

The Sense and Reality of Personal Identity

Thomas Sattig
University of Tuebingen

Published in *Erkenntnis* (online September 2017)

The vast majority of philosophers of personal identity since John Locke have been convinced that the persistence of persons is not grounded in bodily continuity. Why? As numerous ‘textbooks’ on personal identity attest, their conviction rests, to a large extent, on an objection to the bodily approach, which concerns episodic memory. The objection invites us to a thought experiment in which we meet a person who experientially remembers events from the past of a person with a different body. The nature of such first-personal memory-links is viewed as strongly suggesting that the rememberer is identical with the remembered, and hence, given the possibility of such a case, as suggesting that a person can transgress its bodily limits. The memory objection is as influential as it gets in the metaphysics of personal identity. Textbooks often portray it as the starting point of the contemporary debate about personal identity. And it has been widely perceived as a success. As everyone who has taught an introductory course on personal identity knows, the recognition of episodic-memory links in body-switching cases has the power to turn any group of novice students against bodily criteria of personal identity. In this essay, I shall specify and undermine the memory objection. I shall attempt to establish two theses. The first thesis (Sections 1-2) is that the memory objection is only viable if construed as resting on the view that episodic memory contains a sense of personal identity, which teaches us about the reality of personal identity. The second thesis (Sections 3-4) is that there is no such sense of personal identity, that episodic memory teaches us nothing at all about personal identity.

1 The Metaphysics of Personal Identity and the Memory Objection

What grounds a person’s persistence over time? This is the central problem in the metaphysics of personal identity. The contemporary debate about this problem may be parsed as a debate between proponents and opponents of the *bodily approach* to personal identity. According to proponents, a human person’s persistence conditions are biological in nature. The view is that an ordinary person like you and me persists along lines of biological continuity; its beginning and end are the beginning and end of its biological life. The person goes where its body, or its organism, goes. If a person’s temporal boundaries are biological, then a person begins to exist when a foetus begins to exist (or perhaps even when a single cell does?) and it goes out of existence when its biological life-sustaining functions shut down.¹

¹ For an extended discussion of the bodily approach and references, see Olson (1997).

Opponents of the bodily approach deny that the persistence of human persons is grounded in bodily continuity. Most opponents adopt the standard psychological approach to personal identity—now often called *Lockeanism*, as it is arguably the view John Locke proposed in his *Essay Concerning Human Understanding* of 1690—according to which a person’s persistence conditions are psychological in nature. The view is, roughly, that a person persists along lines of psychological continuity; its beginning and end are the beginning and end of its psychological life. There is controversy among Lockeans over which psychological features get to ground personal persistence. The mainstream view seems to be, roughly, that a present person is that future being that can remember experiences of the present person and that inherits a significant range of psychological features from the present person, such as certain beliefs, desires, and character traits. Whichever psychological features are emphasized by different camps, Lockeans agree that if a person’s temporal boundaries are psychological, depending on the continuation of higher mental capacities, then a person begins later than a foetus, and a person may survive biological death.² Other opponents of the bodily approach have been less optimistic about the prospects of grounding personal persistence in psychological facts about persons, adopting a primitivist picture instead, according to which personal persistence is ungrounded.³ All of these alternatives, and numerous others that have not been mentioned here, are united by the rejection of the bodily approach.

Prima facie, it seems quite natural to identify human persons with human animals that are individuated biologically. We often conceive of persons in a third-personal way—from ‘the outside’—as objects instantiating physical properties and as exhibiting behaviour that may be explained and predicted like the properties and behaviour of tables and trees. When persons are conceived of as material objects in this way, it seems rather natural to expect them to have biological persistence conditions.

Yet the majority of philosophers of personal identity since Locke have been convinced that the persistence of persons is not grounded in bodily continuity. Whatever the precise metaphysical ground of personal persistence, if it has a ground at all, it has seemed clear to most philosophers working in this area that persons can transgress their bodily limits. Accordingly, the metaphysics of personal persistence since Locke has been dominated by the task of finding and developing the best alternative to the bodily approach. What motivates this conviction?

To begin with, the conviction is clearly not justified by a direct ‘metaphysical insight’ that one rival or other is superior to the bodily approach. Lockeans will surely agree that it is not just evident that personal persistence is grounded in psychological continuity. If we had a rational insight to this effect, the bodily approach would not have the prima facie appeal that it does in fact possess. It takes some philosophical work to present the anti-body metaphysics of persons in a favourable light.

² Lockeans include Baker (2000), Lewis (1976), Noonan (2003), Nozick (1981), Parfit (1984), and Shoemaker (1963; 1970; 1984; 1997).

³ Primitivists include Butler (1736), Reid (1785), and Merricks (1998).

A massively influential criticism of the bodily approach concerns specific cases of personal persistence in which we intuitively judge a person to persist along a certain path that is not marked by bodily continuity. I shall call them *body-switching cases*. Such cases are, at least at this point, only to be encountered in thought experiments.⁴ Locke's case of the prince and the cobbler is an early instance. A famous modern-day instance is Sydney Shoemaker's (1963) Brown-Robinson-Brownson case. Suppose that Brown's brain is transplanted into Robinson's body. The result is Brownson, who is biologically but not psychologically continuous with Robinson—Brownson and Robinson share the same body—and psychologically but not biologically continuous with Brown—Brownson shares Brown's personality traits, beliefs, desires, and has memories of certain past events, which causally derive from Brown's experiences of those events, and thus link Brownson to Brown. Who is Brownson in this case? That is, how is this case to be described in terms of strict personal identity? Is he identical with Brown or with Robinson or with neither? The view shared by most is that Brownson is identical with Brown. Given that the scenario described by this thought experiment is possible, and given that the bodily approach is characterized by the thesis that persons necessarily follow lines of bodily continuity, the bodily approach faces a counterexample, since Brownson and Brown are one person without being linked by a line of bodily continuity.

What about this sort of body-switching case has compelled most of us to judge that personal persistence transgresses bodily limits? By assumption, Brownson has first-personal memories of things Brown did before the operation, such as worrying about the life-changing events ahead of him, discussing the perils of brain-transplant surgery with his family, and so on. Psychologists call states of this sort *episodic memories*. Opponents of the bodily approach have viewed the existence and nature of such episodic-memory links between Brownson and Brown as providing strong evidence for their view that Brownson is Brown. So we are looking at an objection to the bodily approach from episodic memories in possible body-switching cases. I shall call it the *memory objection*. While not constituting the only avenue of opposition to the bodily approach, the memory objection undoubtedly possesses a distinguished status. I think it is fair to say that considerations of episodic memory in body-switching cases have given rise to the most influential objection to the bodily approach. The objection has won over legions of philosophers, past and present, assuring them that persons are not individuated by their bodies, and sending these philosophers on a quest for alternatives, which quest has dominated metaphysical discussions of personal identity since Locke.⁵

The memory objection is based on a philosophical intuition about episodic memories in body-switching cases. This intuition is a major reason for opposition to the bodily approach. In what follows, two questions will be addressed: What exactly

⁴ If psychological and biological continuity both obtain, it is more or less trivial that personal identity obtains, but it is hard to tell which type of continuity personal persistence (if grounded at all) consists in. Since in most everyday cases there is no divergence between psychological and biological continuity, these are useless as tie-breakers.

⁵ As attested by many textbook-style introductions to personal identity, including Noonan (2003), Olson (2015), and Sider (2007).

is the intuition about? And can the intuition be trusted? In other words, what exactly is it about episodic memories in body-switching cases that justifies the judgment that the rememberer is identical with the remembered? And how strong is this justification? Standard presentations of the memory objection are silent on these questions. The latter cannot be ignored, however. Philosophical intuitions of the present sort are a mere opening gambit. They cannot be upheld in their initial state of innocence as philosophical analysis of their subject matter progresses. We cannot allow our intuitions to be confused or defeated yet still trust them. Naive credulity about intuitions is unpalatable, or so I shall assume.

Let us begin with the first question. In light of the complex profile of episodic memories the memory objection must be focused. That is, opponents of the bodily approach must specify which aspects of episodic memories in body-switching cases their objection is sensitive to. Specifying the memory objection and distinguishing between different versions of it will be my first task.

2 The Memory Objection and the Sense of Personal Identity

As the memory objection concerns certain states of episodic memory, I shall, as a preliminary to specifying this objection, begin with some basic and fairly uncontroversial assumptions about this type of memory. A state of episodic memory is standardly viewed as an experiential state that is caused by an earlier perceptual state. I shall focus on visual memories that causally derive from visual experiences, and I shall adduce three standard representational characteristics of such episodic memories. First, an episodic memory represents a certain event, or scene, in the past. Suppose, for example, that a subject has an episodic memory of a sunrise. The sunrise represented in memory is the same sunrise that was perceived in the earlier state from which the memory causally derives. The memory is a way of re-experiencing that past event. Second, the memory-state represents the sunrise as past, whereas the visual perception represents the same sunrise as present. (The representation of presentness in perception and of pastness in memory is philosophically puzzling but will only play a minor role here.) Third, the memory-state represents the sunrise from a certain visuospatial perspective, which causally derives from the visuospatial perspective of the past perception. Somewhat more precisely, the memory-state represents the qualitative attributes of the sunrise as perspectively formatted: the memory-state represents the sunrise in a way that is governed by lines of sight converging at a single vantage point. The lines of sight governing the representation indicate a subject at the point of the lines' convergence. I shall, for simplicity, call this subject the *centre* of the remembered event. Since the representational scheme of the episodic memory is derived from the representational scheme of the past experience, the represented centre of the remembered event is the same as the represented centre of the perceived event.

So far, I have only described certain representational aspects of visual memories. What is the content of such memories? There is widespread agreement that states of episodic memory have a *propositional content* that can be intuitively characterized as *what* is remembered. There is, on the other hand, a debate about

whether states of episodic memory have an *overall content* that includes but goes beyond their propositional content, containing additional components that concern not what is remembered but rather *how* the propositional content is entertained. These additional components concern the *mode* of the attitude towards propositional content. I shall only speak of representational *components* of the overall contents of states of episodic memory, assuming that any representationally significant property of an episodic memory is encoded in its overall content, but sidestepping the question whether it is encoded in the attitude's propositional content or in the attitude's mode.⁶ Accordingly, I shall assume that the overall contents of episodic memories contain at least the following three components: an *event-component*, representing a certain event as having certain 'tenseless' and non-perspectival qualitative attributes; a *pastness-component* that formats the event component temporally, representing the event as past; and a *perspective-component* that formats the event-component perspectively, representing the event as centred. Finally, I shall call these components in overall memory-content the *basic components*.

With these assumptions in place, let us focus on the case of Brownson and Brown. Suppose that Brownson remembers a certain sunrise. This memory-state represents the sunrise as seen by someone from a certain subjective point of view in the past—in short, Brownson remembers a centred event. This centeredness is the perspective-component modulating the event-component in the overall content of Brownson's memory. Further, as part of the thought experiment, we are assuming that the basic components of Brownson's memory-state are veridical, in that they represent past facts. So the sunrise really happened and was visually perceived by someone in exactly the way Brownson remembers it. Now it turns out that this someone was Brown. Brownson's memory derives from Brown's visual perception per brain-transplant, with the consequence that Brownson remembers the sunrise exactly as seen by Brown.

This scenario will provide the background for the following discussion of the memory objection to the bodily approach. The central question is what about this scenario justifies the belief or intuition that the rememberer, Brownson, is identical with the remembered, Brown. To begin with, I take it to be obvious that the overall content of Brownson's memory-state, as characterized so far, is *representationally neutral* on the question of whether Brownson has memory-access to his own past experience or whether he has memory-access to someone else's past experience.⁷ For the memory-content's correctness is clearly compatible with both options, the first being the option endorsed by the opponent of the bodily approach, and the second being the option endorsed by the proponent. Owing to this neutrality, the fact that Brownson has an episodic memory with this qualitative content *by itself* provides no reason for believing that Brownson remembers his own past experience, and that he is therefore identical with Brown. How, then, should it be possible for an opponent of the bodily approach to reach a clear-headed judgment about personal identity in this

⁶ For discussion of this issue, see, *inter alia*, Recanati (2007: part 5).

⁷ In the case of first-personal cognitive access to someone else's experience in the past, some prefer to speak of *quasi-memory-access*. I do not. More on the notion of quasi-memory in n. 9.

scenario, which is the point of the memory objection? I shall consider two specifications of the memory objection, both of which focus on the role of identity in episodic memory, and I shall argue that only the second version deserves serious consideration.

2.1 First version: Objecting from factual reflexivity

Let me start by introducing a notion of reflexivity in episodic memory. I shall say that an episodic memory is *factually reflexive* (or *factually de se*) just in case the centre of the remembered event is (in fact) identical with the subject of the memory—that is, just in case the events that I remember from the inside are (in fact) events that happened to me.⁸ Note that I shall not assume the concept of episodic memory to be partially defined by factual reflexivity. Factually non-reflexive states of episodic memory, whose subject is distinct from the centre of the represented event, are thus conceptually possible.⁹

The first version of the memory objection focuses on factual reflexivity. Suppose again that Brownson remembers a certain sunrise. The memory-state represents the sunrise as seen by someone from a certain subjective point of view in the past. Assuming, further, as part of the thought experiment, that the basic components of Brownson’s memory-state are veridical, the sunrise really happened and was visually perceived by someone in exactly the way Brownson remembers it, namely, by Brown. As I just pointed out, owing to its representational neutrality with respect to personal identity, Brownson’s cognitive access to Brown’s first-personal point of view on certain events in the past by itself provides no reason for believing that Brownson is identical with Brown. Seeking a basis for this identification, one might reflect on the metaphysics of episodic memory, specifically on how it links individuals at different times, while not relying on any further representational aspects of memory, and one might thus come to hold the belief—by some route or other—that Brownson’s memories are factually reflexive: their subject, namely Brownson, is identical with the centre of the remembered events. Since Brownson has memories of past events as perceived by a subject who was in fact Brown, it follows that Brownson is identical with Brown, which is incompatible with the bodily

⁸ Notice that factual reflexivity in this sense does not settle the question whether the subject of the memory is also identical with the subject of the perception from which the memory derives, as the subject of the perception is to be distinguished from the centre of the perceived event. Assuming, in addition to the factual reflexivity of episodic memory, that in typical cases a perception is also factually reflexive, in that the subject of the perception is identical with the perceived event’s centre, then it follows that episodic memory typically traces a causal link between the memory state of a subject and a past perception of the same subject.

⁹ In the literature on memory and identity, past-directed experiential states lacking factual reflexivity are often called *quasi-memories*—*q-memories*, for short—though this is not the only notion of q-memory in play in familiar discussions (see below for a different notion). (The label ‘quasi-memory’ was introduced by Shoemaker (1970).) I shall have no need for this factual notion of q-memory, as I shall be happy to speak of factually non-reflexive memories.

approach to personal identity, given that the imagined scenario marks a possibility, and that the bodily approach puts forth a claim of necessity. This is the first version of the memory objection, as applied to our sample case.

This version of the memory objection is ineffective. The central assumption, concerning factual reflexivity, takes a stand on the question whether certain states of episodic memory involve personal identity of rememberer and remembered. The main problem with this assumption is that to find out whether certain episodic memories involve such an identity, we must also determine how the subjects of these memories and of the perceptions from which they derive persist. Whether the subject of a memory is identical with the subject of the experience from which the memory derives depends on what personal identity consists in. So there is no way of settling issues of factual reflexivity of memories in independence of any metaphysical account of personal identity. According to the bodily approach, our Brown-Brownson case is one in which a person remembers another person's experiences. In whichever way the factual-reflexivity assumption in the present objection is supported—I deliberately refrained from addressing this question in detail—adopting the assumption requires rejecting a factually non-reflexive understanding of the case, and hence requires rejecting the bodily approach. Since the assumption presupposes an alternative to the bodily approach, the argument begs the question against this approach. Considerations about the factual reflexivity of episodic memories cannot support judgments about personal identity.

2.2 Second version: Objecting from representational reflexivity

In order to get off the ground, the memory objection must focus on a different kind of reflexivity in episodic memory. While factual reflexivity of memory-states concerns the identity of rememberer and centre of remembered event—their identity in reality, one might add—*representational reflexivity* concerns the representation of the identity of rememberer and centre of remembered event in the overall contents of episodic memories. Supposing that I episodically remember a certain sunrise, my memory is representationally reflexive just in case my memory represents the sunrise as seen by someone from a certain subjective point of view, and, in addition, my memory represents this subject as *me*—or, what amounts to the same, my memory represents the past experience as *mine*. That is, in a representationally reflexive state of episodic memory the centre of the remembered event is represented as being identical with the subject of the memory. The overall content of a representationally reflexive memory thus contains a special *identification-component*, which must be distinguished from the basic components.¹⁰

¹⁰ This thesis about the presence of an identification-component in overall memory-content must not be read as presupposing any substantive account of criteria of self-identification, such as the demanding criteria that Evans (1982) associates with identification-components. Specifically, to discern an identification-component in memory is not automatically to hold that this component is separable from all other components. Whether such separability obtains is a further issue, to be addressed below.

Those who understand states of episodic memory as representationally reflexive typically take the identification-component in the overall memory-content to be phenomenally salient. That is, they take the identification-component as grounding a phenomenal *sense of ownership* (or sense of mineness), a sense that the remembered scene is a scene that *I* experienced in the past.¹¹ As the notion of a phenomenal sense of ownership has in the literature not been exclusively applied to a phenomenal aspect of states of episodic memory that is grounded in a representational identification-component, I shall refer to the identity-based sense of ownership in memory as the *sense of personal identity*, which one might want to distinguish from a weaker sense of ownership that is not grounded in representational self-identification.

The second version of the memory objection focuses on representational reflexivity manifesting itself in a sense of personal identity. We suppose, as before, that Brownson remembers a certain sunrise. The memory represents the sunrise as seen by someone from a certain subjective point of view in the past. We also assume, as part of the thought experiment, that the basic components of Brownson's memory-state—that is, the event-component, the perspective-component, and the pastness-component—are veridical, and hence that the sunrise really happened and was visually perceived by someone, namely, by Brown, in exactly the way Brownson remembers it. As we saw, none of this yet provides a good reason for identifying Brownson with Brown. The key move in the second version of the memory objection is to assume, in addition, that Brownson's memory state is characterized by a sense of personal identity that is grounded representational reflexivity. On this assumption, Brownson's memory represents the centre of the remembered sunrise as being identical with the subject of the memory—in other words, Brownson's memory represents the remembered experience as his own experience. Thus, the overall content of Brownson's memory-state contains an identification-component in addition to the basic components. Since the basic components are identification-free, we may assume their veridicality, as we have done above, without assuming the very facts of personal identity that need to be established. By contrast, simply assuming the veridicality of the identification-component, and hence simply assuming that Brownson's memories are factually reflexive, would beg the question against the bodily approach. (Recall the first version of the memory objection.) There is, however, a plausible way of motivating the veridicality of the identification-component without begging the question. It is a widely accepted epistemological principle that we are *prima facie*, though defeasibly, entitled to trust our episodic memories as sources of information about the way the world was, just as we are *prima facie* entitled to trust our perceptions as sources of information about the way the world is. Since the identification-component is part of the overall contents of typical episodic memories, and since we can trust these memories to teach us facts about the past, we can trust the identification-component to teach us facts about personal identity. (This is not the place to demand a defence of this epistemological principle. It is familiar enough to be invoked here without extensive commentary. At

¹¹ I use the label 'sense of ownership/mineness' because it has gained some currency in the memory literature. See, *inter alia*, Klein and Nichols (2012) and Schechtmann (1990).

any rate, I shall not attack the memory objection from this corner.) If, by this epistemological principle, we have a good reason to take Brownson's mnemonic sense of personal identity to track facts of personal identity, then we have a good reason to hold that he is identical with Brown, given that Brownson's memory is a correct representation of a certain sunrise in the past, and given that the centre of the remembered sunrise is in fact Brown. Finally, adding the plausible assumption that the Brown-Brownson scenario is possible, as well as the assumption that the bodily approach puts forth the thesis that human persons necessarily follow lines of bodily continuity, then we end up with a good reason to reject the bodily approach.

I consider this specification of the memory objection as the only version that deserves serious consideration. It rests on the assumption that we possess a sense of personal identity in memory, which teaches us about the reality of personal identity. The objection says, in a nutshell, that in body-switching cases memory teaches us that personal identity extends beyond the limits of bodily persistence. Given that such cases are possible, and given that we can trust the memories involved, we have reason to reject the bodily approach. This account of the memory objection should not come as a big surprise. For the assumption of the representational reflexivity of episodic memories has ranked as orthodoxy for a long time. And many have awarded it a central role in discussions of personal identity. Here, to adduce a well-known example, is Thomas Reid:

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 20 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: 318)¹²

Resting the memory objection on representational reflexivity also helps to explain the argument's remarkable resilience. Critics have been busy pointing out negative theoretical consequences of any body-hostile picture on the market.¹³ Yet none of these attacks seems to have managed to weaken the dominant body-hostile spirit throughout the debate. Part of an explanation of the firmness of the majority's conviction might be that the latter has an unusually convincing source. For here a substantive metaphysical view—that personal persistence is not grounded in bodily continuity—can seemingly be read off fairly directly from intimately familiar episodic memories. When I view the Brown-Brownson case from 'the inside', by taking the point of view of Brownson after the brain-transplant, then I have excellent prima

¹² Speaking of historical precursors, the view that episodic memory is representationally reflexive also seems to have been adopted by Locke (1690), Butler (1736), and Hume (1739).

¹³ Lockean opponents to the bodily approach face, for instance, the well-known problem of fission and the too-many-thinkers problem, though there is the issue of whether the fission problem arises for the bodily approach, as well. See Olson (2015) for references.

facie evidence that I was Brown, because my mnemonic sense of personal identity tells me so. And if philosophers say otherwise, something must be wrong with their reasoning.

Having focused the intuition that drives the memory objection, which shall henceforth be understood along the lines of the second specification, I shall proceed to address the second question posed at the outset: Can this intuition be trusted? I will show that the intuition cannot be trusted, by presenting a range of clear undermining defeaters.¹⁴ To be sure, there may well be other considerations against the bodily approach, which are left unscathed by the considerations to follow. My aim is only to undermine a particular, historically significant objection to the bodily approach, an objection that has become a classic and is widely taken for granted. The target of my criticism will be the assumption of an identification-component in the overall contents of typical states of episodic memory, manifesting itself in a phenomenal sense of personal identity. For ease of presentation, I shall call this assumption *reflexivism*. In what follows, I shall distinguish between two versions of reflexivism, and I shall argue that neither of these versions is able to sustain the memory objection. It is doubtful with respect to the first version of reflexivism that the proposed sense of personal identity, if it exists, teaches us anything about the reality of personal identity. And it is doubtful with respect to the second version of reflexivism that we possess the proposed sense of personal identity in the first place.¹⁵

3 Deflationary Reflexivism

The phenomenal sense of personal identity is grounded in an identification-component in overall memory-content. This is reflexivism. Different versions of reflexivism may be distinguished in response to the following question: What grounds the identification-component? That is, what explains how the centre of a remembered event is represented as being identical with the subject of the memory-state?¹⁶

According to *deflationary reflexivism*, the identification-component in overall memory-content, which manifests itself in the sense of personal identity, is completely grounded in the perspective-component. According to all forms of reflexivism, a subject *S*'s state of episodic memory represents the centre of the remembered event, *C*, as being identical with *S*. According to deflationary reflexivism, *S*'s episodic memory represents the centre of the remembered event, *C*, as being identical with *S*, *in virtue of* representing this event from the subjective

¹⁴ See Pollock (1987) for the notion of an undermining defeater.

¹⁵ Raising the second type of doubt pushes me towards *non-reflexivism*, which is the denial of reflexivism. Non-reflexivists are hard to find in the literature. I read Velleman (1996) as developing a form of non-reflexivism about episodic memory, by recourse to Williams' (1973) discussion of imagining being someone else.

¹⁶ I am here making the now-common assumption that grounding, or constitution, backs non-causal explanations.

perspective of *C*. In virtue of having cognitive access to a subjective perspective on a given event—that is, in virtue of remembering the event ‘from the inside’—the rememberer represents the centre of this perspective as herself. In other words, the representation of an event as centred on a subject completely explains the representation of this subject as being identical with the subject of the memory-state. Self-identification in memory, on this view, has no life of its own; it rests purely on the perspectival formatting of remembered events. It is a direct consequence of deflationary reflexivism that I cannot have cognitive access to a subjective perspective on a given event in the past without also representing the centre of this event as myself.^{17, 18}

According to *robust reflexivism*, as I shall call the reflexivist denial of deflationary reflexivism, the identification-component in overall memory-content, which manifests itself in the sense of personal identity, is not completely grounded in the basic components. The representation of the centre, *C*, of a remembered event as oneself cannot be fully explained in terms of the representation of this event as being centred on *C*, nor in terms of the representation of the event as having a certain non-perspectival qualitative profile, nor in terms of the representation of the event as being past, nor in terms of a combination of these representations. Self-identification in memory, on this view, has a life of its own. My identification of the centre of a remembered event as myself is at least partly independent of the basic components in memory-content.

Notice that the identification-component’s constitutive independence of the basic components, and especially of the perspectival component, allows the robust reflexivist to recognize the possibility for someone to represent a past event from a certain point of view without also representing the centre of this event as herself. As some reflexivists may shy away from calling such identification-free, past-directed experiential states ‘memories’—perhaps because the assumption of an identification-component is taken to be so entrenched in our ordinary use of the term ‘memory’—

¹⁷ I am here making the standard assumption that grounding, and hence metaphysical explanation, implies necessity.

¹⁸ This version of deflationary reflexivism may be called *narrow*, in contrast to *broad* deflationary reflexivism, according to which the identification component in memory-content is completely grounded in the basic components, while the perspectival component plays only a partial or even no explanatory role. I mention the broad version only to set it aside. A serious problem with this alternative is that the event-component and the pastness-component in the overall contents of episodic memories do not seem to be relevant for explaining the identification-component. It is *prima facie* plausible to expect, as the narrow deflationary reflexivist does, that a subject *S*’s representation of a sunrise from the point of view of *C* explains *S*’s representation of *C* as identical with herself. But why should *S*’s representation of a sunrise as having certain non-perspectival qualitative attributes, for example as undergoing a certain change in shape or colour, be relevant for explaining the presence of *S*, as opposed to someone else, in the event? What does the non-perspectival shape of an event have to do with *who* experienced it? Similarly, why should the representation of the sunrise as past be relevant for explaining *who* experienced it? As I cannot see why these aspect should have any explanatory relevance at all, I shall, henceforth, focus on the narrow version when I speak of deflationary reflexivism.

it will be useful to have the notion of a *representational q-memory* ready for this sort of state.¹⁹ The robust reflexivist can then be portrayed as accepting the possibility of representational q-memory, while the deflationary reflexivist denies this possibility.²⁰

I shall now advance two objections to premising the memory objection on deflationary reflexivism, both of which appear avoidable by invoking robust reflexivism instead.

3.1 First objection: The unreliability of the deflationary sense of personal identity

The gist of the second and only viable version of the memory objection is that since memory is representationally reflexive, it tells us something about personal identity over time, and since we are entitled to trust our memories, we can take our memories to teach us facts about personal identity. If representational reflexivity is deflated, however, then we cannot trust our memories anymore. Here is why.

According to deflationary reflexivism, the identification-component in memory-content is completely grounded in the perspective-component: subject *S*'s episodic memory represents the centre of the remembered event, *C*, as being identical with *S*, *in virtue of S*'s memory representing this event from the subjective perspective of *C*. Now recall that, owing to its representational neutrality, the perspective-component in memory-content by itself fails to motivate a denial of the bodily approach. That Brownson has cognitive access to Brown's first-personal point of view in the past says us nothing about whether Brownson is Brown. But if our deflationary sense of personal identity rests entirely on purely non-reflexive, perspectival aspects of episodic memories, which obviously fail to provide information about facts of personal identity, then the deflationary sense cannot be trusted to deliver information about facts of personal identity either. As a consequence, the deflationary sense of personal identity cannot sustain the memory objection.

To put the worry another way, the assumption of the veridicality of the identification-component, employed in the memory objection, was motivated by the principle that we are defeasibly justified in taking our memories to teach us facts about the past. But now it turns out that the identification-component in episodic memories entirely derives from the perspective-component, and that the latter fails to teach us anything about who we were. This deflationary account of the nature of

¹⁹ While factual q-memory is free of factual reflexivity, representational q-memory is free of representational reflexivity. Unfortunately, discussions of q-memory are occasionally unclear on whether failure of factual reflexivity or of representational reflexivity or of both are under consideration. I shall focus on the latter sort.

²⁰ While my way of distinguishing between deflationary and robust reflexivism is not a standard one, I read Parfit (1984) and Klein & Nichols (2012), *inter alia*, as robust reflexivists, and I read Schechtman (1990), *inter alia*, as a deflationary reflexivist. The issue whether there is a distinctive phenomenal sense of ownership in memory is in many ways similar to the issue whether there is a distinctive phenomenal sense of bodily ownership. See, *inter alia*, the recent exchange between Bermúdez (2013, 2015) and de Vignemont (2013).

the identification-component is a clear defeater of our prima facie trust in our representationally reflexive memories as corresponding to facts. Hence, the deflationary account of representational reflexivity undermines the veridicality assumption about self-identification in the memory objection.²¹

Premising the memory objection on robust instead of deflationary reflexivism points to a way out of this predicament. To reiterate, while our mnemonic sense of personal identity is supposed to teach us about the reality of personal identity, this sense is unreliable if grounded entirely in mere representational perspectivity. If, on the other hand, the sense of personal identity is viewed as somehow going above and beyond the mere first-personal representation of events in episodic memories, then the unformativeness of representational perspectivity regarding the facts of personal identity does not immediately undermine our prima facie trust in our mnemonic sense of personal identity.

3.2 Second objection: Deflationary reflexivism and the case of R.B.

There is a further reason why the memory objection should not be premised on deflationary reflexivism. Deflationary reflexivism faces countervailing empirical evidence, as it clashes with a recent neurological case described by Stanley Klein and Shaun Nichols (2012). Patient R.B. suffers from a peculiar form of memory dissociation. As a result of head trauma following an accident, R.B. seemed, during a certain period, to have experiential memories of events from his past without describing these events as having been experienced by himself. As Klein and Nichols present the case, R.B. was able to recall particular incidents from his life in way that strongly suggests that these memories are episodic, as opposed to semantic. For he was able to remember events as experienced from a certain subjective point of view. Yet he had the impression that these past events were not experienced by him. For instance, following a report of memories from his childhood, R.B. comments:

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. (Klein and Nichols 2012: 686)

²¹ While I have directed this reliability objection against narrow deflationary reflexivism, it also applies to broad deflationary reflexivism. For if the identification component rests entirely on the basic, non-reflexive components in memory-content—no matter which exact explanatory role the perspective-component plays here—then the sense of personal identity cannot be trusted to deliver information about facts of personal identity, since the basic components obviously fail to provide information about facts of personal identity. We saw that Brownson's representation of an event from Brown's first-personal point of view is neutral on whether Brownson is Brown. The representation of an event as having a certain (tenseless and non-perspectival) qualitative profile, and the representation of an event as past are similarly neutral on questions of personal identity. We do not learn anything about the reality of personal identity on this basis.

R.B. gives a similar characterization of memories from his time in graduate school:

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. (Klein and Nichols 2012: 686)

R.B. can be described as having mental states that causally derive from past visual perceptions of his own. The overall contents of these states, unlike the contents of semantic memories, contain a perspective-component. Moreover, the overall contents of these states, unlike the contents of perceptions, contain a pastness-component. However, the overall contents of these states do not seem to contain an identification-component: ‘It did not feel like it was something that really had been a part of my life’. So R.B. seems to have what a reflexivist who holds that proper memories come with an identification-component would call representational q-memories.²²

What does this case tell us about reflexivism? The most straightforward consequence of the case seems to be that the identification-component in the contents of typical states of episodic memory, if there is such a component, is not completely grounded in the perspective-component. For if the identification-component were completely grounded in the perspective-component, then R.B.’s past-directed states, whose overall contents contain a perspective-component, would represent the centre of the remembered events as being identical with the subject of these states—that is, they would represent R.B. as having been there. And so the overall contents of his states would contain the same identification-component to be found in typical memories. This, however, does not seem to be the case. R.B.’s states have a perspective-component but lack an identification-component. His states represent the qualitative attributes of certain past events as perspectively formatted, as governed by lines of sight that indicate a subject at the point of the lines’ convergence, without representing this subject as himself. Thus we seem to have a counterexample to deflationary reflexivism. The objection may be summarized thus: while the deflationary reflexivist denies the possibility of representational q-memory, R.B. seems to have had q-memories of precisely this sort.

Robust reflexivism seems less threatened by the case of R.B., because the robust variant allows for an account of the case in terms of an impaired robust sense of personal identity. Very roughly, the account is that while we typically have a distinctive sense of personal identity grounded in a positive identity-representation, this sense is impaired in R.B.’s case. R.B. has the impression that someone else experienced those events in the past, and hence his sense of personal identity fails to

²² Since R.B.’s states are still factually reflexive, he does not have factual q-memories.

correspond to a positive identity-representation, corresponding to a negative one instead. R.B. thus ends up with representationally non-reflexive q-memories instead of full-blown, representationally reflexive memories. While this case raises further important issues about memory and representational reflexivity, I shall, for present purposes, rest content with pointing out that robust reflexivism seems in a better position to handle it than its deflationary rival.

4 Robust Reflexivism

The foregoing considerations show that the memory objection holds little promise if premised on deflationary reflexivism. On the face of it, the memory objection is in better shape when premised on robust reflexivism instead. On closer inspection, however, going robust will not do. Recall that according to the second and only viable version of the memory objection, episodic memory tells us something about personal identity over time, because memory is representationally reflexive. And since we are entitled to trust our memories, we can take them to teach us facts about personal identity. The main problem for robust reflexivism is that if there were a reliable robust sense of personal identity, meeting the demands of the memory objection, then this sense would be truly magical.

In order to see this, suppose that I am endowed with a non-deficient capacity of episodic memory, and that I have a memory of a sunrise in scenario c_1 , where I am also the centre of the remembered event, which means that the memory is factually reflexive. Now consider a different scenario, c_2 , in which I have a memory of a sunrise that is indistinguishable with respect to any aspects of the perspectively formatted qualitative content from my memory in the first scenario. In the two scenarios the sunrise is represented in exactly the same qualitative, first-personal way. Suppose, however, that in c_2 I am not the centre of the remembered event, but someone else is, and hence that my memory is factually non-reflexive.

Let us now consider a version of robust reflexivism, according to which a subject S 's episodic memory represents the centre of the remembered event as being identical with S in a way that cannot be explained at all in terms of the memory's perspectively formatted qualitative content. According to this version, my identification of the centre of a remembered event as myself completely 'transcends' the basic components in the overall contents of states of episodic memory. Accordingly, if my identification of the centre of a remembered event as myself is reliable, as demanded by the memory objection, then I should be able to determine which of the remembered events in the sketched case was experienced by me. That is, I should be able to determine in which of the two scenarios the centre of the past event is identical with the subject of the memory: the centre of one event should feel like me, whereas the centre of the other should not, while this feeling of who is the centre of the remembered event is constitutionally independent of any aspects of the perspectively formatted qualitative memory-content. I find it very hard to believe that we possess such a mysterious ability of detecting our own presence in memory.

In response, one might wonder whether such cases of factual non-reflexivity are even possible. If they are not, is the worry alleviated? No. For present purposes,

it is irrelevant exactly how the factually non-reflexive memory in c_2 came about, nor is it relevant whether this case is really possible. The case is merely invoked to illustrate how a reliable transcendent sense of personal identity is supposed to function. This sense is designed in a way that it would determine when I have memory-access to my own experiences and when to someone else's, in a qualitatively indistinguishable pair of scenarios differing only with respect to factual reflexivity, if the pair were possible—pretend that it is, and observe the self-detector at work. My point is that it is highly implausible to suppose that our episodic memories are characterized by a sense with such a function.

A second version of robust reflexivism carves out a somewhat weaker alternative: the identification of the centre of a remembered event as myself relies in part on certain aspects of the perspectively formatted qualitative memory-content and in part on additional aspects. So the internal self-detector is only partly transcendent. This version of robust reflexivism obviously does not escape the mystery objection to the first version. Partial transcendence is just as implausible as complete transcendence, as it would still allow me to determine which of the remembered events was experienced by me in the two qualitatively indiscernible scenarios, c_1 and c_2 , sketched above.

The remaining reflexivist option is to concede that the robust sense of personal identity is unable to determine which of the remembered events was experienced by me in the two qualitatively indiscernible scenarios, while holding that this is no counterexample to our possessing such a sense, if in ordinary circumstances my mnemonic representation of the centre of a remembered event as being me does correspond to the facts. In other words, under normal conditions our robust sense of personal identity works fine, but in exotic circumstances it breaks down.

The main deficiency of this alternative understanding of robust reflexivism is that it renders the sense of personal identity unreliable, and hence unsuited to play the role required by the memory objection. For if the sense of personal identity were reliable, then, owing to its constitutional independence of the perspectively formatted qualitative memory-content, it would be able to self-detect in qualitatively indistinguishable scenarios. Given the robust nature of the sense of personal identity, the task of distinguishing such scenarios would not be an extraordinary task for it. The task would be business as usual. Since the sense breaks down in these cases, it cannot be trusted to teach us facts about personal identity.

Let me conclude. According to the massively influential memory objection, reflection on episodic memories in body-switching cases strongly suggests that personal identity extends beyond the limits of bodily persistence, and hence that the bodily approach to personal identity is false. In this essay, I have argued for two theses. The first thesis is that the memory objection is only viable if construed as resting on the view that episodic memory contains a sense of personal identity, which teaches us about the reality of personal identity. The second thesis is that there is no such sense of personal identity, that episodic memory teaches us nothing at all about personal identity. The outcome of my considerations is that the memory objection is

misguided. A confusion about memory has put legions of metaphysicians of personal identity on the wrong track.²³

References

- Baker, L. R. 2000: *Persons and Bodies*. Cambridge: Cambridge University Press.
- Bermúdez, J. L. 2015: ‘Bodily Ownership, Bodily Awareness, and Knowledge without Observation’, *Analysis*, 75: 37-45.
- 2013: ‘Bodily Awareness and Self-Consciousness’, in S. Gallagher (ed.), *Oxford Handbook of the Self*. Oxford: Oxford University Press.
- Butler, J., 1736, *The Analogy of Religion, Natural and Revealed, to the Constitution and Course of Nature*, London: J. and P. Knapton, 2nd corrected edition.
- de Vignemont, F. 2013: ‘The Mark of Bodily Ownership’, *Analysis*, 73: 643-51.
- Evans, G. 1982: *Varieties of Reference*. Oxford: Oxford University Press.
- Hume, David 1739: *A Treatise of Human Nature*, ed. David Fate Norton and Mary J. Norton. Oxford: Oxford University Press (2000).
- Klein, S. B. and Nichols, S. 2012: ‘Memory and the Sense of Personal Identity’, *Mind*, 121: 677-702.
- Lewis, D. 1976: ‘Survival and Identity’, in A. Rorty (ed.), *The Identities of Persons*, Berkeley, CA: University of California Press; reprinted in his *Philosophical Papers* vol. I, New York: Oxford University Press (1983).
- Locke, J., 1690: *An Essay Concerning Human Understanding*, ed. P. Nidditch. Oxford: Clarendon Press (1975).
- Merricks, T. 1998: ‘There Are No Criteria of Identity Over Time’, *Noûs*, 32: 106–124.
- Noonan, 2003: *Personal Identity* (2nd edition). London: Routledge.
- Nozick, R. 1981: *Philosophical Explanations*. Cambridge, MA: Harvard University Press.
- Olson, E. 2015: ‘Personal Identity’, SEP.
- 1997. *The Human Animal: Personal Identity Without Psychology*, New York: Oxford University Press.
- Parfit, D. 1984: *Reasons and Persons*. Oxford: Oxford University Press.
- Pollock, J. 1987: ‘Defeasible Reasoning’, *Cognitive Science*, 11: 481-518.
- Recanati, F. 2007: *Perspectival Thought: A Plea for (Moderate) Relativism*. Oxford: Oxford University Press.
- Reid, Thomas 1785: *Essays on the Intellectual Powers of Man*. Edinburgh: Bell & Robinson.
- Shoemaker, S. 1997: ‘Self and Substance’, in *Philosophical Perspectives* (Volume 11), J. Tomberlin (ed.): 283–319.

²³ For comments on the material presented in this paper I am indebted to Adrian Alsmith, Sven Bernecker, Andrea Bottani, Valerio Buonomo, Carl Craver, Richard Davies, Bahadır Eker, Brendan de Kenessey, Jean Moritz Mueller, Krisztina Orban, Alfredo Paternoster, François Recanati, Katia Samoiloova, Alfredo Tomasetta, Hong Yu Wong, and audiences at the University of Bergamo, the University of Milan, the Max Planck Institute for Biological Cybernetics, Tuebingen, and at GAP.9 at Osnabrueck University.

- 1984: 'Personal Identity: A Materialist's Account', in Shoemaker and Swinburne, *Personal Identity*, Oxford: Blackwell.
- 1970: 'Persons and Their Pasts', *American Philosophical Quarterly*, 7: 269-85.
- 1963: *Self-Knowledge and Self-Identity*. Ithaca: Cornell University Press.
- Sider T. 2007: 'Personal Identity', in T. Sider & E. Conee, *Riddles of Existence*, Oxford: Oxford University Press, 7-21.
- Velleman, J. D. 1996: 'Self to Self', in *Philosophical Review*, 105: 39-76.
- Williams, B. 1973: 'The Imagination and the Self', in *Problems of the Self*, Cambridge: Cambridge University Press.