whether episodic memory might show some functional adaptation to facilitate commitment. Specifically, might there be a propensity for stronger encoding of or retention of episodic memories that are commitment related? This could be tested – following a procedure developed by Conway (2009; cf. Williams et al. 2008) – by asking people to list as many specific memories as possible for yesterday, two days ago, three days ago, and so on, and measuring the frequency of memories in which social commitments are generated. In order to determine whether there is commitment-specific facilitation, it would be important to compare the effects of commitment on memory with other factors such as generalized personal or social significance. Should commitment-specific facilitation be found, the further question would be whether this is an evolutionary adaptation, and it would be important to rule out alternative explanations such as enculturation operating on developmental plasticity. It is in general challenging to infer back from current function to evolutionary adaptation, and especially so in the case of multifunctional traits.

With this cautionary note in mind, we may venture to observe that Conway's theory of episodic memory also raises interesting considerations for understanding the origins of commitment in human evolution. Conway's hypothesis is that the function of episodic memory is to maintain a record of progress in relation to short-term goals (Conway 2009). In this respect, it is important to note that active working memory has limited capacity, which means that over the course of temporally extended goal-directed activities, task-related information must be stored and retrieved from long-term memory (LTM). There are compelling reasons to think that the form of LTM involved is episodic memory (self-involving) and not mere event memory (no self present). After all, Alan must be aware that X is *his* goal and that *he* has performed a particular set of task-related actions up to this point. Merely remembering that *some* agent was performing a task, which the individual somehow egocentrically remembers, isn't enough to carry on with the task.

Complex, temporally extended goal-directed activities arguably play an important adaptive role for a number of nonhuman species, and clearly were extremely important in human evolution. This lends considerable plausibility to Conway's theory. More recently (in phylogenetic terms), human lifeways have been shaped by the importance of coordinating with others in *joint* goal-directed activities. It is therefore tempting, against the backdrop of Conway's theory, to speculate that episodic memory may have come to support the function of keeping track of who is committed to what within the context of joint goal-directed activities. Could such information be encoded directly to semantic memory, bypassing episodic memory? Possibly, but on Conway's account, episodic memory forms the basis for higher-order conceptual memory structures (e.g., a conceptual frame like *a day at work*) that provide narrative structure which organizes specific episodes. Thus, according to Conway, episodic memory plays a foundational role in the development of higher levels of narratively structured memory.

Conway's theory offers an attractive framework for understanding the evolution of social commitment and, in so doing, provides an illuminating backdrop to M&C's analysis. As M&C point out, episodic memory is important in grounding social commitments. This is surely true. It is also true that, to make a commitment, you have to already be capable of engaging in temporally extended goal-directed activity – otherwise there is nothing to make a commitment about. Furthermore, the social regulation of commitments (especially to the extent that the commitment is implicit and/or indeterminate) is likely to involve the detailed narrative memory for goal-directed activity described by Conway's theory.

Autonoesis and reconstruction in episodic memory: Is remembering systematically misleading?

doi:10.1017/S0140525X17001431, e22

Kourken Michaelian

Department of Philosophy, University of Otago, Dunedin 9054, New Zealand.
kourken.michaelian@otago.ac.nz **http://phil-mem.org/**

Abstract: Mahr & Csibra (M&C) view autonoesis as being essential to episodic memories and construction as being essential to the process of episodic remembering. These views imply that episodic memory is systematically misleading, not because it often misinforms us about the past, but rather because it often misinforms us about how it informs us about the past.

Mahr & Csibra (M&C) argue that the function of episodic memory is to enable a subject to persuade others to endorse the subject's descriptions of past events. Although the authors build an impressive case for this communicative account, it turns out to be committed to a counterintuitive claim, namely, that episodic memory is systematically misleading. Other accounts, including the future-oriented account (e.g., Schacter & Addis 2007), likewise turn out to be committed to this *misleadingness claim*. The future-oriented account sees episodic memory, along with episodic future thought (Szpunar 2010), as a form of mental time travel (MTT) (Suddendorf & Corballis 1997), with future-oriented MTT or episodic future thought being primary, in the sense that the function of the MTT system is to enable the subject to imagine future events, while the ability to engage in forms of past-oriented MTT, including episodic memory, emerges as a by-product. Although they differ on the question of the function of the memory or MTT system, the future-oriented account and the communicative account agree on two claims that together imply the misleadingness claim: (1) that episodic memories necessarily involve autonoesis (the *autonoesis claim*) and (2) that episodic remembering is necessarily a constructive process (the *construction claim*).

The autonoesis claim: M&C understand autonoesis in metarepresentational terms (cf. Dokic 2014; Fernández 2016), characterizing the content of a retrieved memory as having two components: a first-order component informing a subject about an event and a second-order component informing him or her that the information provided by the first-order component originates in the subject's own experience of the event. If retrieved memories are indeed metarepresentational, then, when retrieval results in the formation of a belief, the subject believes not simply that such-and-such an event occurred but rather that he or she knows that such-and-such an event occurred because of having experienced its occurrence. Crucially, the second-order component of a memory belief might be inaccurate – and hence the belief as a whole might be false – even if the first-order component is accurate, simply because there are sources of accurate information about an event other than one's own experience.

The autonoesis claim is essential to the communicative account: In making a memory claim, a subject claims epistemic authority over the event in question, and autonoesis is normally the subject's only ground for doing so. The claim might not, strictly speaking, be essential to the future-oriented account: Because autonoesis may not play a role in episodic future thinking (Perrin 2016), the future-oriented account might replace it with a weaker claim, namely, that although autonoesis typically plays a role in episodic remembering, it is not a necessary feature of retrieved memories (Michaelian 2016b). Even this weakened claim is, however, sufficient to commit the future-oriented account to the misleadingness claim.

The construction claim: M&C understand construction as occurring through Bayesian prediction of features of past events based on evidence provided by both episodic traces and semantic information (De Brigard 2014a). Alternative understandings are available (Michaelian 2016b), but they concur that, at least in typical cases, not all of the content of a given retrieved memory originates in the subject's experience of the remembered event. This, in turn, implies that retrieved memories will often be, to some degree, inaccurate with respect to remembered events. But construction does not make inaccuracy inevitable: The incorporation of nonexperiential information into a retrieved memory representation, in particular, does not necessarily imply inaccuracy, simply because incorporated information may itself be accurate (Michaelian 2013).

The construction claim is essential to the communicative account: If the point of making memory claims is not to convey accurate descriptions of past events but rather to convey descriptions that the subject wants an audience to endorse, a constructive memory process is needed to enable the subject to generate suitable representations of events. The claim is likewise essential to the future-oriented account: The MTT system must be able to constructively recombine and modify information from various sources in order to generate representations of possible events in episodic future thinking, and if episodic remembering is carried out by the same system, it is bound to be constructive in the same sense.

The misleadingness claim: Together, the (weakened) autonoesis claim and the construction claim imply the misleadingness claim. If the autonoesis claim is right, a memory might be false even if the event that it represents occurred exactly as the belief represents it as having occurred. In particular, the belief will be false in cases in which its first-order content originates at least in part in a source other than the subject's own experience of the event. If the construction claim is right, such cases occur frequently. Indeed, because, as M&C acknowledge, episodic remembering is driven as much by current beliefs as by episodic traces, they are the rule rather than the exception. Therefore, the second-order component of a memory belief – and the belief as a whole – will frequently be false. In short, both the communicative account and the future-oriented account are committed to the claim that episodic memory beliefs are frequently false, not because construction results in inaccurate representations of events, but rather because autonoesis results in inaccurate metarepresentations of the relationship between representations and the sources in which they originate, both where events are represented inaccurately and where they are represented accurately.

We might, in principle, attempt to avoid the misleadingness claim by rejecting either the construction claim or the autonoesis claim, but we have good reason to accept both of these claims. We might also attempt to avoid it by modifying the metarepresentational understanding of autonoesis so that the autonoesis claim says that the second-order component of a retrieved memory informs a subject only that part of the first-order component of the memory, as opposed to the first-order component as a whole, originates in the subject's experience of the event, but it is unclear whether this is compatible with the roles assigned to autonoesis by the communicative and future-oriented accounts. We may thus be forced to accept the counterintuitive conclusion that episodic memory is indeed systematically misleading.

Autonoesis and dissociative identity disorder

doi:10.1017/S0140525X17001558, e23

John Morton

Institute of Cognitive Neuroscience, University College, London WC1N 3AR, England.

j.morton@ucl.ac.uk **https://johnmorton.co.uk/**

Abstract: Dissociative identity disorder is characterised by the presence in one individual of two or more alternative personality states (alters). For such individuals, the memory representation of a particular event can have full episodic, autonoetic status for one alter, while having the status of knowledge or even being inaccessible to a second alter. This phenomenon appears to create difficulties for a purely representational theory and is presented to Mahr & Csibra (M&C) for their consideration.

A good test of a framework is the way in which it handles rare cases. The challenging example I wish to introduce for Mahr & Csibra's (M&C's) consideration is that of the episodic memory of individuals with dissociative identity disorder.

The *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.; *DSM-V*) diagnostic category for dissociative identity disorder (DID) has two main criteria:

> A. Disruption of identity characterized by two or more distinct personality states. …
>
> B. Recurrent gaps in the recall of everyday events, important personal information, and for traumatic events that are inconsistent with ordinary forgetting. (American Psychiatric Association 2013, p. 292)

Other criteria include the ruling out of cultural factors and general medical conditions. Any gap in the recall of everyday events is usually filled by the recall of another personality state. Thus, detail of the previous day's activities might be traced by piecing together the (non-overlapping) episodic recall of three or four alters.

With DID patients, then, the phenomenon of interest relates to what one alternative personality state (alter) knows about what happened to another alter. One experimental demonstration of this involves an alter learning 24 nouns. A second alter, who denies all knowledge of the preceding procedure, is taught a different set of nouns. A week later, without warning, the second alter is brought out and asked to follow a recognition memory test with the 48 stimuli together with distractors. Huntjens et al. (2003; 2007) found that their DID subjects responded to the words presented to the other alter as though they had previously seen them, in spite of having no recollection of the presentation. These authors conclude that "dissociators … seem to be characterised by the *belief* of being unable to recall information instead of an actual retrieval inability" (2007, p. 788, their italics). This situation, where there is no phenomenal experience of an event, but where the event is exerting a clear influence on behaviour, matches the phenomenon of post-hypnotic amnesia (e.g., Smith et al. 2013). Here, subjects claim no recollection of recent experiences which, nonetheless, affect current behaviour. Smith et al. (2013) have suggested that executive processes are responsible for controlling the initial access to material and then determine whether retrieved information is allowed into consciousness. However, material that has been accessed will exert some influence on processing even though it is not allowed into consciousness. Morton (2017) gives a similar account for the results of Huntjens et al. (2003; 2012) described above.

Using the same experimental procedure as Huntjens et al. (2003), Morton (2012; 2017) found two individuals with DID where one alter responded to the words that had been presented to another alter in exactly the same way as they responded to the control words. In other words, this material could not even be accessed by the second alter despite being a full part of the first alter's phenomenal past.

Similar results have been shown with more complex material. Reinders et al. (2003) studied DID patients who were in either a trauma-related identity state or a neutral identity state. The former generated an autobiographical traumatic memory that the latter failed to recognise as relating to themselves. These memories were contrasted to neutral memory scripts, which both states accepted as autobiographical. The two scripts were put into the third person and read in a neutral tone to the patients while they were in a scanner. The scans were similar with the neutral script for the two states, and there were only small differences between the scans of the two scripts for the neutral identity state. The big difference occurred when the trauma-related state