1996, p. 289).

We propose, as a natural extension of the basic characteristics of memory outlined above, that retrieval, re-aggregation, and inferential narrative smoothing of traumatic memories work by the same principles as retrieval, re-aggregation, and inferential narrative smoothing of any other autobiographical memory. Specialized representations are retrieved from modules and re-aggregated; inno-
cent confabulation, driven by inferential processes, enhances or elaborates the re-aggregated mass through a narrative. Once the narrative is supplied, a sense of familiarity or autobiographical fit occurs. The only thing that distinguishes a recovered memory of trauma from a recovered memory of an ordinary event is what can be retrieved. For a traumatic memory, the information from one module, the appropriate spatiotemporal frame, is missing.

If the "true" context of a set of representations cannot be retrieved, some other context will be fit to the task. Memories without a context, which we take as equivalent to intrusive memories, lie at the root of dissociative phenomena and are the substrate for all recovered memories following amnesia. Memories retrieved in this way are necessarily discontinuous, a characteristic that triggers inferentially driven innocent confabulation. Confabulation, in conjunction with the bits and pieces that have been retrieved, completes a coherent narrative, which is experienced as an autobiographical memory.

Memory Unraveled

Until recently, the term memory was used only in mentalistic contexts. During the past 50 years or so, a neurobiologically based model of memory has been elaborated and although there remain many points of contention, its central features are accepted by most investigators. To restate: Information is disaggregated and stored as representations in specialized modules. These representations can influence the ways in which subsequent events are interpreted and stored. They may be recalled directly or re-aggregated with other representations in ways that can yield behavior or full-blown autobiographical recollections.

Stress-induced amnesia, the starting point for any concept of "recovered" memory, is an involuntary, adaptive biological function that occurs without awareness. It reflects failures of encoding and storage in the presence and aftermath of severe traumatic events and involves only traumatic material. Critically, it does not involve self-deception, which distinguishes it from two other processes that have been proposed to account for amnesia: denial and repression.³ It has, for ethical reasons, been most extensively studied naturalistically. We have taken these naturalistic data seriously for several reasons, the most important of which is that these, like all good naturalistic studies (e.g., Lorenz, 1965; Tinbergen, 1951; von Frisch, 1955), draw the attention of experimentalists and theoreticians to ecologically important phenomena. These conclusions have notable connections to, and implications for, folk, forensic, and clinical psychology.

Folk Psychology

The folk theory of human psychology is reflected in our every-day language. Folk language contains several intriguing hints about human memory, as reflected in the meaning of the words used to describe memory and remembering. That there

³Unlike the notion of stress-induced amnesia that we have outlined here, denial and repression have been used primarily in mentalistic contexts, imply self-deception, occur after encoding and storage have been completed, involve intentionality and some level of awareness, produce amnesia for both traumatic and nontraumatic material, and have been produced in the laboratory (e.g., Erdelyi, 1990). We believe that stress-induced amnesia is fundamentally different from both denial and repression.
might be a relationship between the folk language used to discuss memory and current biological and clinical theory is not surprising, because scientific thinking is frequently constrained by the framework ordinary language provides.

Dictionaries do not generally contain a reference to memory as a characteristic of a physical system. Instead, memory is referred to as a characteristic of mind. According to the on-line version of Webster's Dictionary, 10th ed. [available to subscribers from http://www.eb.com:180], “memory” is a noun that refers to “the power or process of reproducing or recalling what has been learned and retained especially through associative mechanisms.” It is also “the store of things learned and retained from an organism’s activity or experience as evidenced by modification of structure or behavior or by recall and recognition.” Thus, memory refers both to the ability to remember and to what is remembered. Both storage and retrieval are embraced within these definitions.

Focusing for the moment on the retrieval aspect of memory, Webster's highlights a series of words to describe the re-expression of previously learned skills or information. The most important of these are remember, recollect, recall, and remind. “Remember implies a keeping in memory that may be effortless . . . Recollect implies a bringing back to mind what is lost or scattered . . . recall suggests an effort to bring back to mind and often to re-create in speech . . . Remind suggests a jogging of one’s memory by an association or similarity.”

These words fall into two classes: the first implies bringing and aggregating a collection of individual “memories” to awareness (re-member and re-collect); the second implies bringing a complete memory directly to awareness (re-call and re-mind). Whereas each may mean to bring an image or idea from the past into awareness, they imply that the quality of the memories can be quite different. This distinction, captured in the folk language about memory, is very similar to what we have proposed in this article. Thus, we have argued that autobiographical memories require a process of complete re-aggregation, whereas emotional memories do not. Remembering and recollecting, like autobiographical memories, can be subject to a variety of distorting influences. Recalling and reminding, like emotional memories—or what van der Kolk (1994) has termed body memories—are not.

Such conclusions are counterintuitive. We observe our own memory, see that it agrees with the accounts of others, and assume the accounts are veridical, brought directly to mind. Theories describing autobiographical memory as a complex reconstructive process do not match this view. If, under normal conditions, autobiographical reports are a reconstruction of past reality, subject to distortion, then memories can be composed of nonveridical accounts of personal history. If trauma and traumatic stress prevent certain kinds of information from being stored, then any autobiographical recollection of such events must be at least partially a fabrication—a confabulated account of past episodes. Such conclusions raise difficulties for those taught to respect historical accounts as sincere and veridical. Nonetheless, modern cognitive and neural science shows that, at least under the conditions we have outlined, personal memory is subject to distortion from sources that include the way in which it is encoded and represented, processes that may subsequently influence and alter these representations, and both internal and external forces at the time of recollection.
Forensic Psychology

For the forensic psychologist, both the data and the theoretical conclusions based upon them must occasion pause, because it is the forensic psychologist who testifies about the veracity of recovered traumatic memories. What must be determined in assessing the veracity of any recovered memory is whether the context that has been fitted to the details bears any relation to the context in which the events unfolded. If normal memory is readily subject to incorrect re-aggregations, as the case study and analysis of confabulation we have provided suggest, how else are we to determine whether a recovered memory related to trauma is veridical or not?

To restate the premise: Information represented in the hippocampus provides a spatiotemporal frame of reference that organizes the storage and retrieval of information represented in other modules appropriate to an event that happens at a particular time and place. Under certain circumstances, these spatiotemporal markers are missing. We have argued that all representations laid down without an appropriate spatiotemporal frame are fair game for recall by "reminders" of one or more traumatic events (see Jacobs et al., 1996); in what follows we summarize that argument.

When conditioning lays down emotional representations without significant contributions from the hippocampal system, they enter a "pool" where they remain isolated, temporally disordered, and stimulus bound. Consider only emotional representations that might be in such a pool. We might think of this as an aggregation of emotional representations, each equally privileged for retrieval. Internal or external cues (CSs) that resemble appropriate stimuli in the original traumatic episode(s) may retrieve these emotional representations. The strength of a retrieved representation is, of course, described by the shape of the generalization gradients around the original conditional stimuli.

A parsimonious way to think of the retrieval process is that each cue samples the population of representations making up the pool. Most natural conditions will retrieve more than one representation because of the many cues available to elicit these emotional representations. Thus, sampling driven by a complex naturalistic situation will not be random; instead, it will be driven by extant stimuli. Nonetheless, we use the notion of random sampling as a heuristic to illustrate our point.

Characteristics of the Pool

The collection of emotional representations in a pool may show little, moderate, or wide variability, depending on the number, quality, and type of traumatic experiences the person has encountered. In addition, the distribution of the collection may be normal, skewed, or multimodal. These characteristics affect the probability of an individual representation being drawn in the face of a set of CSs.

Because we store and retrieve emotional memories from modules, one might think of the retrieved material as formally equivalent to a limited random draw from an isolated pool (module) of numbers. On retrieval from a module, a representation "intrudes" on ongoing experience not as a normal memory, but instead as fragmented and disorganized, without the beginning, end, and internal spatiotemporal structure of an autobiographical event (Thomas et al., 1995). Once
drawn, these representations integrate within an ongoing spatiotemporal frame of reference and stabilize, partly through a confabulated narrative construction. As described above, humans blend such material by effortless, innocent, and unintentional narrative confabulation (e.g., Gazzaniga, 1985), organized by an internal or an external schema (e.g., Bartlett, 1932; Mandler, 1984).

For the sake of simplicity, let us characterize the fully integrated experience of retrieval and stabilization as the mean of repeated but limited random draws from the pool of emotional representations. With such a simplifying assumption, we can represent the veracity of a fully re-aggregated memory by comparing the obtained mean with a number that represents the experience. Assuming each draw is unbiased, we can use this mean to describe various patterns of veracity. We first generate these descriptions in the abstract and then apply them to specific examples.

If the numbers in the pool are homogeneous, the sample mean (the recovered memory) will represent each number (the experiences) well. There will be some differences (error) between the sample mean (recovered memory) and any given number (actual experience) in the pool, but the recovered memory is an adequate characterization of most of the actual experiences (representing a "theme"). Although the memory misrepresents each experience, it represents the overall structure of the experience adequately.

If the numbers in the pool (experiences) vary widely, or if the distribution is not normal, then the mean (recovered memory) misrepresents most of the numbers (experiences) in the pool: It is a poor characterization of any theme(s).

Finally, if we bias the draw in some way (e.g., we retrieve the representations through structured experiences), then the mean (recovered memory) may accurately or inaccurately represent the numbers (experiences) in the pool. The accuracy of a recovered memory depends on the imposed bias but not on the variability or shape of the distribution in the pool.

To make these ideas more concrete, suppose a person has experienced one massive traumatic event, such as an intense, long-lasting firefight in combat. Under this condition, a homogeneous pool of emotional representations exists. Given the random sampling, a recovered memory of the actual event should be relatively accurate because there is a limited set of representations to re-aggregate and innocent confabulation, bound by appropriate spatiotemporal inferences, should (or at least can) organize the representations coherently. We would have some, but not complete, confidence in a memory recovered under these circumstances.

Suppose a perpetrator has repeatedly assaulted a person in one place, at a fixed time of night, with one act. As before, that person will form a homogeneous pool of emotional representations. Although a memory recovered from this pool may not represent an individual assault, it has a reasonable probability of capturing the "theme" of the assaults. Again, we would have some confidence in memory recovered under these circumstances.

Now suppose a person has experienced distinct traumatic events, such as physical and sexual assaults by several different perpetrators, a variety of traumatic war experiences, and several automobile accidents. Under these conditions, a person will form a heterogeneous pool of emotional representations. A recovered traumatic memory may badly misrepresent each traumatic event. Because the risk
of misrepresentation under these circumstances is high, we would have little confidence in memory for events recovered under these circumstances.

Finally, suppose a person's experience has produced a pool of emotional representations with reasonable variance and two distinct themes. Suppose also that each draw from the pool is not random, but guided by another person with an agenda. Perhaps the guide is (a) committed to a particular model of psychopathology, (b) has been shaped to interpret memories for trauma in a particular way, or (c) has a factual account of the traumatic event. Under such guidance, the person would retrieve certain types of representations but not others. Thus, a recovered memory would reflect a combination of the person’s emotional representations, the person's schema, the guide's schema, and the statistical characteristics of the pool. Under these circumstances an imposed schema would have a small effect given homogeneity in the pool (e.g., repeated stereotypical assault, a single massively traumatic experience), but large effects given heterogeneity in the pool (e.g., many distinct traumatic experiences). In the latter case, the (unknown) relation between the imposed schema and the “target” traumatic event determines the impact of the guide's schema on the veracity of recovered memory. Unless the guide has access to a factual account of this event, one would do well to remain suspicious of memory for events recovered under these circumstances.

Our analysis suggests that the retrieval of memory for normal and traumatic events from the past does not involve simply pulling them out in their entirety from single storage locations. Like all other autobiographical memories, a recovered memory for traumatic events involves independent representations re-aggregated within a contextual framework identifying the fragments as part of the same event, and a “filling in” or “smoothing out” of the information accomplished through inferential processes. This suggests that, like a normal autobiographical memory, a recovered traumatic memory reflects a fundamentally reconstructive and inferential activity—a process that produces a seamless blend of retrieved information (that which is recalled) and knowledge (that which is known through inferential processes), that is organized by a narrative—taking a story form with a beginning, end, and thematic development. By this account, all memories, normal or traumatic, are “recovered” memories.

However, when memory storage fails because of stress-induced disruption of the hippocampus, the spatiotemporal frame of the traumatic event is not properly represented. Without such representation, the frame cannot be linked to the entities that comprised the event. Hence the frame, the entities, and the linkage among them are not available at the time of retrieval. Instead, memory fragments—physical, emotional, cognitive, and behavioral memories derived from the full complement of traumatic (and possibly other) experiences—are independently available. As before, innocent confabulation works to produce a seamless blend of retrieved information (memory fragments) and knowledge (in this case spatiotemporal context and narrative), which is experienced as an autobiographical memory. The veracity of such a memory depends on the retrieval of event-appropriate fragments and inferring a spatiotemporal frame that approximately (a) represents the original spatiotemporal frame, (b) links the representation with event-appropriate fragments, and (c) is available when the act of remembering is required. Thus, disrupted storage of the spatiotemporal frame can have profound consequences, including at one end of the scale the creation of false event
memories, but also including at the other end of the scale the creation of more or less true event memories.

Conclusion

We have shown that much has been learned about how memories are stored and retrieved, yet our discussion of the veridicality of recovered memories indicates that much more remains to be determined. The characteristics of a "pool" or "collection" of modularized representations of trauma are unknown (if such a pool exists at all). The ways that elements of such a pool might be sampled, how they might be aggregated, and how they might be smoothed into a coherent narrative are only now beginning to be explored. Although a psychologist might attest to the intensity, the frequency, or even the circumstances under which emotional reactions arise, a psychologist cannot, on the basis of psychological methods alone, testify to the absolute veracity of any given autobiographical memory. It may become possible to use psychological methods to determine broad probabilities about the veracity of memory and hence to invoke various standards of proof, but without independent corroboration by others, by records, and by physical evidence, modern cognitive and neural science cannot unequivocally determine the truth of recovered memories.

References


Received March 14, 1997
Accepted September 24, 1998