Memory-Based Personal Identity Without Circularity

Thomas Sattig
University of Tuebingen

Published in Valerio Buonomo (ed.), *The Persistence of Persons. Studies in the Metaphysics of Personal Identity over Time* (2018), Neunkirchen-Seelscheid: Editiones Scholasticae.

1 A Circularity Objection to Lockeanism

What are the criteria of personal identity over time? In other words, what grounds the persistence of persons? The standard psychological approach to personal identity was arguably first proposed by John Locke in his *Essay Concerning Human Understanding* of 1690. According to *Lockeanism*, as the approach has come to be called, a person's persistence conditions are psychological in nature: a person persists along lines of psychological continuity; its beginning and end are the beginning and end of its psychological life. While there is controversy among Lockeans over exactly which psychological features get to ground personal persistence, Lockeans seem to agree that veridical states of episodic memory play a central role. The core picture seems to be that a present person is that future being that can veridically remember experiences of the present person and that inherits a significant range of psychological features from the latter, such as certain beliefs, desires, and character traits.[1]

A standard objection to Lockeanism is that veridical episodic memories cannot (partially) ground personal identity, because the veridicality of episodic memories is itself grounded in personal identity. The main question of this essay is how Lockeans can avoid the latter type of circularity. Following an introduction to the circularity objection, I shall criticize a recent response to it (Section 2) and then develop a new and superior response (Section 3).

What is episodic memory? Let us begin with some basic assumptions. 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 components in the contents of episodic memories, which I shall call the *basic components* of memory-content. First, there is the *event component*. An episodic memory represents a certain event, or scene, in the past. Suppose, for example, that I have an episodic memory of Hurricane Katrina. The storm represented in memory is the same storm that I perceived in the earlier state from which the memory causally derives. The memory is a way of re-experiencing that past event. Second, there is the *pastness component*. The memory-state

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

represents the storm as past, whereas the visual perception represents the same storm as present. There is, third, the *perspective component*. My memory-state represents the storm from a certain visuospatial perspective, which causally derives from the visuospatial perspective of my past perception. Somewhat more precisely, the memory-state represents the qualitative attributes of the storm as perspectivally formatted: the memory-state represents the storm 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.

What is the role of (numerical) identity in the contents of episodic memories? Supposing that I remember Hurricane Katrina, my memory is *representationally reflexive* (or representationally *de se*) just in case my memory does not merely represent the storm as seen by someone from a certain subjective point of view, but my memory represents the storm as seen by *me*. 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 content of a representationally reflexive memory thus contains a special *de se component*, which must be distinguished from the basic components characterized above.[2]

The view that our episodic memories are representationally reflexive, which I shall henceforth refer to as *reflexivism*, has a long and illustrious tradition. The view seems to have been adopted, *inter alia*, by Butler (1736), Reid (1785), and Hume (1739). Reid (1785: 318), for instance, claims: 'My memory testifies, not only that this was done but that it was done by me who now remember it.' Reflexivism also finds acceptance in contemporary philosophy and psychology of memory. Here many find it plausible that typical episodic memories contain a phenomenal *sense of personal identity*, also known as a *sense of ownership* or *mineness*, which is the sense that the remembered scene is a scene that *I* experienced in the past, and which is grounded by a *de se* component in memory-content.[3]

If the view that episodic memories are representationally reflexive is combined with Lockeanism in the metaphysics of personal identity, yielding *reflexivist Lockeanism*, then Lockeans face the following *circularity objection*. They hold that the ground of personal persistence is psychological continuity, and that episodic memory is an essential element of the required kind of psychological continuity. This is the Lockeans' signature claim. Suppose, for a simple instance of the Lockean scheme, that the fact that a person *P*, which exists presently, existed at an earlier time *t*, partly

---

[2] By contrast, an episodic memory may be called *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.

[3] See, *inter alia*, Klein and Nichols (2012) and Schechtmann (1990). 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).

consists in the fact that *P* currently has an episodic memory of an event at *t*, and partly consists in the fact that this memory is a veridical memory—that is, a memory whose content corresponds to the facts. The veridicality of the memory is a partial ground because non-veridical, merely apparent memories are cheap and unsuited for grounding personal persistence. What makes the memory veridical? What is its veridicality-maker? By the assumption of representational reflexivity, *P* remembers the centre of that event as herself. Thus, the veridicality of the memory is partly grounded in the obtaining of certain facts in the world, including the fact that *P* was the centre of the event at *t*, and hence existed at *t*. But then *P*'s existence at *t* partly grounds *P*'s existence at *t*, which few will accept, as it is standard to understand metaphysical explanation and grounding as irreflexive. The problem, in short, is that a reflexivist Lockean's metaphysical explanation of personal identity ends up running in a circle.[4]

In what follows, I shall explore two strategies for saving Lockeanism from this circularity objection. These strategies rest on different views about representational reflexivity in episodic memory. I shall criticize the first, more familiar strategy and advertize the second, less familiar one.

2 Avoiding Circularity I: Robust Reflexivism

Is it possible for Lockeans to avoid circularity while sticking to reflexivism? In order to understand how conciliation might be achieved, different versions of reflexivism may be distinguished in response to the following question: What grounds the *de se* component in memory-content? That is, what explains how the centre of a remembered event is represented as being identical with the subject of the memory-state?[5]

According to the view I shall call *deflationary reflexivism*, the *de se* component in 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 herself. According to deflationary reflexivism, *S*'s episodic memory represents *C* as being herself, *in virtue of* representing this event from the subjective 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

[4] This may not be the only memory-related circularity problem for Lockeanism. There may also be a problem concerning factual reflexivity that is independent of representational reflexivity. See Shoemaker (1970) and Parfit (1984). Discussions of circularity in the metaphysics of personal identity are not always clear on which notion of reflexivity is meant to be the source of the trouble. Here I am only concerned with straightforward representational-reflexivity-based circularity of the sort first pointed out by Butler (1736) and Reid (1785). As Butler famously puts the objection: 'And one should really think it self-evident, that consciousness of personal identity presupposes, and therefore cannot constitute, personal identity, any more than knowledge, in any other case, can constitute truth, which it presupposes' (Butler 1736: Dissertation I; Perry 1975: 100).

[5] The following distinctions derive from Sattig (2017).

'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. Since deflationary reflexivism makes the *de se* component an essential component of memory-content, 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.[6]

According to *robust reflexivism*, as I shall call the reflexivist denial of deflationary reflexivism, the *de se* component in 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 representation of the centre of a remembered event as myself is at least partly independent of the basic components in memory-content.

Owing to the *de se* component's constitutive independence of the basic components, and especially of the perspective component, the robust reflexivist may 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', it will be useful to apply the notion of a *representational quasi-memory*, or *q-memory*, to this sort of state.[7] The robust reflexivist can then be portrayed as accepting the possibility of representational q-memory, while the deflationary reflexivist denies this possibility.

Now back to the circularity objection. Given the distinction between deflationary and robust reflexivism, it is a natural move for Lockeans to hold on to the prima facie attractive view that typical episodic memories are representationally reflexive, while allowing for the possibility of non-reflexive representational q-memories, which may be invoked as partial grounds of personal identity without generating ground loops. (Here there is no need for details on how Lockeans could explain personal identity with such q-memories.) The key in avoiding the objection, then, is to appeal to robust reflexivism about episodic memory. If deflationary reflexivism is assumed instead, then the *de se* component is inseparable from the contents of episodic memories, and hence it is hard to see how loops in grounding personal identity can be avoided. For the deflationary reflexivist does not recognize

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

[7] 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. The notion of quasi-memory was introduced by Shoemaker (1970).

an episodic-memory-like state that is identification-neutral. Any perspectivally formatted way of re-experiencing a past event automatically represents the centre of that event as identical with the subject of the past-directed state. Robust reflexivism, by contrast, does recognize an episodic-memory-like state that is identification-neutral.

Stanley Klein and Shaun Nichols (2012) have recently supported this type of move by appeal to the neurological case of patient R.B. who suffered from an unusual 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. Here is how R.B. characterizes memories from his childhood and from his time in graduate school, respectively:

> 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)

> 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)[8]

R.B. can be described as having mental states that causally derive from past visual perceptions of his own. The contents of these states, unlike the contents of semantic memories, contain a perspective component. Moreover, the contents of his states, unlike the contents of perceptions, contain a pastness component. However, the contents of these states do not seem to contain a *de se* 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 a *de se* component, would call representational q-memories.[9]

---

[8] Note that while R.B. only failed to self-ascribe remembered experiencies that occurred in the time period preceding his injury, he did not suffer such an impairment in remembering experiences that occured after the accident.

[9] Since R.B.'s states are still factually reflexive, he does not have factual q-memories.

Following Klein and Nichols, I consider the case of R.B. highly relevant for issues concerning reflexivism about episodic memory. However, as this case is currently the only one of its kind discussed in the literature, the somewhat speculative nature of the following discussion cannot be denied. With this methodological caveat in mind, let us ask what the case of R.B. tells us about reflexivism. The most straightforward consequence of the case seems to be that the *de se* 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 *de se* 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 contents of his states would contain the same *de se* 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 a *de se* component. His states represent the qualitative attributes of certain past events as perspectivally 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.

There is more to the case, though. For not only can it be invoked to undermine deflationary reflexivism but also to support robust reflexivism. So far, we have observed that R.B. remembers past events without placing himself at their centre. But that is not all. For R.B. also has the impression that the remembered events were experienced by someone else. (Note that while he seems compelled to the hypothesis that he has cognitive access to someone else's point of view, R.B. does not accept this hypothesis, on the grounds of countervailing evidence.) How should this impression of non-identity, or of *disownership*, be explained?

Here is one explanatory strategy—call it the *ownership account*. Something is missing from R.B.'s episodic memories when compared to typical ones. What is missing is indicated by his impression that what he remembers are not events experienced by himself. One might thus hold that what is missing in R.B.'s memories, but present in typical episodic memories, is a fully functional sense of personal identity. Some impairment in this sense, which is responsible for positive *de se* representations in typical cases, yields R.B.'s negative *de se* representation, his disownership-impression. Since this sense of personal identity is not completely grounded in the perspective component in memory-content, or else it would yield the same *de se* representation in R.B.'s case as it does in typical cases, the ownership account of R.B.'s deficit is committed to robust reflexivism.

I interpret Klein and Nichols as subscribing to the ownership account of the case of R.B, in terms of what I call robust reflexivism. Here is how they outline their view:

> 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. (Klein & Nichols 2012: 689)

Turning to the question of the functioning of the sense of personal identity, Klein and Nichols construe it as 'a special kind of conceptual self-representation' (Klein & Nichols 2012: 690), as a 'specialized neural machinery that inserts the conceptual element *self* into the agent slot of an episodic memory attribution' (Klein & Nichols 2012: 689). And this specialized machinery was compromised by R.B.'s injury, though it was not destroyed, as indicated by his ability to recover the sense of personal identity several months after his injury, and by the availability of his sense of personal identity in episodic memories of experiences that were made after the accident.[10]

We saw that by appeal to robust reflexivism, and hence to representational q-memories, Lockeans can avoid the circularity objection. Klein and Nichols (2012: 690-5) recognize this strategy, and they recommend it as being directly supported by the case of R.B.

The sketched response to the circularity objection will not do, however, since robust reflexivism is highly implausible.[11] My worry is that if there were a robust sense of personal identity, then this sense would be 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 hurricane 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 hurricane that is indistinguishable with respect to any aspects of the perspectivally formatted qualitative content from my memory in the first scenario. In the two scenarios the hurricane 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 episodic memory represents the centre of the remembered event as herself, in a way that cannot be explained at all in terms of the memory's perspectivally formatted qualitative content. On this version, my representation of the centre of a remembered event as myself completely 'transcends' the basic components in the contents of states of episodic memory. Given that my sense of personal identity has this function of a transcendent self-detector, 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

[10] See Klein (2012: 491-4).

[11] The following criticism of robust reflexivism also appears in Sattig (2017), though the context is a different one.

perspectivally 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 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 offers a weaker alternative: the identification of the centre of a remembered event as myself relies in part on certain aspects of the perspectivally 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 the sense of personal identity would still be expected to determine which of the remembered events was experienced by me in the two qualitatively indiscernible scenarios, c_1 and c_2, sketched above.

Another response might be 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 denying that this is a counterexample to our possessing such a sense. The case of the indiscernible scenarios, the robust reflexivist might hold, is just an exotic circumstance. And that our sense of personal identity breaks down in exotic circumstances is nothing to worry about, as long as it works fine under normal conditions, in that in ordinary circumstances my mnemic representation of the centre of a remembered event as being me does correspond to the facts.

I reply that the case of the indiscernible scenarios is far from exotic relative to the expected function of the sense of personal identity. Given the robust nature of the latter sense, the task of distinguishing such scenarios would not be an extraordinary task for it. The task would be just what the sense is 'designed for'. That is, owing to its constitutional independence of the perspectivally formatted qualitative memory-content, the sense of personal identity, if we had one, would be expected to be able to self-detect in qualitatively indistinguishable scenarios. Since the sense would break down in these cases, it would not be able to perform its representational function in 'normal' circumstances. It is therefore unlikely that we possess such a sense.[12]

The mystery charge undermines the reflexivist Lockean's robust response to the circularity objection. The contents of episodic memories should not be viewed as

[12] An interesting follow-up issue is how alleged *de se* components in memory are related to alleged *de se* components in perception, as advocated, *inter alia*, by Peacocke (2014).

containing a robust *de se* component. Construing memory-contents as containing a deflationary *de se* component, however, leads Lockeans straight into circularity. Moreover, and here I agree with Klein and Nichols, the case of R.B. speaks against deflationary reflexivism. These conclusions suggest questioning reflexivism about episodic memory altogether.

3 Avoiding Circularity II: Non-Reflexivism

Non-reflexivism, as I shall use the term, is the view that the contents of typical states of episodic memory are not representationally reflexive, and accordingly do not come with a phenomenal sense of personal identity (while typical states of episodic memory are still assumed to be factually reflexive). The memory-contents are here construed as neutral on whether the remembered event was experienced by the rememberer. According to this position, when I remember Hurricane Katrina, my memory represents the storm as seen by someone from a certain subjective point of view, without also representing this point of view as mine, and hence without representing the centre of the remembered event as being me. That is, the content of an episodic memory does not contain a *de se* component. To be sure, the memory does indicate the centre of the remembered event: the center is the subject occupying the memory image's vantage point and undergoing the visual impression from which the memory-image is derived. The point is that this subject is not also represented as being identical with the subject of the memory.[13,14]

Combining Lockeanism with non-reflexivism straightforwardly avoids the circularity objection. Veridical episodic memories may be appealed to in grounding personal persistence without apparent circularity, as non-reflexivist memory states' veridicality is not grounded in personal persistence in the way deflationary-reflexivist memory states' veridicality is. In light of our critical discussion of deflationary and robust reflexivism, this non-reflexivist response to the circularity objection is to be taken very seriously. Deflationary reflexivism seems to have a counterexample in the case of R.B., while robust reflexivism is unacceptable for a priori reasons.

---

[13] Non-reflexivists are rare 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.

[14] Let me emphasize that reflexivism and non-reflexivism, as understood here, concern the presence or absence of a *de se* component in the contents of states of episodic memory, while they do not concern the presence or absence of a *de se* component in the contents of *beliefs* about the past, which are based on such memories. There is no question that we have self-identifying beliefs that derive from episodic memories, to the effect that the centre of the remembered event is believed to be identical with the subject of the memory. For example, on the basis of my memory of Hurricane Katrina in the past, I believe that this storm happened around me. What is controversial, and what I am focusing on here, is whether there is a *de se* component in the contents of states of episodic memory, manifesting a sense of personal identity, which is *independent* of representations of identity at the doxastic level. The interesting further question as to how episodic memories yield certain beliefs *de se* is not a concern of mine here.

However, non-reflexivism faces a serious challenge, as well. If our episodic memories are representationally non-reflexive, then what is missing from R.B.'s episodic memories when compared to typical ones? How can R.B.'s impression that the remembered events were experienced by someone else be explained, if not by a defect in the sense of personal identity? I call this the *disownership-challenge* to non-reflexivism, to which I shall respond in the remainder of this paper.

For the purpose of explaining R.B.'s disownership-impression in a way that does not invoke a sense of personal identity, I propose to take a widely accepted account of the Capgras syndrome, a well-known delusion of identity, as a guide. Capgras patients believe that people who live with them and with whom they had a close emotional tie, such as a spouse, have been replaced by impostors. While realizing that the person they see resembles their spouse exactly, they still deny that the person is their spouse. Given that the patient's ability to recognize face-similarity is intact, the defect is not a purely visual one. The syndrome rather shows that face-recognition cannot just consist in a visual-similarity detector.

One might think that what is missing is a separate criterion of 'spouse-identification', an internal 'spouse-detector', which combines with the face-similarity detector to yield full-blown face-recognition. The delusion in Capgras syndrom might then be explained by invoking this spouse-detector: in typical cases, the detector yields identity-verdicts, whereas in Capgras cases it yields non-identity-verdicts, such that, as Bayne and Pacherie put it, 'it is part of the representational content of a Capgras patient's visual perception that "This is someone who looks just like my close relative but is not really him/her"' (Bayne and Pacherie 2004: 4). This approach is called the 'expression' or 'endorsement' approach in the literature.[15] I mention this view only to set it aside.

The standard view among cognitive neuroscientists is a different one. According to the approach to the Capgras delusion known as the 'explanation' approach, the delusion represents an explanation of an anomalous experience.[16] Here is how Max Coltheart summarizes the account:

> According to the explanation account put very generally, in any case of monothematic delusion a state of the world has arisen for which the patient has to find an explanation, and the delusional belief provides such an explanation: that is, if the delusional belief were true, then it would follow that the world would be the way it now seems to be to the patient. In this sense, the belief explains why the world is as it now seems to be. Thus, in the specific case of Capgras delusion, the new state of the world that has arisen is that there is a person in the world who looks exactly like one's wife but when that person is seen there is little or no arousal of one's autonomic nervous system—the degree of response is that which is characteristic of observing a stranger. This datum is just what one would expect to observe if the person

[15] See Fine et al. (2005) for discussion and references.

[16] The model goes back to Ellis and Young (1990). For discussion and references, see Fine et al. (2005) and Coltheart (2005).

being seen is indeed a stranger (despite her physical resemblance to one's wife). (Coltheart 2005: 153)

It is now widely agreed that the discrepancy between expected and actual emotional response is not sufficient to explain the Capgras delusion. As Fine et al. (2005: 144-5) put the issue, 'the perceptually odd experience of emotional hyporesponsiveness in the presence of someone emotionally salient to the patient might suggest the hypothesis that that person has been replaced by an impostor. But, it is argued, the model does not explain why the patient would accept a belief that is almost certainly false.' This issue concerning belief-acceptance will not concern me here. The aspect of the explanation approach that is central for present purposes is that a Capgras patient's non-identity hypothesis is generated by a feeling of unfamiliarity. If so, the syndrome provides no support for the view that face-recognition involves an identification-criterion—an internal spouse-detector—that is impaired in Capgras patients. There is no ground for holding that in typical face-recognition, loosely put, the face-similarity detector registers a match and in addition the spouse-detector does. Instead, the delusional non-identity hypothesis is generated by the absence of an affective response, which response is itself identification-free. The affective response, that is, does not constitute a special criterion of determining *who* is the person with given facial properties. It is merely a reaction to seeing a face with such-and-such features (whoever's face it may be). The delusional hypothesis of non-identity—concerning *who*'s face it is—results from the subject's explanation of the oddity of the absence of the expected affective response.[17]

Now back to R.B. I have suggested that episodic memories do not contain a sense of personal identity. As a consequence, I am facing the challenge of explaining what is missing from R.B.'s episodic memories when compared with typical ones. R.B. has the impression that what he remembers are not events experienced by himself. His episodic memories drive him to the hypothesis that the centre of the remembered events is not identical with the subject of the memories. How can this non-identity, or disownership, hypothesis be explained? In light of the standard account of the Capgras syndrome (and other delusions) reviewed above, I want to speculate that the schematic 'explanation' approach partly extends to R.B.'s case. Here is a rough sketch of how the account might go.

In episodic memory a subject recalls a certain past event from a subjective perspective—she remembers the event from 'the inside'—though the memory does not respresent the centre of the past event as being identical with the subject of the memory. Typical episodic memories cause certain affective responses. We might speak of *feelings of familiarity* here. For the purpose of a mere outline, I shall allow myself to remain vague about their nature. These responses, like the memories that cause them, do not represent the centre of the past event as being identical with the subject of the memory. That is, the affective responses do not constitute a sense of personal identity. The feelings of familiarity lack a self-identifying function. Now, in the case of R.B., there is episodic memory without the usual affective response. This

[17] According to Coltheart (2005), this schematic account also applies to other kinds of delusion, including cases of mirrored-self misidentification.

odd feeling that 'something is different' requires an explanation, and R.B.'s hypothesis that he remembers events in which he was not present provides such an explanation. Since this inferential account of R.B.'s disownership-hypothesis is free of any involvement of a sense of personal identity at the level of episodic memory, it is compatible with non-reflexivism about episodic memory.[18]

Note that R.B. is not delusional, since he does not believe that someone else experienced those past events. He asserts that, during his 'unowned period', he knew that it was R.B. who had those experiences in the past: 'Intellectually I suppose I never doubted that [the experience] was a part of my life' (Klein & Nichols 2012: 686). R.B. has plenty of evidence for his own presence in the remembered events from sources other than episodic memory, including semantic memories and statements from relatives and friends. While R.B.'s interpretation of his memories as being about events not experienced by himself may be accounted for as resulting from the attempt of explaining the absence of typical affective memory-associations, the mentioned outside evidence keeps him from believing that this interpretation of his memories is correct. So R.B.'s impairment bears at most a limited resemblance to identity delusions such as the Capgras syndrome, a resemblance that concerns only the inferential link between the absence of feelings of familiarity and non-identity hypotheses.[19]

Could the feelings of familiarity be indicators of 'ownership' of past experiences in a non-identifying sense? I will not campaign for a negative answer. For it has not been my intention to argue that episodic memory fails to contain a sense of ownership in some sense or other. My target has only been the self-identifying sense, the sense of personal identity, that raises the circularity problem for Lockeanism. Though I cannot help finding the labels 'ownership' and 'mineness' misleading when self-identification is not part of their meaning. I will be happy, on the other hand, to assign a feeling of familiarity a role in the generation of hypotheses or beliefs about identity, while viewing this feeling itself as incapable of representing identity. If that makes the feeling a *weak* sense of ownership, that is fine

[18] This picture inherits some critical questions familiar from discussions of the explanation approach to monothematic delusions; see Fine et al. (2005) for an overview. Here is not the place to tackle them.

[19] R.B. not only fails to ascribe the subjective perspective on certain past events to himself, he also fails to describe those events as 'real'. This aspect raises the question whether R.B.'s past-directed states lack a further ingredient, a 'sense of reality', in addition to lacking a sense of personal identity. If so, does the expected sense of reality in typical episodic memory jar with my non-reflexivist account of typical memory involving feelings of familiarity? While this issue cannot be discussed appropriately here, I shall sketch a reason for expecting that there is no incompatibility problem. The mentioned aspect of R.B.'s case is reminiscent of judgements of 'unreality' in cases of derealization disorder. On the standard account of derealization, a core feature of the disorder is a 'persistent diminution or loss of emotional reactivity' (see Medford (2012)). If R.B.'s impression that the remembered events feal unreal, merely imagined, can be explained along these lines, as resulting from the absence of an expected emotional response to the memory of these events, then there is obviously no reason to expect an incompatibility with the suggested account of R.B.'s disownership-impression.

by me. The rough model just sketched only discusses such a feeling's role in generating hypotheses of non-identity. It will be interesting to see whether and how it is involved in generating hypotheses and beliefs about identity.

This is the beginning of a response to the disownership-challenge to non-reflexivism, which strikes me as plausible. The suggested strategy is meant to steer the discussion of the sense of personal identity in memory in a new, non-reflexivist direction, which promises to bring relief to Lockeans without committing them to a mysterious ability of self-detection in episodic memory. I admit, however, that the proposed picture leaves a number of questions unanswered.[20]

References

Baker, L. R. 2000: *Persons and Bodies*. Cambridge: Cambridge University Press.

Bayne, T. and Pacherie, E. 2004: 'Bottom-Up or Top-Down? Campbell's Rationalist Account of Monothematic Delusions', *Philosophy, Psychiatry, and Psychology*, 11: 1-11.

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.

Coltheart, M. 2005: 'Conscious Experience and Delusional Belief', *Philosophy, Psychiatry, and Psychology*, 12: 153-7.

Ellis, H.D. and Young, A.W. 1990: 'Accounting for Delusional Misidentifications', *British Journal of Psychiatry*, 157: 239-48.

Fine, C., J. Craigie, and I. Gold. 2005: 'Damned if you do, damned if you don't: the impasse in cognitive accounts of the Capgras delusion', *Philosophy, Psychiatry, and Psychology*, 12: 143–151.

Hume, David 1739: *A Treatise of Human Nature*, ed. David Fate Norton and Mary J. Norton. Oxford: Oxford University Press (2000).

Klein, S. B. 2012: 'The Self and its Brain', *Social Cognition*, 30: 474-518.

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).

[20] For comments on the material presented in this paper I am indebted to Adrian Alsmith, Sven Bernecker, Carl Craver, Brendan de Kenessey, François Recanati, Katia Samoilova, and audiences at the University of Milan and the Max Planck Institute for Biological Cybernetics, Tuebingen.

Medford, N. 2012: 'Emotion and the Unreal Self: Depersonalization Disorder and De-Affectualization', *Emotion Review*, 4: 139-44.
Noonan, 2003: *Personal Identity* (2nd edition). London: Routledge.
Nozick, R. 1981: *Philosophical Explanations*. Cambridge, MA: Harvard University Press.
Parfit, D. 1984: *Reasons and Persons*. Oxford: Oxford University Press.
Peacocke, C, 2014: *The Mirror of the World: Subjects, Consciousness, and Self-Consciousness*. Oxford: Oxford University Press.
Perry, J. (ed.) 1975: *Personal Identity*. Berkeley and Los Angeles: University of California Press.
Reid, Thomas 1785: *Essays on the Intellectual Powers of Man*. Edinburgh: Bell & Robinson.
Sattig, T. 2017: 'The Sense and Reality of Personal Identity,' *Erkenntnis* (online September 2017).
Schechtman, Marya 1990: 'Personhood and Personal Identity,' *Journal of Philosophy*, 87: 71–92.
Shoemaker, S. 1997: 'Self and Substance', in *Philosophical Perspectives* (Volume 11), J. Tomberlin (ed.): 283–319.
—— 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*. Ithaka: Cornell Unievrsity Press.
Velleman, J. D. 1996: 'Self to Self', in *Philosophical Review*, 105: 39-76.
de Vignemont, F. 2013: 'The Mark of Bodily Ownership', *Analysis*, 73: 643-51.
Williams, B. 1973: 'The Imagination and the Self', in *Problems of the Self*, Cambridge: Cambridge University Press.