Memory and the Sense of Personal Identity

Stanley B. Klein  
University of California, Santa Barbara  
stan.klein@psych.ucsb.edu  

Shaun Nichols  
University of Arizona  
sbn@email.arizona.edu

Memory of past episodes provides a sense of personal identity—the sense that I am the same person as someone in the past. We present a neurological case study of a patient who has accurate memories of scenes from his past, but for whom the memories lack the sense of mineness. On the basis of this case study, we propose that the sense of identity derives from two components, one delivering the content of the memory and the other generating the sense of mineness. We argue that this new model of the sense of identity has implications for debates about quasi-memory. In addition, articulating the components of the sense of identity promises to bear on the extent to which this sense of identity provides evidence of personal identity.

1. Introduction

Memory is at the heart of the way most people think about personal identity. It is because I remember my first kiss that I think I am the same person as that awkward adolescent. If I had no memory of past experiences, the sense that I existed in the past would be dramatically compromised.

Memory is also at the heart of philosophical discussions of personal identity. Perhaps the most prominent account of personal identity, attributed to Locke, holds that these kinds of memories are (part of) what make me the same as the person I was in the past. Memories of past actions go towards constituting personal identity. Locke’s immediate philosophical opponents, Reid and Butler, rejected the constitution thesis. But they did not shrink from relying on memory to ground judgements of personal identity. On the contrary, Reid and Butler took memory to provide the critical evidence of
past existence. Thus, Reid writes:

Our own personal identity and continued existence, as far back as we remember anything distinctly...we know immediately, and not by reasoning. It seems, indeed, to be a part of the testimony of memory. Everything we remember has such a relation to ourselves as to imply necessarily our existence at the time remembered. (Reid 1785, p. 586).

Even philosophers like Hume, who reject the idea that there is an enduring self, still typically acknowledge the force of memorial experience in giving the impression of identity across time. Hence we find Hume saying, ‘As memory alone acquaints us with the continuance and extent of this succession of perceptions, ‘tis to be considered, upon that account chiefly, as the source of personal identity’ (Hume 1739/2000, p. 168; see Baxter 2008 for treatment of Hume’s complex and evolving position on self and memory). Indeed, it is incumbent on theorists who deny persistence of self to prove that memory presents us with an illusion of identity.

Despite the manifest role for memory in philosophical discussions of personal identity, there is remarkably little discussion of the sense of personal identity that memory delivers — the sense that I am the same person as someone in the past. Lockeans use this sense to build a theory of personal identity; Reid and Butler appeal to this sense as evidence for a persisting self; and Humeans maintain that the sense presents us with an illusion of a persisting self. This raises a cluster of philosophical issues about personal identity and memory. Does memory of experiences always bring with it a conviction of identity? Does memory necessarily presuppose personal identity? Does memory provide evidence for personal identity across time? To address these issues adequately, we need a sharper picture of the nature of the sense of identity.

In this paper, we will begin by reviewing the work on memory, and we will present and discuss a neurological case study of a patient who lacks an experienced sense of ownership of his autobiographical memories. We suggest that this illuminates the psychological underpinnings of the sense of identity and that this might have important implications for issues surrounding personal identity.

2. Memory and personal identity

2.1 Contemporary psychological treatments of long-term memory

Prior to presenting the case study, we need to be explicit about how contemporary psychologists conceptualize long-term memory.
Our view of memory is based on the widely held position that long-term memory consists in multiple systems (for reviews, see Foster and Jelicic 1999; Klein et al. 2004; Schacter and Tulving 1994). Most psychologists agree that, at its most general level, long-term memory can be characterized as containing two basic systems: procedural and declarative (e.g. Cohen 1984; Squire 1987; Tulving 1983, 1995). The procedural memory system makes possible the acquisition and retention of motor, perceptual, and cognitive skills (for example, knowing how to ride a bike); it consists in the nonconscious expression of previously acquired behavioural skills and cognitive procedures (see Parkin 1997; Tulving and Schacter 1990). Declarative memory, by contrast, consists in facts and beliefs about the world (for example, knowing that canaries are yellow; knowing that Nolan Ryan pitched seven no-hitters).

Tulving (1972, 1983) further distinguishes two types or systems within declarative memory: episodic and semantic (see also Cermak 1984; Furlong 1951; Gennaro 1996; Parkin 1997). Semantic memory contains relatively generic, context-free information about the world, such as Grapes are edible, 2 + 2 = 4, and Sacramento is the capital of California. Semantic memory lacks a source tag — it does not specify when or where the memory was acquired; rather, it is retrieved as knowledge without regard to where and when that knowledge was obtained (e.g. Klein 2004; Tulving 1983, 1995; Wheeler, Stuss, and Tulving 1997). Although semantic memory typically does not make reference to the self, it can contain propositions expressing facts about the self (for example, Stan Klein was born in New York in 1952), just as it can about other things in the world. But this information is known in the same way that one knows that 2 + 2 = 4; it is not recollected or re-experienced vis-à-vis the context in which it was originally learned.

In contrast to semantic memory, episodic memory records events as having been experienced at a particular point in space and time. On standard characterizations, what distinguishes episodic memory is that an episodic memory represents the ‘what, where, when’ of an event. As such, it is experienced as a memory that makes explicit reference to the time and place of its acquisition. Examples of episodic memory are I remember eating chicken for supper yesterday evening; I recall my meeting with Judith last Monday.¹

¹ There are delicate issues about how the self is represented in episodic memory, and this will be a concern later in the paper.
Not surprisingly, it is the episodic component of declarative memory that historically has been the focus of interest for psychologists (for review, see Klein 2001, 2004; Klein and Gangi 2010) and philosophers (e.g. Brennan 1985; Campbell 2004; Parfit 1984; Northoff 2000; Schechtman 1990) studying the relation between self and memory. This is because retrieval from episodic memory is assumed to have a self-referential quality thought to be largely absent from other types of memory (for discussion, see Klein 2001; Klein et al. 2002). In particular, episodic memory has been thought to involve re-experiencing events from one’s past, thus providing its owner with content by which he or she is able to construct a personal narrative, that is, his or her life stories (see, for example, Eakin 2008; Fivush and Haden 2003; Klein 2001; Klein and Gangi 2010). By contrast, semantic memory is assumed not to be accompanied by awareness of re-experiencing one’s personal past (e.g. Klein 2001; Perner and Ruffman 1994; Tulving 1993a; Wheeler et al. 1997): I may know where I was born, but I do not know this by virtue of having re-experienced my birth. That is why this bit of personal history would be considered semantic memory knowledge, despite its being about oneself.

The self is represented by a number of systems and subsystems within long-term memory (see fig. 1). Within semantic memory, there are (at least) two different kinds of self-related memories. First, we have semantic factual knowledge of the self—for example, the memory that I am 58 and live in Goleta. This kind of memory is

---

2 For reviews, see Gillihan and Farah 2005; Klein et al. 2004; Klein and Gangi 2010; Rathbone, Moulin, and Conway 2009.

---

Figure 1: Standard model of long-term memory systems that implicate self representations
dissociable from a second kind of semantic self-knowledge, knowledge of one’s own traits (for review, see Klein et al. 2008; Klein and Lax 2010). Research over the past twenty years has provided evidence that the semantic memory system contains a specific subsystem that stores information about one’s own personality in the form of trait generalizations (for example, Self: usually stubborn). These trait summaries form a fast-access database, which provides quick answers to decision processes that require trait judgements (for review, see Klein et al. 2002; Klein and Gangi 2010). This system delivers a sense of the self given by pre-computed summaries of the dispositions one manifested in various behavioural episodes (e.g. Klein and Loftus 1993).

This dispositional sense of the self turns out to be resilient across dramatic damage to memory systems. There now exists an extensive database showing that even patients suffering total anterograde and retrograde episodic amnesia can describe their own personal characteristics both reliably and accurately (e.g. Klein, Loftus, and Kihlstrom 1996; Tulving 1993b; for reviews, see Klein 2001, 2004, forthcoming a; Klein and Lax 2010). This, of course, is not unexpected given the extensive literature demonstrating clinical and experimental dissociations between episodic recollection and trait self-knowledge (for reviews, see Klein forthcoming a; Klein et al. 2008; Klein and Lax 2010).

Although self-trait memory provides a critical sense of self, it is not sufficient for a sense of personal identity across time. This is illustrated by a number of patients, but the patient D.B. provides what is perhaps the most dramatic case. Following cardiac arrest and presumed hypoxic brain damage, D.B. had knowledge of his own traits, and yet he was ‘unaware that he had a past and unable to imagine what his experiences might be like in the future’ (Klein et al. 2004, p. 466; see also Klein, Loftus, and Kihlstrom 2002). This presumably is because D.B. had severe damage to episodic memory rendering him incapable of recollecting a single event or experience from any point in his past. This absence of episodic memory in D.B. explains his lack of a sense that he existed in the past.\(^3\) Episodic memory seems to be essential for the sense of personal identity across time (Atance and O’Neill 2005).\(^4\)

\(^3\) Perhaps more strikingly, the absence of episodic memory in D.B. seems to coincide with an inability to imagine himself into the future (Klein, Loftus, and Kihlstrom 2002).\(^4\) This is not the only sense of personal continuity available, however. While it is true that D.B. lacks a memory-based sense of diachronic identity, there is a sense of self—a sense of who he is—that is preserved. Indeed, even when a person experiences cognitive chaos so extreme that it partitions conscious access to his thoughts, memories, and perceptions into
Strikingly, when episodic memory is partially intact, people retain a sense of identity even if there has been dramatic memory loss. Consider the famous case of H.M., who had a large portion of his medial temporal lobes surgically removed to alleviate seizures. After the surgery, H.M. could not form new episodic memories. But he reveled in telling a few stories about his childhood. One of H.M.’s biographers recounts: ‘He tells of living in South Coventry, Connecticut, where he could go behind the house to shoot guns. He had a rifle, “One with a scope!” he says, always enthusiastic at this point in his story. “And I had handguns too! A .38 and a .32”’ (Hilts 1995, p. 138). The difference here between the child and the man is profound. After surgery, H.M. was no longer capable of taking care of himself, he could not remember whether his parents were alive or dead, he did not know where he was living, and he was, as noted, unable to form new episodic memories. And yet he seems to take himself to be identical with the young man with a .38 and a .32.

In his critique of memory-based accounts of personal identity, Reid recognized something like this distinction between episodic and semantic memory. Although Reid famously rejects the Lockean view that memory constitutes personal identity, he maintained that memory provides ‘irresistible’ evidence that I am the very person who did the action (Reid 1785, p. 318). To remember a past action is to identify oneself as existing at the time: ‘Every man in his senses believes what he distinctly remembers, and every thing he remembers intervals of one second (that is, awareness of current thought and perception fades and is replaced each passing second!), a sense of self-unity and continuity reportedly is felt during each momentary slice of awareness. This held for Storring’s ‘Patient B’ (Storring 1936). The patient under discussion was unable to maintain a continued focus on any internal or external topic, event, or scene exceeding his brief limit of continued awareness, yet he had a sense of himself as an existing, continuous entity (for review and discussion, see Klein forthcoming a).

Given the episodic criterion for psychological continuity discussed above, what might provide for this personally felt and unwavering sense of personal continuity? One possibility is that there is a sense of personal continuity that derives from trait self-knowledge, a subsystem within semantic memory (see also Rathbone et al. 2009). An extensive review of the available research (Klein and Lax 2010) shows that knowledge of one’s traits (a) is immune to loss in the face of multiple, often severe, neurological and cognitive insult (including total retrograde and anterograde amnesia, autism, Alzheimer’s Dementia, and Prosopagnosia), (b) is empirically dissociable from trait knowledge of others (even well-known others such as the patient’s family members) as well as from purely factual, non-dispositional self-knowledge, and (c) can serve as a firmly entrenched foundation for one’s sense that one ‘is, was, and will be’ (for a similar view, see Rathbone et al. 2009).

In short, the surprising and unanticipated resilience of trait self-knowledge may serve as the bedrock for one’s sense that one is a continuing, experiencing self even when one’s memory-based personal narrative has succumbed to the ravages of episodic amnesia.
convinces him that he existed at the time remembered’ (Reid 1785, p. 318). Reid here clearly has in mind what we now call episodic memory. He acknowledges that I might know who gave birth to me, but denies that this is achieved through memory. Today, we would classify knowledge of one’s birthmother as a fact retrieved from semantic memory, but Reid simply has a narrower use of ‘memory’ that seems restricted to episodic memory (as is true of many psychological and philosophical treatments of memory prior to the early 1970s—though not all; see, for example, Bergson 1911/1970; De Biran 1803; Furlong 1951; Musant 1966). More important for present purposes, Reid maintains that it is logically impossible to have such a memory without also having the conviction that one existed at that time: ‘There can be no memory of what is past without the conviction that we existed at the time remembered.’ To suppose otherwise is a contradiction (Reid 1785, p. 315). Similarly, in the contemporary literature, Marya Schechtman maintains the ‘mineness’ of a memory ‘cannot be separated from its content’ (Schechtman 1990, p. 87).

Psychologists do not typically trade in issues of logical impossibility, but there is a natural psychological thesis that resonates with Reid’s point. It might well be that the nature of episodic memory is such that it is nomologically impossible to remember a past action without ‘the conviction that one existed at that time’. It is possible that the episodic memory system simply cannot produce any output unless the output includes the sense that the same self had the experience that is being remembered (see, for example, Suddendorf and Corballis 1997). Some psychological systems seem to be tied to very particular kinds of representations. For instance, the facial perception system simply cannot produce a percept of a frontal concave face. As the hollow-face illusion shows, if the facial features are present in a concave mask, it is inevitable that we see the face as convex. It is possible that something similar holds for episodic memory and the representation of the self. One viable psychological hypothesis is that the retrieval of an episodic memory inevitably issues a representation of self-as-owner. On this view, one simply cannot get the content of episodic memory without also getting the ‘mineness’ along with it (see, for example, Klein et al. 2004; Wheeler et al. 1997).

Reid argues that the Lockean attempt to define personal identity in terms of memory is circular because memory presupposes identity. Contemporary theorists invoke ‘quasi-memory’ to address this problem, and we will discuss at length below how the case studies might bear on the notion of quasi-memory.
2.2 The case of patient R.B.

Having situated the discussion within the context of the memory systems, we turn to the case of Patient R.B., a right-handed, forty-three-year-old male with a BA and a Master’s degree in physics. While riding his bike, he was struck by an automobile, resulting in multiple injuries, including broken ribs, multiple fractures of his pelvis, a collapsed lung (pneumothorax), and head trauma. At the time of his accident, which took place in 2002, he was employed as an electronic design engineer.

Of direct relevance to the issues at hand is the nature of the cognitive and memory impairments sustained as a result of his trauma. These included difficulty in maintaining attention, mild aphasia, and retrograde and anterograde amnesia for the events in close temporal proximity to the accident (for example, several days both prior to and following neural insult. After several months, R.B. was able to recover memories for a few of these pre- and post-accident events). Verbal fluency (the ability to generate exemplars of categories in a fixed amount of time — for example, names of zoo animals) and short-term memory span both were slightly (but significantly) below scores provided by neurologically healthy age-matched controls.

While in the hospital, R.B. was placed on morphine (IV drip, followed by pills) and Oxycontin to alleviate the considerable pain he endured. He was switched to Vicodin following release from the hospital, and, as the intensity of pain subsided, he weaned himself off medication. Importantly, at the time of memory testing R.B. was not on any pain medication. After returning home, he was seen regularly by speech and physical therapists to help with his aphasia and to regain his ability to walk (lost due to the shattering of his pubic ramus; for additional case details, see Klein forthcoming b).

Of particular relevance to the question at hand — personal ownership and episodic memory — is that R.B. was able to remember particular incidents from his life accompanied by temporal, spatial, and self-referential knowledge, but he did not feel the memories he experienced belonged to him. In his words, they lacked ‘ownership’. This particular form of memory impairment — episodic recollection absent a sense of personal ownership, is a form of memory dissociation that, to our knowledge, has not previously been documented in the neurological literature. There are, however, two cases that bear some similarity. One is from a case study of an amnesic reported by Talland (1964). Unfortunately, the data available from that brief report, while suggesting a similar dissociation, are too limited to support strong
conclusions. In the other case (Stuss and Guzman 1988), a patient semantically relearned his ‘personal history’ following a case of severe retrograde episodic amnesia spanning most of his past life history. The patient commented that the relearned memories seemed to lack a feel of real happenings in his life. They were, to him, more like stories and facts told to him by others (which, indeed, they were!). In this sense, they were more like semantic facts about himself (e.g. Klein forthcoming a; Klein and Lax 2010) than episodic recollections: the patient knew his memories were about him, but he did not remember them as temporally and spatially acquired in the correct context (that is, when they transpired). Instead, they were memories that temporally and spatially were experienced episodically (they consisted in the episodic recall of information acquired following the onset of his retrograde memory loss) as second-hand stories told to him at a particular time and place.

So, are R.B.’s recollections episodic or semantic? The evidence suggests the former. Almost immediately following his accident, R.B. was able to intentionally recall specific events temporally and spatially situated in his personal past, but his recollection of those events was compromised in an unusual manner — memory for those events, though fitting the standard criteria for episodic recollection, was not accompanied by a sense of personal ownership.

For example, approximately two months following release from the hospital (by which time he was no longer taking pain medication), R.B. offered the following description of what it is like for him to recall personal events:

What I realized was that I did not ‘own’ any memories that came before my injury. I knew things that came before my injury. In fact, it seemed that my memory was just fine for things that happened going back years in the past [the period close to the injury was more disrupted]. I could answer any question about where I lived at different times in my life, who my friends were, where I went to school, activities I enjoyed, etc. But none of it was ‘me’. It was the same sort of knowledge I might have about how my parents met or the history of the Civil War or something like that.

R.B. himself initiated use of the language of ‘ownership’, and he uses this expression throughout his reports. We simply adopt his expression below.

In our review of contemporary theories of memory, we noted that psychologists maintain that episodic memories are temporal, spatial, and self-referential. Semantic memory, by contrast, largely (though, as we have noted, not exclusively) enables retrieval of impersonal
facts (for example, *A bird is an animal*). By these criteria, R.B.’s descriptions of his memory experience leave little doubt that they are personal recollections, appropriately situated in time and space, rather than factual semantic knowledge. For instance, following R.B.’s recall of events from his childhood, he added:

> I was remembering scenes, not facts … I was recalling scenes … that is … I could clearly recall a scene of me at the beach in New London with my family as a child. But the feeling was that the scene was not my memory. As if I was looking at a photo of someone else’s vacation.

Note that these memories were substantiated by third parties as valid renditions of events that actually transpired in R.B.’s life. While this recollection satisfies the criteria for episodic memory — time, place, and self— it simultaneously exhibits a highly atypical absence of experienced ownership. This absence of ownership again is evident in R.B.’s response to instructions to recall personal memories from time spent in graduate school:

> Things that were in the present, like my name, I continue to own. Having been to MIT had two different issues. My memories of having been at MIT I did not own. Those scenes of being at MIT were vivid, but they were not mine. But I owned ‘the fact that I had a degree from MIT’. That might have simply been a matter of rational acceptance of fact.

He continues:

> I can picture the scene perfectly clearly … studying with my friends in our study lounge. I can ‘relive’ it in the sense of re-running the experience of being there. But it has the feeling of imagining, [as if] re-running an experience that my parents described from their college days. It did not feel like it was something that really had been a part of my life. Intellectually I suppose I never doubted that it was a part of my life. Perhaps because there was such continuity of memories that fit a pattern that lead up to the present time. But that in itself did not help change the feeling of ownership.

Once again, R.B.’s memory performance adheres to the definition of episodic recollection: he can remember where the events transpired, when the events took place, and that they directly involved him. Moreover, they are not simply autobiographical semantic knowledge or inferences based on semantic memory (although, as his response makes clear, he is capable of such inference).

An important attribute of episodic recollection is that it can be intentionally called into awareness. While episodic recollections can be spontaneous (e.g. Bernsten 2010; Salaman 1970), their ability to be
auto-cued (Donald 1991) is one of the features that separates this type of memory ability from the types of cue-dependent memories that appear to characterize most non-human sentient creatures (Suddendorf and Corballis 1997). In this regard, R.B. shows clear ability to intentionally bring to mind past personal events. Consider, for example, the following exchange between Klein and patient R.B.:

SBK: Can you recall personally important events from your pre-injury period?
RB: I remember things that came before my injury. In fact, it seems that my memory is just fine for things that happened going back years in the past. I can answer questions about where I lived at different times in my life, who my friends were, where I went to school, activities I enjoyed … To clarify, I am remembering scenes, not facts. Since I was remembering scenes, I think this means I am dealing with exactly what you are asking about.

Note that R.B.’s responses exhibit a clear sense of self both narratively and factually. That he also has an intact dispositional self can be seen from his remarks as the interview continues.

SBK: Can you recall who you are? More specifically, what you were like and what you are like—that is, your trait characteristics. If so, are your traits felt as your own?
RB: Yes, I definitely have no identity problem. And the memories created since the injury I have full ownership of.

SBK: Can you recall for me a personal event concerning your time at college that would involve knowing what happened to you as a personal experience. Or is the recall more of a factual nature?
RB: I can see the scene in my head. I’m studying with friends in the lounge at my residence hall. I am able to re-live it. I have a feeling … a sense of being at there, at MIT, in the lounge. But it doesn’t feel like I own it. It’s like I’m imagining, re-living the experience but it was described by someone else.

SBK: Can you recall memories whenever you desire to do so?
RB: I can recall memories [from the non-ownership period of his life] at will. I have normal control over remembering facts and scenes from my past. But when I remember scenes from before the injury, they do not feel as if they happened to me—though intellectually I know that they did—they felt as if they happened to someone else.

Commenting on his gradual recovery of physical function (for example, walking) and the eventual return of feelings of personal
ownership of his episodic recollections, R.B. observes:

It [that is, loss of experience of personal ownership of his episodic recollections] helped explain the other puzzle: I didn’t feel ‘down’ about not being able to walk, etc. Because it was as if I was learning to walk for the first time. There was no sense of loss. Only a sense of gaining new skills and meeting these interesting new challenges. I knew that I once could walk, but it wasn’t ‘me’ who once could walk.

With respect to the recovery of episodic ownership:

When I did ‘take ownership’ of a memory, it was actually quite isolated. A single memory I might own, yet another memory connected to it I would not own. It was a startling experience to have no rhyme or reason to which memories I slowly took ownership of, one at a time at random over a period of weeks and months.

He continues:

What happened over the coming months was interesting: every once in a while, I would suddenly think about something in my past and I would ‘own’ it. That was indeed something ‘I’ had done and experienced. Over time, one by one I would come to ‘own’ different memories. Eventually, after perhaps eight months or so, it seemed as if it was all owned. As if once enough individual memories were owned, it was all owned.

From a related perspective in the psychological memory literature, R.B. has at least some capacity for what Suddendorf and Corballis (1997, 2007) have termed the act of mental time travel (see also Klein, Loftus, and Kihlstrom 2002; Wheeler et al. 1997) — that is, the ability to recollect one’s personal past as well as imagine or project one’s self into a possible future. Importantly, to date, this ability has only been found in individuals with intact episodic memory (for recent review, see Suddendorf and Corballis 2007). What is unique to R.B.’s case is the phenomenological distance at which he experiences his episodic recollections. He knows they are his in some sense, but he feels as though they are not ‘owned’. The following quote makes clear that even during the period of non-ownership, R.B. retained his capacity to consciously plan for his personal future, an ability, as we have noted, widely considered causally linked to the presence of intact episodic memory function (e.g. Hassabis et al. 2007; Klein, Loftus, and Kihlstrom 2002; Tulving 1985; for review see Klein 2010; Suddendorf and Corballis 2007).

During the un-owned period I was able to plan for the future. Although my working memory loss and lack of skill at compensation made it challenging. When I slowly returned to work, it was hard to plan a complex
strategy. I had to think of useful things to do and then do them. The best compensation I found was to separate the planning of the strategy from the execution. It worked best if I made a list of Things To Do. Then I could handle doing them one at a time.

3. Implications

3.1 A sketch of a model of the sense of identity
As we have seen, the sense of personal identity given by episodic memory is robust—it gives us an ‘irresistible’ sense of being the same person, and it has seemed to many to be a necessary concomitant of episodic memory. But the case of R.B. indicates that this sense of identity is dissociable from episodic memory itself. The sense of identity turns out to be, pace Reid, a contingent feature of memory.

The apparent paradox—episodic memory of past events absent the feeling of personal ownership—can be understood by situating episodic memory in the context of a system of interrelated memory processes, some of which provide the raw data for experience and some of which enable the experience to be ‘mine’ (for extended treatment, see Klein 2004, 2010, forthcoming a, forthcoming b; Klein et al. 2004). R.B.’s recollections during his ‘unowned’ period can be explained in the context of the view that there is specialized neural machinery that inserts the conceptual element self into the agent slot of an episodic memory attribution (Klein forthcoming b; see also Klein et al. 2004). This neural machinery in R.B. seems to have been compromised by his injury, but only for those events that occurred in the time period preceding his injury. That is, R.B. suffered a form of retrograde amnesia that compromised his ability to experience his personal recollections as his own.

During the non-ownership period, R.B. had episodic memory for past events, but lacked a sense of numerical personal identity with the past person. The case of R.B. suggests that the sense of numerical personal identity is quite narrowly circumscribed. R.B. had factual self-knowledge, trait self-knowledge, and knowledge of episodes, but he did not have the sense of personal identity with the past person.

Importantly, during his ‘unowned’ period R.B. had no trouble representing that R.B. had experiences on a beach in New London. He presumably also had no trouble representing that his body was present for those experiences. He knows that the memories are about him rather than his mother. And he can auto-cue the memories at will.
So, in that sense, the memories involve self-reference. However, there seems to be another type of self-reference that typically accompanies episodic recollections (‘ownership’, ‘mineness’) that has been impaired in his case. His apparent deficit was in representing, from the first person, ‘I had these experiences’. The nature of this deficit remains unclear. One natural explanation is that there is a special kind of conceptual self-representation that is missing in R.B.’s episodic memories. There is independent reason to think that there is a special kind of conceptual self-representation corresponding to (some uses of) ‘I’ (see, for example, Rey 1997, p. 291; Nichols 2008, pp. 522–3). To adapt an example from John Perry, if Hector learns that Hector is about to be attacked by a bear, this will not generate the appropriate action unless Hector also thinks I am Hector (Perry 1977, p. 494). The corresponding ‘I’-concept plays a special role in our cognitive economy. It is plausible that this concept is also implicated in typical cases of episodic recollection and that it is this special representation that is somehow missing in R.B.’s episodic recollections, giving rise to a phenomenological difference in the sense of ownership.

The fact that R.B. was able to recover these functions suggests that one facet of the self-referential aspect of his recollections was not destroyed by his injury. This is also implied by the fact that he had a sense of personal ownership of ongoing experiences that occurred after the accident that impaired his memory (with the exception of his temporally moderate anterograde memory loss). Why his mental machinery was able to insert a distinctive I-tag into memories as they were being built but not when recollecting memories of past events is yet another feature that remains unclear (for relevant discussion, see Dalla Barba 2002; Tulving and Lepage 2000).

3.2 The sense of identity and quasi-memory

According to psychological continuity approaches to personal identity, what makes me the same person as someone in the past is a psychological relation between me and that past person. Until recently, discussions of personal identity have been dominated by psychological approaches, and, in particular, memory-based approaches to personal identity (but see Olson 1997; Shoemaker forthcoming). On a simple Lockean memory theory, I am the same person as someone in the past just in case I remember having an experience as that person. Reid and Butler maintained that this proposal entails an obvious circularity, for
I can only remember my own experiences (Butler 1736; Reid 1785). Thus, for the Lockean account, remembering presupposes personal identity. And if memory presupposes personal identity, then trying to give an account of identity in terms of memory seems hopeless.

Although philosophers widely agree that the circularity objection is a serious problem for a simple Lockean theory, a number of theorists still favor a psychological continuity account of diachronic personal identity (as opposed to, say, biological accounts; e.g. Olson 1997, 2007). Accordingly, potential emendations have been proposed to rein in the tautology identified in the simple Lockean account (e.g. Bernecker 2010; Brennan 1985; Collins 1997; Grice 1941; Hamilton 1995; Lund 2005; Quinton 1962; Slors 2001a). The most widely discussed technique for reformulating the Lockean account in the face of the circularity objection draws on the notion of quasi-memory. Originally proposed by Shoemaker (1970), Parfit defines quasi-memory as follows.

I have an accurate quasi-memory of a past experience if:

1. I seem to have an experience,
2. someone did have this experience,
3. my apparent memory is causally dependent, in the right kind of way, on that past experience. (Parfit 1984, p. 220)

Quasi-memories thus do not presuppose a persisting self. With this definition of quasi-memory, a Lockean can maintain, without circularity, that quasi-memory partly constitutes personal identity.

Advocates of psychological-continuity accounts of personal identity tend to take the quasi-memory account as solving the problem of circularity, thus making it safe to hold a Lockean theory of some form. However, in a lovely paper in The Journal of Philosophy, Schechtman offers a powerful contemporary defense of the circularity objection. She argues that it is impossible to specify how memory matters for personal identity without presupposing identity (1990, p. 79). To make her case, she begins by noting that the notion of quasi-memory is supposed to evade the circularity objection by characterizing ‘a nondelusional memory without reference to personal identity’ (Schechtman 1990, p. 79). Schechtman charts the relation between memories, delusions, and quasi-memories as follows: to have a memory, she suggests,

is to have an apparent memory of an experience that one actually had, and to take it (correctly) to be one’s own experience. To have a delusion is to
have an apparent memory of an experience that one did not in fact have, and to take it (incorrectly) to be one’s own. To have a quasi-memory is to have an apparent memory (properly caused) and to hold no view about whose memory it is. (Schechtman 1990, p. 78)

Quasi-memories then, are supposed to be nondelusional apparent memories, and since having a quasi-memory leaves open whose memory it is, quasi-memory does not presuppose personal identity.

With this framework in place, Schechtman develops an ingenious argument against the possibility of quasi-memories, which we will call Schechtman’s dilemma. The dilemma holds that, for any candidate quasi-memory, the state will either be a delusion or it will presuppose personal identity. As a result, no candidate quasi-memory will work to solve the problem of circularity: in each case, the apparent memory will either be a delusion or it will be a state that presupposes sameness of rememberer and experiencer. She writes, ‘simply deleting the “nametag” from a memory is not sufficient to make it nondelusional’ (Schechtman 1990, p. 79). The only way for an apparent memory not to be delusional will be to ‘presuppose the identity of the rememberer with the person who had the experience’ (Schechtman 1990, p. 79). The root of the problem, according to Schechtman, is that ‘The mine-ness of a psychological state cannot be separated from its content’ (Schechtman 1990, p. 87).

We think that the case of R.B. provides a way for the quasi-memory theorist to evade Schechtman’s dilemma.7 Schechtman’s dilemma

---

6 We are simplifying Schechtman’s argument here, since she in fact allows a further alternative—however, that alternative is, in her view, another dead end for quasi-memory. The other way to make the quasi-memory nondelusional would be to strip away much of the content of the apparent memory; but in that case, she says, ‘it is no longer plausible to say that what is relevant to personal identity in genuine memory is preserved in quasi-memory’ (Schechtman 1990, p. 79). Relatedly, in recent work, Schechtman has noted that even if it is possible to transfer some quasi-memories, there remains a question about whether the transferred quasi-memories are the sort that can be constitutive of personal identity (Schechtman 2011, p. 70). Our primary interest in this paper is the sense of identity, and so we focus on the narrower dilemma drawn from Schechtman. We note, though, that responding to this narrower dilemma will be a necessary part of any defence of quasi-memory. The question then arises whether quasi-memories that evade the dilemma can be partly constitutive of personal identity; but we leave that question to others.

7 The debate about personal identity and memory is often framed in terms of conceptual analysis and conceivability (see, for example, Schechtman 1990, pp. 71, 73, 76). Thus, one might maintain that actual empirical evidence on memory cannot be relevant to what is conceivable. We reject this assumption. Empirical work can show that a distinction is empirically viable, and this might change the way we think about what is conceivable. For instance, after the discovery of blindsight, people plausibly find it easier to conceive of a distinction between perception and awareness.
casts doubt on the coherence of the notion of quasi-memory. It seems that there is no space for the relevant notion of quasi-memory because for any apparent memory, it will either (i) be a delusion or (ii) presuppose identity. Now consider R.B.’s recalling ‘a scene of me at the beach in New London with my family as a child.’ The veracity of his account was corroborated (as were the other events R.B. recounted). So these recollections are accurate recountings of events. The original fixation of the event in memory was presumably caused by the normal processes of memory consolidation (for recent reviews, see Roediger, Dudai, and Fitzpatrick 2007). And R.B. reports being able to retrieve these sorts of memories at will, just as we can normally retrieve episodic memories. Furthermore, R.B. does not doubt the content of the recollections. So it is hard to see why these states should count as delusions in any sense at all.

Consider the other horn of the dilemma—that the apparent memory presupposes personal identity. Does R.B.’s apparent memory of the scene on the beach presuppose personal identity with the person on the beach? We think the answer is ‘no’. To see why, note that it is an open theoretical question whether, at the time of recall, R.B. is the same person as the child on the beach. R.B. himself seems to regard the child on the beach in much the way he regards his mother on the beach—as a separate person. Schechtman’s own account might well take R.B. to be a different person after the accident. She writes,

Part of what is involved in a psychological state’s being mine in the sense which is at issue in the self-knowledge question, then, is its coherence with my total psychology—my ability to view it as a comprehensible part of my life, and to take it to be my own. (Schechtman 1990, p. 90, emphasis added)

What we find in R.B. is precisely that he does not take the apparent self-referential memory to be clearly his own. He reports not being depressed by the fact that he could not walk because ‘it was as if I was

Northoff (2000) draws on empirical findings in support of the realizability, and thus conceivability, of quasi-memory. However, to address Schechtman’s dilemma, he relies primarily on work on schizophrenia. He describes a case of a schizophrenic woman who maintained that she was the granddaughter of the last czar. At late stages of the pathology, Northoff maintains that the patient’s apparent memories were clearly delusional (Northoff 2000, p. 205). However, at earlier stages, Northoff maintains, the apparent memories were not delusional because the patient could still distinguish between ‘her own and others’ memories’ (Northoff 2000, p. 206). We are sympathetic to Northoff’s project here. However, the schizophrenic patient’s apparent memories, even in early stages, are clearly delusions in some sense, and this makes the case less than optimal as a response to Schechtman’s dilemma.
learning to walk for the first time’. This would lead some theorists to maintain that R.B. really is not the same person as the person represented in the apparent memory. Of course, some psychological continuity theorists might want to say that R.B. really is the same person as the person who figures in his apparent memories, even if he denies it. Indeed, many philosophers would maintain that R.B. remains morally responsible for his past actions, even if he lacks a sense of ownership over them. The kinds of psychological connections that are critical to identity on some theories — convictions, values, character traits, knowledge of friends and relatives — show great continuity in R.B.’s case. A psychological continuity theory that affirms the persistence of R.B. might be wrong, but it cannot be accused of circularity. R.B.’s recollections provide an existence proof for a kind of memory that navigates between the horns of Schechtman’s dilemma. In R.B., the ‘mineness’ of episodic recollection is separated from the content.

Thus far, we have argued that the case of R.B. offers the quasi-memory theorist a response to an important objection. But the case also raises important questions about quasi-memory. In his definition of quasi-memory, Parfit says that an accurate quasi-memory must be ‘causally dependent, in the right kind of way’ on a past experience (Parfit 1984, p. 220). What counts as a memory being caused in the right kind of way? In section 3.1, we suggested that a special kind of conceptual self-representation (the ’I’-concept) is typically implicated in episodic memory. During his non-ownership period, R.B.’s episodic recollections seem to lack this conceptual element. But even during this time, R.B. retained other kinds of self-representations that he connected with the ‘unowned’ memories, as reflected by his remark, ‘Intellectually I suppose I never doubted that it was a part of my life.’ Thus, in some sense, he was capable of representing himself as having had the experience.

The case of R.B., in conjunction with broader research on memory, suggests that there are critically different kinds of self-representations associated with recollection (e.g. Klein 2010; Klein and Gangi 2010). Typical cases of episodic recollection plausibly implicate both the distinctive ’I’-concept as well as richer representations of the self, including a representation of self as a constellation of traits. The notion of quasi-memory is supposed to undergird a theory of personal identity, but which (if any) of these self-representations is required in order for a candidate quasi-memory to count as being ‘caused in the normal...

---

8 We thank an anonymous referee for pressing this point.
way’? A more specific way to put the point is, do R.B.’s recollections during his un-owned period count as quasi-memories, given that they do not carry the distinctive conceptual self-representation? To answer this question, quasi-memory theorists will need to say more about exactly which aspect of quasi-memory is supposed to constitute personal identity.

3.3 The sense of identity and evidence of personal identity

One of the major challenges facing a theory of personal identity is to explain how there can be identity of self through apparent changes in the self: How can I be the same person as a person who existed forty years ago, given all of the apparent differences between the person now and the person then? This issue has exercised theorists from the earliest days of the debate.

Hume’s reaction to the problem was to maintain that it simply makes no sense to claim that there could be a genuine identity of person across time (Hume 1739). The fact that a person’s psychology is constantly changing entails that there can be no strict identity across time. Reid and Butler drew the opposite conclusion. They maintained that since there is a strict diachronic personal identity, there must be something that really is unchanging. Reid faced the issue quite directly and maintained that memory provides the evidence for a strict identity of self across time. He writes:

My thoughts, and actions, and feelings change every moment — they have no continued, but a successive existence; but that self or I to which they belong is permanent, and has the same relation to all the succeeding thoughts, actions, and feelings, which I call mine … But perhaps it may be said this may all be fancy without reality. How do you know — what evidence have you — that there is such a permanent self which has a claim to all the thoughts, actions, and feelings which you call yours?

To this I answer, that the proper evidence I have of all this is remembrance, I remember that twenty years ago I conversed with such a person; I remember several things that passed in that conversation: my memory testifies, not only that this was done, but that it was done by me who now remember it. If it was done by me, I must have existed at that time, and continued to exist from that time to the present. (Reid 1785, p. 318)

Reid takes his memory of past existence as the evidence that he — the exact same self — did exist in the past. Similarly, Butler writes,

Every person is conscious, that he is now the same person or self he was, as far back as his remembrance reaches; since, when any one reflects upon a past action of his own, he is just as certain of the person who did that
action, namely himself, the person who now reflects on it, as he is certain that the action was at all done. (Butler 1736, p. 295)

Does the sense of personal identity provide evidence of numerical identity? These are large and difficult issues, but future investigations into the matter might draw on the fact that the sense of personal identity is really a byproduct of the episodic memory system. That is just how episodic memory happens to work. The case of R.B. indicates that this sense of numerical personal identity is subserved by a distinctive, contingent, and poorly understood part of memory systems (see also Klein forthcoming a). Knowing the nature of the systems that deliver these judgements might be an important source of information for evaluating the extent to which the sense of identity can be taken to reflect the reality of identity. We leave discussion of those issues for another occasion. For the purposes of this paper, we are satisfied simply to have begun to articulate the systems implicated in the sense of personal identity.  

References


The authors contributed equally to this paper. We would like to thank Sven Bernecker, Carl Craver, John Doris, Brian Fiala, Rachana Kamtekar, David Shoemaker, Galen Strawson, Aaron Zimmerman, two anonymous referees, and the Editor for helpful comments on earlier versions of this paper.


De Biran, Maine 1803: ‘The Influence of Habit on the Faculty of Thinking’, in his 1929.


Furlong, E. J. 1951: *A Study in Memory*. New York, NY: Thomas Nelson and Sons Ltd.


Klein, Stanley and Judith Loftus 1993: ‘The Mental Representation of Trait and Autobiographical Knowledge about the Self’. In Srull and Wyer 1993, pp. 1–49.


—— 1993b: ‘Self-knowledge of an Amnesic Individual is Represented Abstractly’. In Srull and Wyer 1993, pp. 1–49.