

## Further thoughts on memory: replies to Schechtman, Adams, and Goldberg

Sven Bernecker

Published online: 21 November 2010

© The Author(s) 2010. This article is published with open access at Springerlink.com

**Abstract** This is a response to three critical discussions of my book *Memory: A Philosophical Study* (Oxford University Press 2010): Marya Schechtman, “Memory and Identity”, Fred Adams, “Husker Du?”, and Sanford Goldberg “The Metasemantics of Memory”.

**Keywords** Memory · Personal identity · Knowledge · Justification · Meaning · Mental content

The order in which I discuss the critics of my book *Memory: A Philosophical Study* is the order of the chapters they refer to.

### 1 Memory and personal identity: reply to Marya Schechtman

The book argues that the dependence of memory on personal identity is of a contingent rather than a logical kind. Memory presupposes personal identity only when the memory content involves an indexical reference to the rememberer. The veridicality constraint on remembering demands that the bearer of a memory with reflexive content is the same as the bearer of the corresponding past representation. Extroversive memories in the third-person mode, however, don't refer back to the rememberer and hence leave open the issue of the personal identity of the rememberer. Here interpersonal memory transfer is, in principle, possible. Since memory doesn't *logically* presuppose personal identity, it is possible to define

---

S. Bernecker (✉)

Department of Philosophy, University of California, Irvine, Irvine, CA 92697-4555, USA  
e-mail: s.bernecker@uci.edu

personal identity in terms of memory. Contrary to popular belief, memory-based accounts of personal identity are not circular.

Memory is a sub-species of quasi-memory. Quasi-memory is like ordinary memory in all phenomenal and causal respects, except that it is not restricted to experiences of one's own past. Quasi-memory doesn't presuppose that the bearer of the past experience is co-personal with the bearer of the present state of seeming to remember having had that experience.

Marya Schechtman's paper "Memory and Identity" is a rich and sophisticated defense of the logical dependence of memory and personal identity. On her view which I label *constitutive holism*, "the content of memories depends at least in part on their place in a broader psychological context, and so that some content will necessarily be lost when a memory is transplanted into an alien psychology" (Schechtman 2011, Sect. 2). Instead of constitutive holism, I adopt *content externalism*, that is, the view that the individuation of mental contents depends on systematic relations that the subject bears with certain conditions of his physical and social environment. More on content externalism in my reply to Sanford Goldberg.

The book (pp. 52–55) raises a number of concerns about constitutive holism. It points out that cases of amnesia and multiple personality disorder cast doubt on the empirical plausibility of constitutive holism. Moreover, it argues that the possibility of Parfit-type fission cases suggests that it is possible to quasi-remember another person's experiences. Regarding the latter point, Schechtman (2011, Sect. 2) explains that, according to my own theory, not all memories are transferable via fission. When introversive memory and extroversive memory in the first-person mode are reproduced in fission they don't count as quasi-memories because they violate the veridicality constraint. This point is well taken. I am also inclined to agree with Schechtman that it is difficult to draw any kind of firm conclusions from wildly hypothetical cases like fission. At the same time I doubt that there is a principled way of distinguishing between those thought experiments in the personal identity debate that are informative and those that are not.

Schechtman defends constitutive holism against the charge of empirical implausibility by drawing a distinction between kinds of memories. She concedes that there are memories that are logically independent of personal identity and that may be transferred as quasi-memories. Yet these are not the memories in terms of which personal identity is defined, namely *autobiographical experience memories* (Schechtman 2011, Sect. 2). By distinguishing between identity-dependent autobiographical experience memories and identity-independent extroversive memories in the third-person mode Schechtman attempts to acknowledge the empirical data in favor of quasi-memory while holding on to the claim that a memory-based theory of personal identity cannot escape circularity.

It seems to me that Schechtman doesn't fully appreciate the force of the empirical data in support of quasi-memory. Some of the studies suggest that also autobiographical experience memories are logically independent of personal identity. Consider, for example, the case of R.B., a 48 year old male who suffered a serious head injury in a car accident. Almost immediately following his accident, he was able to remember events from the past, yet his recollection of those events was compromised—he could not remember the events as having been personally

experienced. His memory of these events was no different from what it would have been had they happened to someone else. While his memories from before the injury had no sense of personal ownership, the memories from after the injury were normal. This is how R.B. describes what it is like to remember personal events without the experience to be ‘mine’

What I realized was that I did not ‘own’ any memories that came before my injury. I knew things that came before my injury. In fact, it seemed that my memory was just fine for things that happened going back years in the past. ... I could answer any question about where I lived at different times in my life, who my friends were, where I went to school, activities I enjoyed, etc. But none of it was ‘me’. It was the same sort of knowledge I might have about how my parents met or the history of the Civil War or something like that (Klein 2011, Sect. 2).

And this is how R.B. describes the gradual recovery of ownership of his autobiographical memories:

What happened over the coming months was interesting: Every once in a while, I would suddenly think about something in my past and I would ‘own’ it. That was indeed something ‘I’ had done and experienced. Over time, one by one I would come to ‘own’ different memories. Eventually, after perhaps eight months or so, it seemed as if it was all owned. As if once enough individual memories were owned, it was all owned (Klein 2011, Sect. 2).

To be sure, these findings and other like them (Bernecker 2010, pp. 53–54) don’t *prove* that memory is a sub-class of quasi-memory. But they do seem to cast doubt on Schechtman’s claim that the kinds of memories that are constitutive of personal identity logically presuppose personal identity. The case against memory-based accounts of personal identity doesn’t seem to be empirically realistic.

Next Schechtman argues that there is a tension between my cautious approval of the connectionist model of memory traces (Bernecker 2010, pp. 135–137, 194) and my defense of the possibility of interpersonal memory transfers. Given that memory traces consist in a distributed weighting of synaptic connections and that many different memories can be stored on the same set of connections, it doesn’t seem possible to transplant a small number of traces. Schechtman calls this the *engineering version* of the argument from constitutive holism (2011, Sect. 3). Two comments in response. First, if or when memory transplants become a medical possibility, surgeons may find a way to transplant single memory traces. Second, even if should turn out that single memory transplants are not medically possible, this doesn’t prove that memory logically presupposes personal identity.

If it is possible to transfer memories from one person to another then it should also be possible to transfer memories from one person to many others. Furthermore, if a memory trace is nothing but a continuous means of preserving the propositional information contained in a memory, then memories can be transferred not only by means of intracerebral states but also by means of pen and paper or spoken words. Schechtman is agreeable to these consequences provided we focus on propositional memory. For in the case of propositional memory it doesn’t matter whether we

remember our own past experiences or whether our memories are caused by reading or hearing about someone else's experiences. But in the case of what she calls *experiential memories* it is not enough that we have access to propositional information about the past with a reliable pedigree. Since one can experientially remember only what one has personally experienced experiential memories, like autobiographical experience memories, presuppose personal identity. "The causal connection and trace conditions might be perfectly adequate to define some kinds of memories, but there are others for which we need stronger requirements, and these seem inherently linked to facts about identity" (Schechtman 2011, Sect. 3).

Schechtman concludes that my account of memory is incomplete. I agree that more can and ought to be said about the relation between memory and personal identity. What I don't agree with is that my classification of kinds of memories in terms of the grammatical objects of the verb 'to remember' "makes no room" for, and "suppresses" distinctions that are crucial to understanding the nature of memory (Schechtman 2011, Sect. 3–4). This objection is misplaced since I don't lay claim to my taxonomy being the only viable one. I merely claim that the traditional distinction between experiential and propositional memory is neither sharp nor intuitively compelling.

## 2 Memory and knowledge: reply to Fred Adams

According to the widespread epistemic theory of memory, to remember something is to know it where this knowledge was previously acquired and preserved. The book argues against the epistemic theory by showing that, unlike knowledge, memory implies neither belief nor justification (however construed). Fred Adams's paper "Husker Du?" is a careful and astute discussion of my examples and arguments against the epistemic theory of memory. Limitations of space prevent me from examining all of Adams's objections. I will confine myself to the most important ones.

Counterexamples to the claim that memory implies justification can be divided into two groups. According to *gettierized justification cases* (examples # 1 and 2), one can remember that  $p$  despite the fact that one's past or present belief that  $p$ , though true and subjectively justified, is gettierized and thus doesn't qualify as knowledge. According to *defeated justification cases* (examples ## 3, 4 and 5), one can remember that  $p$  even though there is (was) some defeating information such that, if one became (had become) aware of it, one would no longer be justified in believing  $p$  and hence wouldn't know. Let's start with Adams's objections to gettierized justification cases.

Example #1 deals with someone who recalls that a friend has borrowed a certain book from the library ( $p$ ). The point of the example is that the subject can remember at  $t_2$  that  $p$  even though his belief at  $t_1$  that  $p^*$ , though true, was gettierized. Example #2 is like #1 except that the belief at  $t_1$  that  $p^*$  was false. The gettierization takes place sometime between  $t_1$  and  $t_2$  by  $p^*$  becoming accidentally true.

Adams complains that example #1 is under-described. This may be true. But Adams's (2011, Sect. 3) adaptation of #1 is clearly uncharitable, for it changes the

character of the example: it turns a Gettier case into a case of genuine knowledge (on the tracking style analyses). It is thus not surprising that Adams's adaptation of #1 is fully compatible with the epistemic theory of memory.

Adams raises two objections regarding example #2. The first objection is that it is a case of 'remembering  $p$ ' rather than 'remembering that  $p$ ', where the former, but not the latter, allows for  $p$  to be false (Adams 2011, Sect. 3). As far as I can see, this distinction lacks phenomenological or linguistic support. We wouldn't say, to use Adams's example, that someone can remember green ideas sleep furiously. What we would say instead is that someone can remember having believed that green ideas sleep furiously. What is more, there is no reason to deny the protagonist of example #2 memory that  $p$ ; after all  $p$  is true at the time of recall.

Adams's second objection concerns my thesis (Bernecker 2010, pp. 38–39, 74, 150–151, 215) that the authentic reproduction of a formerly false proposition which in the meantime has become true due to good fortune may qualify as memory. Not requiring that the proposition be true already at the time of input makes the concept of propositional memory too 'promiscuous' for Adams's (2011, Sect. 3) taste. Two comments in response. First, most real-life cases where the proposition fed into the memory process is false but the proposition emerging from the process is true fail to qualify as instances of memory because they violate the connection condition or the content condition. Second, requiring that the remembered proposition be true already at the time of input would make sense if memory's primary function of memory was *truth* preservation. But it is not. In the first instance, memory is meant to preserve *content*, regardless of its truth value. This is why the cognitive process underwriting so-called 'false memory' may be the same as the process underwriting memory (Bernecker 2010, pp. 198–201, 214–217). Adapting an aphorism by Wittgenstein (1969, § 505) we can say, it's always by the favor of nature that one remembers something. In other words, a person's memory claim can be false without his memory being at fault. The memory claim can be false because the content was already false at the time of input or because it was made false in the meantime. One source of blameless memory failure is slow switching discussed in my reply to Sanford Goldberg.

By authentically preserving content the faculty of memory goes some way toward ensuring that the recalled content is true, provided the past content was true. For if the contents fed into the memory process are veridical and if there are no external circumstances changing the truth values of the contents while they are stored, then truth preservation is a gratuitous by-product of content preservation.

Next turn to defeated justification cases, where a subject remembers that  $p$  but where there is some defeating information such that, if the subject became aware of it, he would no longer be justified in believing  $p$ . In example #3, it is the past belief that is defeasibly justified; in example #4 it is the present belief.

Example #3 is presented as a case where someone remembers at  $t_2$  that  $p$  even though he wasn't justified at  $t_1$  in believing (and hence didn't know) that  $p^*$  because he had plausible yet misleading reasons to suppose that  $\sim p$ . Adams claims that I am mistaken in thinking that this is a case of memory without past justification. According to Adams, the protagonist of example #3 knows at  $t_1$  that  $p^*$ , loses his knowledge between  $t_1$  and  $t_2$  due to the presence of undercutting defeaters, and then

comes again to know at  $t_2$  that  $p$ . He concludes that #3 “is not a counterexample to the epistemic theory of memory” (2011, Sect. 3).

Even if I were to agree with Adams’s reading of example #3, it would still count against the epistemic theory of memory. For the epistemic theory of memory holds not only that remembering is a form of knowing but also that memory can only preserve but not generate justification and knowledge (Bernecker 2010, pp. 65–67, 96–97). Yet on Adams’s reading of example #3, memory does generate justification and knowledge: it transforms at  $t_2$  a belief into knowledge that had qualified as knowledge at  $t_1$  but that has lost its status as knowledge sometime between  $t_1$  and  $t_2$ . So example #3 definitely contradicts the epistemic theory of memory. The question is only which aspect of the epistemic theory it calls into doubt.

Example #5 is a case of defeated present justification. The protagonist of this example has plausible yet misleading reasons to suppose that  $\sim p$  and is therefore not justified in believing that  $p$ . The presence of an undefeated defeater robs him of the ability to know that  $p$  but it doesn’t affect his ability to remember that  $p$ . According to Adams, given that the defeating evidence prevents the subject from knowing  $p$ , it would be irrational for him to continue to believe that  $p$ . “To be rational he should at least withhold belief until he learns more” (Adams 2011, Sect. 3). The protagonist of example #5 neither believes nor remembers that  $p$ ; or so Adams claims. Two comments in response. First, Adams assumes, without argument, that belief is a necessary condition for memory. As we will see shortly, the book argues against this view. Given that belief isn’t necessary for memory, it is possible that the subject fails to believe that  $p$  but still remembers that  $p$ . Second, it need not be irrational to hold on to a belief in light of defeating evidence. The defeater that prevents someone from knowing  $p$  could be a partial defeater. Partial defeaters are defeaters that don’t require withholding a belief but only require holding it less firmly; they defeat justification without defeating belief (Bergmann 2006, p. 155n6).

The cases of memory without justification discussed so far have in common that the content of the past belief is type-identical to the content of the memory belief. Yet it is not essential for memory that the content of memory belief be of the same kind as the content of the past belief. Memory requires only that the diachronic content tokens be sufficiently similar. The book argues for the *entailment thesis*, whereupon a belief at  $t_2$  is memory-related to a belief at  $t_1$  only if the content of the belief at  $t_2$  is relevantly entailed by the content of the belief at  $t_1$ .

The entailment thesis, in combination with tracking-style accounts of knowledge, provides further support for the view whereupon one can remember at  $t_2$  that  $p$  in virtue of having believed at  $t_1$  that  $p^*$ , and where the belief at  $t_1$  was justified and known but the memory belief at  $t_2$  is not. In the book (pp. 80–81) I use Kripke’s red-barn example (#6) to bring this point home: the subject who is driving through Kripke’s fake-barn country can’t *know* of a red barn he is looking at that it is a barn (because he can’t track the fact that it is a barn) but he can *remember* that is a barn. Adams, however, maintains that if the subject driving through fake-barn country cannot know that something is a barn, then he can’t remember it either. Following Dretske and Yourgrau (1983), he claims that the protagonist of example #6 “forgot that it is a barn, despite continuing to believe that is a barn” (2011, Sect. 3) and despite satisfying the connection, content and truth condition on memory. The

reason the subject can't remember that  $p$  is the he can't track the facts that make  $p$  true. Adams grants that this move is counterintuitive but he thinks that it is not any more counterintuitive than my non-epistemic account of memory. I beg to differ. Adams's proposal renders memory a very rare commodity indeed. If memory required tracking-style knowledge at the time of input and at the time of recall, as Adams claims, then most of the mental states we call *memories* would be so-called 'false memories'. The tracking account of memory leads into skepticism about memory. And it strikes me as obvious that memorial skepticism beats the non-epistemic theory in terms of counterintuitivity (Bernecker 2008, pp. 122–126).

Granted that knowing implies believing, the epistemic theory of memory has it that remembering too implies believing. In the book (pp. 88–90) I use an example adapted from Malcolm (example #10) to argue that it is possible to remember that  $p$  without believing that  $p$ . The protagonist of example #10 remembered at  $t_2$  that he was kidnapped at  $t_1$  but since it takes considerable convincing (he is shown police and news accounts) until he accepts at  $t_3$  that he was kidnapped at  $t_1$  it would be wrong to say that he believed at  $t_2$  that he was kidnapped. Adams (2011, Sect. 3) defends the epistemic theory of memory against this example by proposing a different reading. On Adams's reading, the subject doesn't remember that he was kidnapped at  $t_1$  until he believes and knows it at  $t_3$ . I disagree. It is surely correct to say that the subject comes to know at  $t_3$  that he was kidnapped as a child. But at the same time he ceases to remember that he was kidnapped. The reason he ceases to remember at  $t_3$  is that he violates the counterfactual connection condition (Bernecker 2010, pp. 144–151). This condition states that for a subject's representation at  $t_3$  that  $p$  to be memory-related to his representation at  $t_1$  that  $p^*$ , the causal chain connecting the two states must be such that if he hadn't represented at  $t_1$  that  $p^*$  he wouldn't represent at  $t_3$  that  $p$ . The protagonist of example #10 violates the counterfactual connection condition because when he is presented with police and news accounts at  $t_3$  he comes to represent that he was kidnapped regardless of whether he had a corresponding representation at  $t_1$ .

In the section "Questions for Bernecker" Adams asks whether, on my non-epistemic account of memory, one can remember that  $p$ , remember that  $p$  entails  $q$ , and not remember that  $q$ . The answer is yes. A necessary condition for remembering  $q$  is representing  $q$ . It seems perfectly possible that one doesn't put two and two together and thus fails to represent that  $q$  while remembering that  $p$  and remembering that  $p$  entails  $q$ . Another reason for why one may fail to remember  $q$  even though one remembers  $p$  and that  $p$  entails  $q$  is that one's representation that  $q$  doesn't satisfy the connection condition for memory. Suppose the representation that  $q$  is hard-wired: one represents that  $q$  come what may. Given that one would represent at  $t_2$  that  $q$  even if one hadn't represented at  $t_1$  that  $q^*$ , the representation violates the counterfactual connection condition and hence doesn't qualify as memory.

Adams paper is in large parts a defense of the epistemic theory of memory, that is, the thesis that to remember something is to know it. If truth is a necessary condition of knowledge, as it certainly is, then by transitivity of implication, truth is a necessary condition of memory. At the end of his paper Adams suggests, however, that one can remember propositions that are not true. If he is right, this would undermine both the epistemic and the non-epistemic account of memory. The cases

of so-called ‘false memory’ Adams discusses concern semantically incomplete or gappy thoughts, for instance, remembering that Pegasus is a winged horse. According to the direct reference theory, whose truth Adams assumes, the content of the thought that Pegasus is a winged horse is neither true nor false. But then I can’t remember that Pegasus is a winged horse, since what I remember isn’t true. This strikes Adams as false. He insists that “there is some memory phenomenon going on here” (2011, Sect. 4). I agree. When I claim to remember that Pegasus is a winged horse I may remember that I believed that Pegasus is a winged horse. Memory reports involving vacuous singular terms may express what I call *introspective memory*, that is, propositional memory of one’s own mental states.

### 3 Memory and meaning: reply to Sanford Goldberg

Sanford Goldberg’s paper “The Metasemantics of Memory” is a detailed and insightful discussion of aspects of my externalist account of memory content. The book defends pastist externalism, that is, the view whereby the content of a memory-constituting representation is fixed, once and for all, by the physical and social environment the subject was in at the time he had the original representation. Goldberg expresses his support for content externalism in general and pastist externalism in particular. From his point of view, the difference between his position and mine concerns the interpretation of slow switching scenarios. While I agree with some of Goldberg’s specific points, I believe he is mistaken about the overall dialectical situation. The real dispute between Goldberg and me concerns the interpretation of content externalism not the interpretation of slow switching scenarios.

The main motivation for content externalism are Twin Earth thought experiments due to Putnam and Burge. Consider two individuals who are molecular duplicates. One of them lives on Earth while the other lives on Twin Earth, a planet which is a duplicate of Earth except for the fact that the liquid that the Twin Earthians call ‘water’ looks, tastes, and behaves like water but has the chemical composition XYZ and not H<sub>2</sub>O. When an Earthian expresses the thought that water is wet, his Twin Earthian twin expresses a different thought—the thought that twater is wet. This difference in concepts and thought contents, according to the externalist, is due to the difference in environments.

In slow switching thought experiments we are invited to imagine an Earth-bound subject who is unwittingly switched to Twin-Earth and remains there for some time. While all proponents of content externalism agree that the environmental change eventually brings about a conceptual change opinions differ concerning the kind of conceptual change. In the *conceptual replacement view*, the slow switching causes the subject to adopt new concepts and to lose the old ones. In the *conceptual addition view* (Goldberg calls it the *supplement view*), the slow switching causes the subject only to adopt new concepts but not to lose the old ones.

A putative problem of the replacement view is that an unwitting environmental shift can bring about an unwitting conceptual shift, that, in turn, can rob the Earthian expatriate residing on Twin Earth of the concepts needed to reactivate certain



thoughts from before the switch. For instance, if the subject used to think the thought that water is wet and if his *water* concept has since been replaced by the *twater* concept, then he is unable to reactivate his Earthian ‘water’ thought—except perhaps by means of a description, such as “the clear liquid in Lake Tahoe in the summer of 1978”. The environmental shift causes memory failure. Some philosophers find the context-dependence that follows from the combination of pastist externalism and the replacement view too implausible to be acceptable and therefore opt for the addition view. Given the addition view, unwitting slow switching doesn’t cause memory failure, for the switched subject retains the concepts needed to reactivate his thoughts from before the switch.

Goldberg takes me to endorse the replacement view and to reject the addition view which he endorses. But I don’t reject the addition view. Rather my point is that proponents of pastist externalism are not *compelled* to adopt the addition view but are free to embrace the replacement view (Bernecker 2010, p. 171, 195). Both views have their advantages and disadvantages. When pastist externalism is combined with the replacement view, there is the issue of semantically-induced memory failures. Chapter seven of the book shows that this issue need not be taken seriously, for there are no good reasons to doubt that environmental shifts can bring about memory failure. When pastist externalism is combined with the addition view, the issue of semantically-induced memory failure doesn’t arise. But this advantage comes at a cost, for the addition view has problems of its own.

To see the problems of the addition view let’s assume its truth: a switched subject’s term *water* can mean *water* or *twater*. For slow switching not to give rise to semantically-induced memory failure we have to assume that our mind is equipped with a mechanism that selects the contextually appropriate meaning. This mechanism assures that the term ‘water’ employed by an Earthian expatriate residing on Twin Earth means *water*, when it is used in a memory claim, and means *twater*, when it is used to refer to the liquid that he presently swims in. But how should we conceive of such a mechanism for filling in the appropriate meaning for ambiguous terms? This is where Tyler Burge’s (1998) account of *preservative memory* comes into play. Burge’s idea is that just as the reference of an anaphoric pronoun is secured by the chains inherent in the discourse, the reference of the expatriate’s use of ‘water’ in a memory claim is fixed by the causal chains leading to the appropriate past meaning.

The book (pp. 193–196) presents three arguments against the addition view or the preservative model of memory, respectively. Contrary to what Goldberg says, these arguments are not meant to amount to a conclusive refutation of the addition view; instead they are meant to motivate the defense of the context-dependence of memory that follows from the combination of pastist externalism and the replacement view. Given that the addition view has problems of its own, we ought to take a second look at the *prima facie* counterintuitive consequences of the replacement view and weigh them up against the problems of the addition view.

The lion’s share of Goldberg’s paper is a critical discussion of my arguments against the addition view or the preservative account, respectively. I will comment

on Goldberg's objections to my arguments before I turn to his version of content externalism.

Goldberg's response to my first argument rests on a misunderstanding. The point of my argument is that the theory of preservative memory fails to account for the fact that memory is also an active device for processing stored contents rather than solely a passive device for reproducing contents. Using pronominal back-reference as a model for understanding how memory states get their meanings works only in the rare cases where memory is a purely passive device and where consequently the memory content is type-identical with the content of the past propositional attitude. Just as the meaning of the anaphoric pronoun is the same as the meaning of the textual antecedent, the content of the present propositional attitude is said to be the same as the content of the past propositional attitude. Even though the thesis that memory requires identity of diachronic contents is widespread among philosophers it is fundamentally at odds with what cognitive psychology tells us about the workings of human memory. Our memory is not only a passive device for reproducing contents but also an active device for processing stored contents (Bernecker 2008, pp. 144–154; 2010, pp. 217–221). But when there is no perfect match between the past and present content tokens the pronominal back-reference model doesn't help us to understand how states of recall get their meanings.

Since Goldberg says that he doesn't understand my second argument against preservative memory, I give it another go. The preservative account of memory draws an analogy between the way an anaphoric pronoun inherits its reference from a textual antecedent and the way a state of remembering inherits its reference from a past representation. For this analogy to be of explanatory value we have to assume that representations are language-like in nature, that they are symbolic structures with semantically evaluable constituents. If the structure of representations didn't resemble that of natural language, the phenomenon of pronominal back-reference couldn't be used to shed light on the memory process. But it is still very much an open question whether memories and other representational states are sentences in the language of thought or whether they are connectionist structures. Moreover, it is an open question whether there is a neat correspondence between linguistic structures (e.g., pronouns) and activity patterns over a neural network. Thus, it is not prudent to settle for an interpretation of slow switching scenarios that seems to be incompatible with connectionist models of learning and memory.

My third objection to the addition view is that it is unable to account for the Kripkean modal intuitions: suppose an Earthian who never knew that water is  $H_2O$  is unwittingly switched to Twin Earth and is told that there is another planet, just like this one, except that it contains  $H_2O$  instead of XYZ. The subject is then asked whether the liquid the inhabitants of the other planet designate with 'water' is really water. According to Kripkean intuitions, the answer is no. On the addition view, however, this answer is false. Goldberg objects that it is not the addition view per se that is incompatible with Kripkean intuitions but only a variety of this view. The addition view comes in two flavors. In the *ambiguity view*, slow switching causes the agent to take on two distinct concepts (*water* and *twater*) for different sorts of things ( $H_2O$  and XYZ). In the *amalgamation view* (Goldberg calls it the *supplemental-cum-amalgam view*), switching causes the agent's old concept to

broaden by applying to H<sub>2</sub>O and XYZ simultaneously. The new concept is an amalgam made out of the two concepts *water* and *twater*, perhaps *water-or-twater*. Goldberg agrees that the amalgamation view violates Kripkean intuitions but he insists that the ambiguity view does not.

Given the ambiguity view, when a resident of Twin Earth who originated on Earth answers in the negative to the question “Is the liquid that people on Earth designate as ‘water’ really water?”, the answer is false if the second occurrence of ‘water’ expresses the concept *water*. For the subject’s negative answer to be correct—the way the Kripkean intuition has it—the second occurrence of ‘water’ must be interpreted as expressing the concept *twater*. Goldberg argues that this is also the natural interpretation of the above question: “insofar as [our subject] intends to grasp the question that the speaker meant to pose, [the subject] must interpret the relevant occurrence of ‘water’ as expressing *twater*” (2011, Sect. 3.3). Even if we grant Goldberg this point, it is nevertheless possible that the subject deploys the wrong concept—the *water* concept—when interpreting the above question and hence violates the Kripkean intuition. Such mistakes are unavoidable since the subject unwittingly possesses both concepts and is unable to discriminate between them. What is more, Goldberg owes us an explanation of the mechanism that enables the subject to select the correct concept when interpreting the question at hand. Note that the preservative account is of no use here since the interpretation of the question at hand doesn’t involve memory.

The book (pp. 195–196) also contains a brief discussion of *knowledgeable* slow switching. I mention in passing that if we follow Gibbons (1996) in combining content externalism with functional role semantics, the appeal to the addition view doesn’t save us from postulating semantically-induced memory failures in knowledgeable slow-switching cases. Goldberg has a lot to say about Gibbons’s version of externalism and my reading of Gibbons. In the end, the dispute comes down to the kind of addition view one takes as a basis. Goldberg shows that for functional role externalism to have the consequence that an environmental shift brings about memory failure we have to interpret knowledgeable switching cases along the lines of the amalgamation view. But if we assume the ambiguity view, considerations of functional role cannot be used to show that knowledgeable switching gives rise to semantically-induced memory failures. I agree. What is curious, however, is that Goldberg agrees with me that Gibbons’s position seems to be incompatible with the ambiguity view. Thus the solution Goldberg proposes on Gibbons’s behalf is at odds with his reading Gibbons’s position.

Goldberg rounds his paper off with a critique of my reading of the replacement view whereby it gives rise to semantically-induced memory failures. The critique proceeds in two steps. First, Goldberg points out that, even on the replacement view, a subject who has been switched from Earth to Twin Earth is able to refer to H<sub>2</sub>O via a description (e.g., the clear liquid in Lake Tahoe in the summer of 1978) and to pick out the lost *water* concept via a description (e.g., the concept through which ordinary people in the summer of 1978 thought about the clear liquid in Lake Tahoe). Second, Goldberg’s argues that, if we grant a switched agent the ability to refer to the lost *water* concept, we should also grant him the ability to ‘re-deploy’ (Goldberg’s term) the lost *water* concept. The idea is that the switched agent’s

intention to use ‘water’ to refer to H<sub>2</sub>O overrides the default reference to XYZ. “Here the speaker’s descriptive semantic intentions serve to fix the referent” (Goldberg 2011, Sect. 4). Goldberg’s conclusion is that I am wrong in claiming that, given the replacement view, slow switching causes semantically-induced memory failures.

I would like to make two comments in response to the second step of Goldberg’s critique. First, given content externalism, being able to pick out a concept by means of a description isn’t sufficient for possessing the concept. To see this consider the following case. A young child overhears his parents having a heated debate about whether combustible bodies contain phlogiston. The child is able to pick out the concept *phlogiston* via a description (e.g., the concept my parents thought about last night), yet no one would conclude that the child possesses the concept *phlogiston*. But if the child doesn’t possess the concept *phlogiston*, then the switched agent who picks out his lost *water* concept via a description doesn’t possess the concept *water* either. For there is no semantically relevant difference between the child referring via a description to a concept he never possessed and the twin world traveler referring via a description to a concept he once possessed but has since irretrievably lost.

Second, Goldberg’s criticism rests on an idiosyncratic reading of content externalism. The idea that the speaker’s descriptive intentions can somehow override the content-determining powers of the physical and social environment, though ingenious, isn’t part of content externalism. In fact Twin Earth thought experiments show that the meaning of a term such as ‘water’ can be radically different from the meaning the speaker thinks the term has or the meaning he intends the term to have. I thus see no reason to retract my claim that, given content externalism, if a switched subject’s *water* concept is replaced by the *twater* concept, he thereby loses the ability to reactivate his Earthian ‘water’ thoughts.

**Acknowledgments** An earlier version of this paper was presented at the Pacific Division Meeting of the APA in San Francisco in 2010. The APA session was organized by Patrick Rysiew and chaired by Rebecca Copenhaver. I would like to thank my critics for the time they have invested in my work. There are many places where they have helped me to understand the issues more clearly and, hopefully, deal with them more adequately. I would also like to thank Rebecca Copenhaver and the journal editor for the opportunity to discuss the book and to develop some thoughts where more need to be said. Work on this paper was supported by a research grant (#A/10/07576) from the German Academic Exchange Service (DAAD) and a research grant (#AY2009-2010) from the Academic Senate Council on Research, Computing and Libraries (CORCL) at the University of California, Irvine.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Non-commercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Adams, F. (2011). Husker du? *Philosophical Studies*. doi:10.1007/s11098-010-9663-4.
- Bergmann, M. (2006). *Justification without awareness*. Oxford: Oxford University Press.
- Bernecker, S. (2008). *The metaphysics of memory*. Dordrecht: Springer.

- Bernecker, S. (2010). *Memory: A philosophical study*. Oxford: Oxford University Press.
- Burge, T. (1998). Memory and self-knowledge. In P. Ludlow & N. Martin (Eds.), *Externalism and self-knowledge* (pp. 351–370). Stanford: CSLI Press.
- Dretske, F., & Yourgrau, P. (1983). Lost knowledge. *Journal of Philosophy*, 80, 356–367.
- Gibbons, J. (1996). Externalism and knowledge of content. *Philosophical Review*, 105, 287–310.
- Goldberg, S. (2011). The metasemantics of memory. *Philosophical Studies*. doi:[10.1007/s11098-010-9642-9](https://doi.org/10.1007/s11098-010-9642-9).
- Klein, S. B. (2011). The two selves: The self of conscious experience and its brain. In M. R. Leary (Ed.), *Handbook of self and identity* (2nd ed.). New York: Guilford Press (forthcoming).
- Schechtman, M. (2011). Memory and identity. *Philosophical Studies*. doi:[10.1007/s11098-010-9645-6](https://doi.org/10.1007/s11098-010-9645-6).
- Wittgenstein, L. (1969). In G. E. M. Anscombe & G. H. von Wright (Eds.), *On certainty*. New York: Harper and Row.