Episodic Memory, Simulated Future Planning, and their Evolution

Armin Schulz* & Sarah Robins

Department of Philosophy

University of Kansas

Lawrence, KS 66045

awschulz@ku.edu / skrobins@ku.edu

http://people.ku.edu/~a382s825/ / https://www.sarahkrobins.com

[* corresponding author]

Conflict of Interest - None

Episodic Memory, Simulated Future Planning, and their Evolution

---

Abstract

The pressures that led to the evolution of episodic memory have recently seen much discussion, but a fully satisfactory account of them is still lacking. We seek to make progress in this debate by taking a step backward, identifying four possible ways that episodic memory could evolve in relation to simulationist future planning—a similar and seemingly related ability. After distinguishing each of these possibilities, the paper critically discusses existing accounts of the evolution of episodic memory. It then presents a novel argument in favor of the view that episodic memory is a by-product of the evolution of simulationist future planning. The paper ends by showing that this position allows for the maintenance of the traditional view that episodic memory operates on stored memory traces, as well as explaining a number of key features of episodic memory: its being subject to frequent and systematic errors, its neural co-location with the capacity for simulationist future planning, and the potential existence of non-human episodic memory.

Keywords: episodic memory; episodic thinking; simulation; constructivism; by-product; evolutionary psychology

Episodic Memory, Simulated Future Planning, and their Evolution

1. Introduction

The pressures that led to the evolution of episodic memory (EM in what follows) have recently seen much discussion and controversy (see e.g. Mahr & Csibra, 2018; Boyer, 2008; Boyle, 2019; Schwartz, 2020). On the one hand, there is agreement on two prominent facts: (1) EM, far from being a first-personal movie of the past, is subject to frequent and systematic errors (Loftus, 1997; Loftus & Pickrell, 1995; Roediger & McDermott, 1995), and (2) EM and the capacity for "simulationist future planning" (SFP in what follows) appear to be neurally co-located (Schacter & Addis, 2007; Szpunar, Watson, & McDermott, 2007; Benoit & Schacter, 2015). On the other hand, there is no consensus as to how EM should be understood—i.e. what it *is*—or what factors influenced its evolution (Craver, 2020; Cheng & Werning, 2016; Michaelian 2016). The upshot is a somewhat confused state of the field. Indeed, this confusion is severe enough that a number of major options for the evolution of EM have not been considered. In this paper, we take steps towards remedying this situation.

We begin, in section 2, by characterizing EM, focusing especially on its relation to SFP. This then gives us the space to develop a new lay of the land concerning its evolution. Specifically, in section 3, we identify four possible ways that EM could evolve in relation to the related ability for SFP. After distinguishing each of these possibilities, we then, in section 4, present arguments in favor of one of them—namely, the view that EM is a by-product of the evolution of the psychological disposition for SFP. In section 5, we present some implications of this view and distinguish it from alternatives. We conclude in section 6.

2. Episodic Memory: What It Is

In this section, we clarify the question being asked about the evolution of EM by first making clearer how this trait should be characterized. Endel Tulving introduced the concept of EM in 1972, contrasting it with semantic memory (Tulving 1983, 1986). EM is memory for experiences; semantic memory is memory for facts. Remembering a family trip to the Grand Canyon is episodic. Remembering that the Grand Canyon is 277 miles long is semantic.[1] Tulving's distinction has had a considerable impact on the study of memory in psychology and neuroscience (see Renoult & Rugg 2020 for an overview).

However, as research on EM expanded, researchers have shifted from a focus on the distinctions between it and semantic memory to EM itself. As many have noted (e.g., Mahr & Csibra 2018), EM continues to be understood in different ways by different researchers. The most prominent understanding, also promoted by Tulving, characterizes EM as involving a particular type of awareness—what Tulving has called "autonoetic consciousness" (2002). Semantic remembering involves only noetic consciousness, awareness of what is being remembered. Episodic remembering includes autonoetic features, providing awareness of what is remembered *and* the subjective experience of the event being remembered.

Of course, this then raises the question of what exactly this kind of "autonoetic consciousness" consists in. A range of proposals are available. Some characterize autonoesis as a distinctive form of mental imagery (McCarroll 2019) or an awareness of subjective time (Hoerl 2001; Carvalho 2018). Others identify particular metacognitive feelings (Dokic 2014; Fernandez

[1] Note that the fact that a memory is self-involving crosscuts the distinction between semantic and episodic memory. I can have many memories that involve only general knowledge about my past—being born in London, for example, or growing up alongside three siblings. Such autobiographical memories (Williams et al., 2008) can include both episodic and semantic information. Similarly, I can have memory of specific events without remembering what the experience of those events was like. I can remember that I once touched stinging nettle on a hike, and that it hurt, without remembering how the pain of doing so felt at the time or which hike it was. Such event memories are often highly particular, but do not involve experiential or subjective details (Rubin & Umanath 2015).

2020) judgments (Hopkins 2014), or monitoring (Michaelian 2016) that accompany the remembered information. Rather than entering this debate, our account is guided by what is required for accommodating the two lines of empirical evidence that have prompted and guided questions about the evolution of EM.[2] These lines of evidence are not per se "explananda" of an account of the evolution of EM; rather they are empirical constraints that such an account will have to respect. We introduce them below and then explain how we use these features to set the contours of EM's autonoetic features.

2.1 False Memory

The last several decades of memory research have been devoted to the study of memory errors, and in particular the overwhelming evidence that our episodic 'memories' can be partially or fully false (Loftus & Pickrell, 1995; Loftus 2003). This evidence reveals that our memory is subject to systematic biases and easily influenced by competing sources of information (see e.g. Suddendorf & Corballis, 2007). Indeed, memory errors are easy to generate in laboratory conditions, as exemplified by prominent methods like the DRM (Roediger & McDermott 1995) and Misinformation Paradigm (Loftus 1978). It's also clear that the false memories produced in these settings resemble errors in everyday experience—swapping and omitting details, mistaking the experience of a friend or loved one for an experience of one's own, etc. Much of the subsequent theorizing about memory, in psychology and philosophy, has been focused on accounting for these errors.

[2] Of course, there are also other important features of EM—such as that they are often negative in valence or appear unbidden (Boyer, 2008). However, from an evolutionary biological perspective, the ones cited in the text are central (though see also section 5 below).

The possibility of false memories is well-established. The *pervasiveness* of such false memories, however, is not. In particular, what has not been established is how often such errors occur relative to instances of successful episodic remembering. False memories can be prevalent without being predominant. They can be easy to induce in experimental conditions without necessarily being easily induced in everyday circumstances (Gallo 2006). Indeed, some memory researchers have begun to argue more stridently for seeing these errors as the exception rather than the rule (e.g., Michaelian 2016, Mahr & Csibra, 2018).

Fortunately, settling this issue is not so important here. What is important for present purposes is, first, *that* EM is an error-prone system. Exactly *how* error-prone it is matters less than the fact that, in an inquiry about the evolutionary pressures on this system, it cannot be presumed that it produces fully accurate autonoetic representations of the past (nearly) all the time. (However, this is no different from what is the case with many other psychological traits, which tend not to operate fully accurately or reliably either—Gigerenzer & Selten, 2001.)

The second important feature of false memory research that impacts on the nature of our inquiry is that it is well-established that in many instances of false memory the error is not detectable to the rememberer herself. False EMs are often subjectively indistinguishable from genuine episodic memories (Dewhurst & Farrand, 2004, Chua et al., 2012). This constrains both the autonoetic features of EM and its plausible evolutionary explanations. First, the autonoetic features of EM cannot be accounted for by the fact that one did previously have this experience (as the experience can also occur when there is no such previous experience). Second, the value of retaining subjective experience cannot be cashed out in terms of the role of such experience in definitively guiding humans toward certainty, evidence, or truth.

2.2 Neural Overlap for Episodic Simulation

The second line of empirical evidence that impacts the discussion of EM's evolution is the well-documented discovery of the shared neural structures that support both autonoetically remembering the past and future-directed autonoetic imagination and planning (Addis et al., 2007; Szpunar et al., 2007). Researchers are increasingly interested in characterizing this distinctively autonoetic way of envisioning possible events. While it is possible to make distinctions among different forms of autonoetic future thought (Szpunar et al., 2014), doing so is not relevant here, and we will therefore refer to them collectively as *Simulationist Future Planning* (or SFP). What is relevant here is that an ever-expanding series of fMRI studies report that EM and SFP recruit the same 'core network', including the medial temporal lobes, hippocampus, retrosplenial cortex, medial prefrontal cortex, and the intraparietal lobule (Schacter et al. 2015, De Brigard et al., 2013).[3]

Many researchers have assumed the overlap between EM and SFP reveals that these two abilities are instantiations of the same psychological trait and must thereby share an evolutionary history. If correct, this would revise the evolutionary question about EM. Instead of asking why our ability to store autonoetic representations of past events evolved, we should be asking why our ability to store autonoetic representations *more generally* evolved.

However, changing the question in this way moves too quickly. Sharing a neural implementation does not make EM and SFP the same trait, nor does it compel the understanding of these two abilities as having a shared evolutionary trajectory. The fact that both the olfactory and the gustatory system employ the same neural regions and mechanisms of chemoreception

[3] Importantly, though, as Schacter et al. (2015) note, this core network is *engaged differently* by different versions of these tasks. For example, the frontopolar cortex is more active during imagining the future than during remembering (Schacter et al., 2012).

does not mean that they are the same sensory modality or that their evolutionary history is the same—neither of which is true.[4] Hence, the fact that EM and SFP recruit the same neural regions should not be taken to imply that they must be the same trait, or that their evolutionary history must be the same.

Of course, it is possible that, once these two systems and the evolutionary pressures on them are better understood, they turn out to be the same trait (as has been argued by De Brigard, 2014), or at least to have evolutionary histories that are closely intertwined. This would need to be established independently, though; the (assumed) neural overlap between these two systems does not by itself settle this question.

From an evolutionary biological perspective, therefore, the more fruitful connection between EM and SFP to be explored concerns just the fact that these two systems are both widely recognized to have autonoetic features (Addis et al., 2007; Szpunar et al., 2007, Schacter et al. 2015, De Brigard et al., 2013). When engaged in SFP, I imagine or simulate what a certain hypothetical situation would feel like. Also like EM, SFP is thus to be distinguished from the non-autonoetic representation of possible ways the world might be: when deciding whether to take my umbrella for the walk to the museum, I can consult the weather report, see that there is 25% chance of rain, note that my clothes are dry-clean only, and decide to take the umbrella. In a case like this, I do not (need to) simulate what it would be like to get caught in the rain without an umbrella; I can just consider that it may rain with a certain probability. There is no question that we often do something very much like this. However, there is also no question that we often rely on a different future planning system, which relies on the production of detailed, experiential representations of ways the world might be—the SFP (Addis et al., 2007; Szpunar et al., 2007,

[4] Note also that cases of synesthesia are interesting precisely because they bring together otherwise separate sensory modalities (Harvey 2013; Niccolai et al., 2012).

Schacter et al. 2015, De Brigard et al., 2013).[5] This is what is key here: a core feature of both EM and SFP is not that of activating and using a particular kind of information, but of activating and using information from an autonoetic perspective.

For this reason, we resist providing a detailed account of the experiential nature of EM. What matters, and thus provides the contours of our account, is just that the experiential features be such that they could also play a role in other cognitive processes like that of SFP. Exactly what this experiential quality is can be left open.[6] Put differently, it is the similarity in the kinds of representations that EM and SFP rely on that is key here. While this similarity does not, on its own, tell us how the evolutionary histories of these two abilities are related, it does imply that a joint exploration of their evolutionary history is warranted. Focusing on the potential biological role of these subjective features focuses our inquiry while also leaving open whether or how it could manifest in a broader set of organisms.[7]

3. Four Possible Evolutionary Relationships between Episodic Memory and Simulationist Future Planning

From the point of view of natural selection, there are four main ways in which EM and SFP could be related. Laying out these four ways is the aim of this section; the next section evaluates which of them is most plausible. It is useful to start with surveying the possible options, as many of them have not yet been properly characterized, recognized, or investigated.

[5] Note that, as with EM, we (of course) do not assume that these representations necessarily need to be accurate: humans cannot see in the future, and what they think will happen is subject to (more or less) systematic biases.

[6] We also note that the many ways of spelling out the details of autonoetic representations are compatible with our proposal.

[7] Note that our point here is not one of semantics or definitions. It is entirely possible to use the term "episodic memory" in a broader way to refer to first-personal, what-where-when, event-memories that may or may not have an autonoetic quality. It is just that this is not how we are using the term here: what we are interested in is investigating the evolution specifically of autonoetic memories. While we think our usage is quite in line with the literature, we are happy for readers to substitute the term "autonoetic memory" wherever we use "EM."

Before we begin, it is worth noting that evolutionary processes are complex, and have different elements. Apart from selection, the evolutionary trajectory of a trait is affected by its heritability, the structure of the population the trait is part of (e.g. whether it is divided into groups or neighborhoods), the size of the population, the genetic and epigenetic relations underlying the trait, as well as the developmental system the trait matures in. Here, though, the focus will be (largely) just on the selective value—or lack thereof—of EM and SFP.[8]

This is not because we think that these other elements of the determination of evolutionary trajectories are unimportant. Rather, it is in the spirit of such analyses of complex issues. For a full evolutionary biological account of EM and SFP, questions of heritability, population structure, etc., will need to be addressed. Such an account, however, does not need to be given in one fell swoop. It can be built up piecemeal. Filling out the remaining elements of the full account of the evolution of EM and SFP is left for a future occasion. (For a related defense of work in evolutionary psychology, see also Schulz, 2018.)

Furthermore, it is of course also true that selection pressures can change: a trait T may not be selected for until time t_0 and then become selected for feature F until time t_1, after which it becomes selected for feature G. For present purposes, though, we restrict ourselves to considering the most recent set of selection pressures only (noting the potential of divergent selective regimes where appropriate). It is also important not to confuse the selection *of* T with the selection *for* T, and neither of these with the question of whether T evolved by drift or selection. If T does not increase the expected reproductive success of its bearer, but if it is

[8] A terminological point: in the biological literature, it is common to use the term "adaptive" to refer to traits that are under selection: these are traits that increase the expected reproductive success of their bearers. However, in the literature on EM, it has become common to use the term "adaptive" to refer to the general usefulness of EM—independently of whether this usefulness has biological value. To avoid confusion, we therefore frame the discussion here without using the term "adaptive."

closely tied to another trait T' that *does* increase the expected reproductive success of its bearer, there will be selection *of* T, though no selection *for* T. In that case, the connection to T' can also imply that the evolution of T may *not* be impacted much by random, drift-like factors—despite there not being direct selection for T. Conversely, a trait that is being selected for can still be subject to many random, drift-like influences—especially in small populations.

3.1 EM and SFP as Distinct Traits with Separate Selective Histories

The first and most straightforward scenario to be considered conceives EM and SFP as distinct traits with individual selection-based evolutionary histories. On this scenario, organisms with SFP had a relatively higher fitness than those without, and the same is true for organisms with EM—but these two increases in fitness were unrelated.

So, it may have been the case that the relevant organisms faced many decision situations in which evaluating their options required close consideration of the details of each choice and its consequences. Consider, for example, an organism of this kind needing to decide whether to join a hunting party that is forming or whether to continue foraging on its own. *Simulating* these options—that is, representing them autonoetically with an SFP-system, rather than merely abstractly evaluating them—might have been the most effective way to decide what to do. In particular, this simulation may have allowed the organism to use its emotional reactions in an off-line manner as a tool for the evaluation. The organism can react to the possible scenario *as if* it were real, and then decide whether to *actually make it real* on this basis (Nichols & Stich, 2003; Picciuto & Carruthers, 2016). Assuming—not unreasonably—that the organism's emotional reactions are correlated with its biological advantage, reliance on an autonoetic SFP-system would be selected for in situations where the features that determine whether a choice is biologically advantageous depend on details that are difficult to represent and assess abstractly, or where such an abstract representation would take too long. The SFP's autonoetic nature (Addis et al., 2007; Szpunar et al., 2007, Schacter et al. 2015, De Brigard et al., 2013) enables efficient and fast decision-making in situations that need to be assessed carefully, but where such an evaluation can be done well using the organism's emotional reactions (Nichols & Stich, 2003; Picciuto & Carruthers, 2016). (We return to the details of this argument in section 4.1 below.)

Further, it may *also* have been selectively advantageous for organisms to autonoetically represent at least some of their past experiences. For example, this may have prevented them from discounting the future in a problematic, time-inconsistent manner by bringing past experiences closer to the mind of the organism (Boyer, 2008). Or, it may have allowed organisms to ascertain epistemic authority over some issues that can then be offered as reasons to others (Mahr & Cisbra, 2018). Or, autonoetically representing the past may have allowed organisms to learn from the details of their experiences long after they have taken place (Boyle, 2019).

While all of these possibilities require further elucidation and discussion—which we provide in the next section—what matters for now is just that it may have been the case that having an SFP system was selectively advantageous *and* that having an EM system was selectively advantageous, but for independent and unrelated reasons. Both of these systems may develop in the same organisms, simply because each system is selectively advantageous on its own, without there being any deep or interesting evolutionary connection between them.

Now, given that both of these systems happen to involve some of the same psychological competencies—viz., the ability to produce autonoetic representations of the world—it is unsurprising that the two systems employ some of the same neural resources. As noted earlier, this would not be the first instance of this happening: for example, it seems something similar has occurred when it comes to language and music appreciation, among other traits (Peretz et al., 2015). The fact that the EM system and the SFP system share neural resources is thus not an outlier, nor sufficient for establishing a deep (or particularly notable) evolutionary connection between these two traits. Indeed, on this scenario, the fact that humans evolved both SFP and EM is highly contingent: it is entirely conceivable that one, but not the other, of these two traits gets lost over evolutionary time, or that one, but not the other, fails to evolve in some lineages. In short, on this scenario, the evolution of EM does not have direct implications for the evolution of SFP, and vice-versa.[9]

3.2 EM is a By-Product of a Selectively Advantageous SFP

The second possibility to consider is that there was selection on organisms to make (some) decisions by relying on SFP, but that EM is a by-product of this reliance on SFP that was not itself selected for.

In this scenario, assume that there was selection on a type of organism to have an SFP system, for the reasons laid out above. That is, assume this type of organism sometimes found it selectively advantageous to simulate the experiences that are likely to result from the decision options open to it, as this allowed it to evaluate these options using its emotional reactions. Next, note that, in virtue of the fact that the SFP system functions as an off-line choice-evaluator, it gives the organism the ability to distinguish what it is *in fact* experiencing—what sounds, sights, smells, etc. it is encountering—from what it *could be* experiencing, but is not. After all, it would

[9] It is possible that there are some indirect implications though: given the fact that these two systems require some of the same psychological competences, the evolution of one can be expected to make the evolution of the second slightly more likely (see also Schulz, 2018, chap. 6). However, this does not mean that the evolution of one of them *brings with it* the evolution of the other, as on the other scenarios below.

not be selectively advantageous for the organism to act on all the simulated scenarios; the organism is only constructing these scenarios as evaluative tools (Nichols & Stich, 2003).

Furthermore, in order to make the SFP operate efficiently (or at all) the organism is bound to at least temporarily store some of these simulated scenarios. There will often be a time-delay between the organism's simulation of a future decision and when it can in fact act on that decision. The organism may also encounter similar decisions several times, making it beneficial to store simulated decisions rather than re-generating these from scratch every time. Finally, the organism may need to use temporarily stored simulations to fine-tune its emotional evaluation systems: if the world turns out to be substantially different from how it was simulated, the organism can use this divergence to change its evaluative dispositions (Glimcher et al., 2005).

This ability to store autonoetic representations that are different from the way the world is currently experienced matters, as it further implies that the organism is now also in a position to store autonoetic representations of how it in fact experienced the past. That is, since the SFP system comes with the ability to store autonoetic representation tagged as *different* from the current state of the world, organisms with such a system also have the ability to store autonoetic representations of what they *did* experience in the past but are not currently experiencing.

Importantly, this ability to store autonoetic representations of past experiences may be put into action even if there was no particular advantage to doing so. So, maybe the organism does not or cannot use stored autonoetic representations to prevent problematic discounting. Or maybe the organism does not or cannot use stored autonoetic representations to increase its epistemic authority. Or maybe the organism does not or cannot use autonoetic representations for learning.

However, the fact that the organism does not *need to* store these representations does not mean that it *will not* store them. Given that the SFP inherently comes with the storage of autonoetic representations different from the way the world is currently experienced, it is entirely possible that the organism ends up accumulating stored autonoetic representations of its actual experiences as well. That is, in virtue of the fact that the organism is storing many similar such representations as part of its SFP system already, it may end up storing autonoetic representations of the past as well. In such a case, the EM system emerges as a by-product of the SFP system.

Of course, if such storage comes with major costs, natural selection would push for its cessation. Similarly, if this storage is not selectively advantageous, we would expect it to become corrupted sooner or later. However, both of these possibilities can take significant periods of time to materialize. Until this happens, the relevant organisms would have an EM that is merely a non-selected by-product of a selected-for SFP system.

3.3 SPF as a By-Product of a Selectively Advantageous EM

The third case reverses the relationship from the previous scenario. Here it is supposed that there was selection for EM, but that SFP is just a non-selected by-product of this reliance on EM.

So, assume that there was selection on a type of organism to have an EM system, for some of the reasons laid out in the first scenario presented. That is, assume this type of organism sometimes found it selectively advantageous to store autonoetic representations of the past, as this allowed it to avoid problematic, temporally-inconsistent discounting of the future, or because this storage of autonoetic representations of the past allowed it to increase its epistemic authority, or because it allowed the organism to learn from its past experiences long after these experiences have taken place (or a combination of these reasons). Next, note that, since EM is *memory*, the organism cannot straightforwardly assume that these EM-produced autonoetic representations still match the world as it is now. There may be many aspects of the world that are unchanged, but there are also likely to be many that now differ—and some drastically. The organism needs to be able to produce autonoetic representations about what the world is actually like—i.e. representations of what it is actually experiencing now—as well as autonoetic representations about what the world *was* like, and then keep these two apart from each other.

Given this, though, it is then possible that, as the organism makes decisions about how to interact with its environment, it starts producing autonoetic representations of what would be the case if it did this or that, even if this does not have a selective value per se. So, while it may be true that its decision making is not biologically enhanced by *simulating* the decision options—perhaps there are quicker ways of evaluating the decision options, or perhaps the organism's emotional reactions are not triggered well or at all by simulated scenarios—the organism might still use its EM-derived autonoetic representational abilities to generate these kinds of simulations. While these simulations are not actually helpful for the organism in making its decisions, they are a natural outgrowth of the fact that the organism needs to consider ways the world might be. Given its dependence on EM, the consideration of ways the world might be could simply trigger the autonoetic representation of the relevant scenarios, even if there is no need to or advantage in doing so. In this case, therefore, the organism has an SFP system, but this system evolved just as a non-selected by-product of the selected-for EM system.

Of course, as before, if the production of autonoetic representations of ways the world might be comes with costs, natural selection should be expected to push for its cessation. Similarly, if the SFP system plays no functional role for the organism, we would expect it to become corrupted sooner or later. In the time before either of these options develops, however, the relevant organisms would have an SFP system merely as a non-selected by-product of a selected-for EM system.

3.4 EM and SFP as Selectively Neutral

The final possibility is that EM and SFP are *both* non-selected traits, or non-selected aspects of some other trait. This could be for several different reasons.

On the one hand, EM and SFP could just be by-products of some other trait without having been under direct selection themselves. For example, it is possible that, once brains get sufficiently complex, a general form of consciousness evolves (Hasker, 1999). Aspects of this kind of consciousness could be or lead to the autonoetic representation of aspects of the organism's past and potential future behaviors (and some combination thereof), without either EM or SFP being selectively advantageous in and of themselves.[10] On the other hand, it could also be that both EM and SFP independently evolved purely by drift, or that one of these two traits evolved by drift, and led to the other as a by-product as on scenarios 2 and 3 above. In any of these scenarios, neither SFP nor EM has been under direct selection.

Note that, as before, if these traits come with costs, they would be expected to be lost in the future, and even if not, there is a chance that they would get corrupted sooner or later. Also, note that it is possible that one or both of them would become selectively advantageous at a future point in time. Until this happens, though, both of these traits should be seen as non-selected traits.

[10] The debate surrounding the nature of consciousness is famously complex and without a resolution. Here, though, we do not take any position on this debate, and just note that the idea that EM and SFP might be aspects of a generally non-selectively advantageous form of consciousness is a possibility to consider—independently of exactly what *consciousness* is.

In sum: EM and SFP may have evolved independently—selectively or not—or the evolution—selective or not—of one may have necessarily led to the evolution of other. Laying out these four possible evolutionary scenarios for EM and SFP brings with it a method by which to determine the most plausible amongst them. To sort between these options, the selective value of EM and of SFP must be considered individually. If there is reason to doubt that EM was selected for, this calls into question options 1 and 3. If there is reason to doubt that SFP was selected for, then options 1 and 2 lose plausibility. If there is reason to presume at least one of SFP or EM was selected for, this rules out option 4.

4. An Argument for EM as a By-Product of a Selectively Advantageous SFP

Of the four evolutionary scenarios laid out in the previous section, the second is most plausible—at least when it comes to humans. To show this, we proceed in two steps: first, we show that there are reasons to think that, at least in humans, SFP *is* likely to have been selected for, and second, we show that EM is likely *not* to have been selected for.

4.1 The Selective Value of Simulationist Future Planning

In humans at least, it is plausible that SFP was selected for. This is so for two reasons. First, humans develop and live in environments of a distinctively social kind. Humans need to not just keep track of what other organisms do, but also what these others organisms think, want, and feel (Byrne & Whiten, 1997; Sterelny, 2003; Henrich, 2015, Schulz, 2018, 2020). This makes human environments complex to navigate: the details of the consequences of the available decision options matter greatly for their evaluation.

For example, it may be that it does not just matter if action A makes conspecific C1 angry, but it matters exactly how C1 looked when it got angry (*who* it was angry with, and *how* angry was it), while keeping track of *exactly how* C2 smiled (Was it a sign of being put in control? Or was it an expression of happiness for someone else?). Moreover, giving appropriate weight to C1's anger and its potential consequences—as opposed to, say, the weather at the time—may be best ensured by simulating its occurrence (rather than just supposing it occurs). Similarly, it may be that person A's joining a hunting party is not always selectively advantageous, and depends on whether conspecific B is also part of the hunting party—but only if A and B are sufficiently socially and psychologically aligned. Are A and B sufficiently well supported by the rest of the community to make their participation in the hunting party smooth and non-disruptive? However, whether the latter is the case depends on a myriad of details that can differ from case to case: it depends on how A and B have interacted with each other (and the group as a whole) in the past, and on how they and others expect each other to behave in the future. Whether it is advantageous going forward may change after each hunting trip.

In turn, this often makes it difficult to rely on hard and fast rules about how to react to a given situation (Sterelny, 2003; Schulz, 2018). It is often more selectively advantageous for organisms to think through and evaluate each option individually and in turn (Schulz, 2018). More generally, in the kind of complex social environments in which humans evolved, simple heuristic rules are unlikely to be selectively advantageous. Instead, the best way of dealing with these environments is by using time, concentration, and attention to evaluate the details of the given decision options in light of a very abstract decision rule such as "Do what makes you happy" (Schulz, 2018; Sterelny, 2003). Hence, at least when it comes to human social living, the specific features of the individual decision options matter greatly, and need to be taken into account as such for humans to interact with each other in ways that are selectively advantageous.

The second reason for why the SFP system plausibly was selectively advantageous in human evolutionary history is that in humans (as in many other organisms), it is plausible to think that emotional reactions are a good guide to biological fitness. In order to react biologically appropriately to a given situation, organisms might need to engage in a whole host of physiological, behavioral, and psychological changes. They might need to attend to certain aspects of their sensory experiences (a specific type of sound, say), they might need to ready their body for fast movement (e.g. by increasing their heart rate), and they might need to recall specific information (such as the frequency of rain at this time of year). Emotional reactions are useful, as they initiate and coordinate this wide set of responses. Indeed, it is widely agreed that the reason why organisms have emotions in the first place is that the latter bring together a wide set of bodily, behavioral, or psychological changes so as to enable the organism to respond biologically appropriately to a given situation (Tooby & Cosmides, 2008; Al-Shawaf et al., 2015, LeDoux, 2012).[11]

Note that emotions need not be perfectly correlated with biological fitness for them to play this role. All that is needed is that they are sufficiently positively correlated with biological fitness to make them a useful guide to biologically advantageous ways of acting in that scenario. Of course, for a full account of the evolution of emotions, the required degree of correlation would need to be made precise. For present purposes, it is enough that it is reasonable that there

[11] This point is independent of the controversy surrounding the existence of basic emotions (Ekman, 1989; Izard, 2011, Fridlund, 1994), or of the nature of emotions (LeDoux, 2012). Whatever exactly emotions are and exactly how much they are impacted by cultural factors, selection for emotions is widely thought to, at a minimum, flag to the organism which situations to avoid or approach (the *valence* part of the emotion) and—perhaps—also how to approach them (the affect program or *content* part of the emotion). This is all that matters here.

is *some* such correlation: what matters for the inquiry into the evolutionary pressures on the SFP is that it is plausible and widely accepted that emotional reactions to many biologically important scenarios are reasonably closely tethered to the selectively appropriate ways of responding to these scenarios.

Among humans, it is furthermore plausible that we should expect social scenarios to be among the ones to which emotional reactions are well tailored (Fessler, 2010; Al-Shawaf et al., 2015). Given the importance of the social environment for human living, social situations are a prime candidate for the kinds of cases in which emotional responses are well correlated with biologically appropriate behaviors.

Because of these two points—the selective vale of attention to detail in the evaluation of social decisions and the selective value of emotional responses—the foundations of the argument for the selective value of SFP sketched in the previous section are met. For humans (at least), there likely *have been* important decision situations in which the evaluation of the options required close consideration of the details of the consequences of these choices: namely, social decisions (i.e. decisions about how to interact with others in their social group). Furthermore, it is plausible that this kind of evaluation is especially efficiently done by *simulating* the decision options. Since humans already have a system in place that allows them to determine which situations to avoid or approach—their emotional system—they are well advised to use this system to evaluate a number of complex decision options (see also Schulz, 2011). That is, in humans, the virtual, autonoetic evaluation of decision options is selected for due to its being biologically advantageous for humans (a) to rely on their emotional responses to react to their *actual* social environment, and (b) to assess social decisions by attending to the details of the available choices.

All in all, therefore: there are good reasons to think that the SFP system was, in fact, selected for in humans. Hence, this suggests that scenarios 3 and 4 above—where SFP is just a non-selected by-product of EM or some other trait—are not plausible at least for humans. However, this leaves scenarios 1 and 2 open still.

4.2 Episodic Memory Was Not Selected For

To see why EM is unlikely to have been selected for, it is useful to begin by noting that this system has some surprising features. EM produces representations of exceptional richness, but these representations are about highly specific events, often at a great temporal distance from the time at which they are represented. This means many of these representations are not straightforwardly useful for navigating the *current* environment.

To see this, recall the three major accounts of the evolution of EM in the literature sketched above: the view that EM evolved to help humans avoid the detrimental consequences of hyperbolic discounting (Boyer, 2008), the view that EM evolved as a way of ascertaining epistemic authority over some issues that can then be offered as reasons to others (Mahr & Csibra, 2018), and the view that EM makes it possible to learn something from experiential sources that have long passed (Boyle, 2019). Each of these accounts faces major problems that stem from the remoteness of EM representations.

When considering Boyer's (2008) account, it first needs to be noted that it often *is* selectively valuable to discount the future (Soman et al., 2005). In an uncertain world, being biased towards present enjoyment is biologically advantageous. The problem is only with some kinds of discounting: namely, hyperbolic ones, which can lead to temporally inconsistent choices. For Boyer's account to work, therefore, it needs to be the case that EM does not simply prevent humans from discounting the future by bringing the present closer to the past—but that it does so in an extremely fine-tuned manner that affects the *rate* at which the future is discounted only. It is not clear how this might work (and Boyer, 2008, does not make it clearer).

Second, Boyer's proposal requires that EM is closely tagged to a time: to reliably avoid hyperbolic updating, the same event would need to be represented differently—with different degrees of vividness, say—depending on how long ago it was. There is no indication that human EM actually has this feature, nor any proposal for how this resource-dense continuous updating would be supported (much less advantageous).

Third, and perhaps most persuasively, evidence from the amnesia patient KC indicates that it is possible to retain temporal discounting abilities in the absence of EM. KC was a neuropsychological patient with profound episodic memory loss as a result of a motorcycle accident. He has retained much of his semantic memory and general cognitive abilities, but has effectively no autonoetic representations of his past experiences. Nonetheless, KC seems to have a rich understanding of time and is susceptible to the same ways of discounting the future as others who possess EM (Kwan et al. 2012; 2013).

As far as Mahr & Csibra's (2018) account is concerned, many issues with the proposal have been pointed out in the comments published with the main essay. Here, we restrict ourselves to making two points. First, if the purpose of EM is to generate epistemic authority that can be used to support reason-giving practices, we would expect EM to be largely accurate—which, as noted earlier, appears false (Robins, 2018).

Second, it is not at all clear that the reason-giving practices that people actually engage in match what Mahr & Csibra (2018) claim. That is, it is not obvious that people only offer reasons for things that they can episodically remember doing, or that these are the reasons found most compelling. It is true is that humans evolved in an inherently social environment, and—as just noted—it is also true that it is plausible that the human SFP system evolved in response to the pressures generated by this social environment. However, there is no good reason to think that this will translate directly into the reason-giving practices in which people engage with their peers. People's EM's may be biased, they are inherently perspectival, and they are limited in extent and accuracy. It is not obvious that they make for good epistemic reasons (cf. the fact that witness testimony is a famously problematic sort of legal evidence). In short: the extent to which epistemically normative reasoning matches up with the people's communications surrounding their EM's is highly unclear (at best).

Finally, as far as the account of Boyle (2019) is concerned, recall that, according to this account, rich autonoetic representations of the past can help us learn useful things long after an experience. Suppose I try a strategy for storing food and it doesn't work and I have no idea why. However, if I keep a representation of this experience around, then when I later observe something about food preservation in another context, I can revisit this representation and learn something from it—something that I can then use to guide future decision-making.

This account is unconvincing, for two reasons. First, at least when it comes to humans—the prime focus of EM-using organisms—many of the relevant environments change quickly. After all, how should my reaction to seeing the Grand Canyon for the first time as a five-year-old be relevant to my decision-making now? My cognitive, physical, and social situation is completely changed compared to when I was five. So, in order to be selectively advantageous, EM would need to *only* be operative in cases where the past is a sufficiently useful guide to the future. Quite apart from the fact that it is not clear how humans (or any other organisms) could solve this problem—which is effectively the problem of induction—this focused form of EM is empirically implausible. People seem to episodically remember things that seem quite clearly not a good basis for future learning just as much things that are valuable for learning.

Second and most importantly, Boyle's (2019) argument at most supports the selective value of a *detailed* form of long-term memory. Even assuming it is biologically valuable to store representations of the past to learn from them in the future, it is not clear why these representations would need to be autonoetic. That is, why can't I just remember that I went to Grand Canyon at age 5, that the weather was sunny, etc. Why would humans (and other organisms) find it selectively advantageous to *autonoetically represent* this information? This, though, is exactly what needs to be answered here: as noted in section II, the issue to explain when it comes to the evolution is EM is precisely why a system producing autonoetic representations of the past evolved—not merely why a system producing *detailed* representations of the past evolved.

Note that this situation is quite different from that in the case of SFP. In the latter, autonoetic representation helps with the *evaluation* of decision situations. In the case of Boyle's (2019) defense of the selective value of EM, though, this is not the case—an appeal to emotional responses to the past is not made. This is not surprising, since the past cannot be affected now: organisms don't need to make decisions as to what pasts they should have brought about. Hence, the autonoetic nature of EM—unlike that of SFP—is not well explained by Boyle (2019)'s argument. Note that this point of course does not preclude the possibility that EM, once it has evolved, could not, at times, be used to learn from past experiences. Our point is just that learning from the past is not well seen as a selective pressure on EM. (Compare: once humans evolved the ability to domesticate plants, they could sometimes use this ability to signal status or group membership—e.g. by making jack-o-lanterns or planting decorative gardens. However, the latter were not major selective pressures on the domestication of plants to begin with.)[12]

More generally, we do not think that the failures of the three accounts of EM's supposed biological value are surprising. The problem is, quite simply, that it is difficult to see what biological function EM *could* have. Situating the question of its selection alongside SFP, for which the possible selective advantages are more straightforward, makes the point especially clear. Given its rich autonoetic and specific nature about temporally remote events, EM is an excellent candidate for being a by-product of SFP. Hence, the fact that various proposed accounts of the biological value of EM fail to be convincing is actually to be expected.

All in all, therefore, we consider scenario 2—i.e. the view that EM is a non-selected byproduct of a selected-for SFP system—the most plausible hypothesis about the evolution of EM and SFP. However, to fully understand this view, it is important to be clear about what implications it has for the workings of EM and SFP—and what implications it does not have. Bringing this out is the aim of the next section.

5. Implications

Our proposal that EM is a non-selected byproduct of a selected-for SFP system has a range of implications for how EM and its features are understood, which provides further support for this scenario. These implications are worth noting for their own sake, but they also serve to add further contrast between our account and those presently available in the literature.

[12] Relatedly, our argument does not imply that loss of the EM is not at all detrimental for humans now. That said, the issues here are complex. It is true that persons with various forms of dementia often experience significant reductions in their autonomy and quality of life. However, it is not clear what this means for the issues at stake here, as it is far from clear to what extent these cases involve selective loss of the EM specifically, rather than loss of memory abilities or SFP more generally.

5.1 The Explanation of the (Sometime) Inaccuracy of EM

As we discussed earlier, concerns about how to best explain the inaccuracies of EM have played an important role in motivating the discussion of EM's evolution. Our account provides an explanation for why EM is frequently inaccurate and unreliable. Given that this system was not itself selected for, organisms cannot be assumed to have evolved mechanisms that ensure EM accuracy. Recall also that our proposal for how EM might emerge from SFP involved the incidental storing of simulations on which the organism may or may not have acted—thus predicting the existence of "false EM's." Our account can thus explain the (sometime) inaccuracy of the EM system, which is otherwise quite puzzling—and does so in ways that are importantly different from other accounts.

So, unlike De Brigard (2014), we do not infer the lack of selection for EM *from* the fact that it is currently producing errors. Rather, we infer the lack of selection for EM from other reasons—viz. its costly autonoetic representational richness that lacks a compelling countervailing benefit—and *use* this fact to explain why EM is error-prone now.

This matters, as the inference from EM's current error-prone state to its not having been selected for is problematic. One the one hand, as Millikan (1984) has noted, an ability can be selectively advantageous even if it only rarely succeeds (a point acknowledged by De Brigard, 2014). On the other hand, as Schwartz (2020) argues, there is no necessary connection between current trends toward memory errors and the survival value of EM. Given that the evolutionary conditions during which EM was created may not now be in operation, errors detected now need not be seen as strong evidence regarding the role of errors in shaping the initial ability. It is thus important to note that nothing in our analysis of the evolution of EM relies on its current rate of successful remembering. Indeed, the fact that our account can *explain* the fact that EM is error-prone—rather than *building this into* the foundations of the account—is one of its key advantages.

In this way, our account provides an important middle ground between accounts that are built around EM's lack of reliability and accounts developed in opposition to this idea (e.g., Michaelian 2016). Debates between these two accounts are often mired in discussions of which notion of reliability to use and how it should be calculated (Robins 2019; Michaelian 2020). Our account makes it possible to sidestep these concerns.

5.2 EM Is Not Purely Constructive

Second, our approach makes it possible to acknowledge the errors involved in EM without endorsing a purely constructive account of its operation. Many theories of EM now characterize this ability as *constructive*—a system that builds plausible representations of past events "on the fly" rather than storing representations of past events in the memory system (Michaelian 2016; De Brigard 2014; Sant'Anna, 2020). Constructive accounts have grown in popularity in response to the perceived need to explain the kind of memory errors identified just above and additional empirical evidence demonstrating the influence of the retrieval context on the representations produced in the act of remembering (Robins 2016).

Purely constructive accounts encounter difficulties, though, because while EM is sometimes inaccurate and unreliable, this is not always the case. There are numerous instances in which EM produces accurate representations of past experiences and, in many of these cases, where those experiences are unique enough that the information could only derive from that experience. The best explanation of such cases is that the information is stored in EM. And so it must be the case that EM can store information from past experiences—i.e., that remembering is not merely the construction of possible past scenarios but, at least on some occasions, involves information retained from the prior experience and is not derived from construction alone (Robins 2016; 2019). Purely constructive accounts are limited in their ability to explain this range of human EM performance—and insofar as the alternative proposals for understanding the evolution of EM compel this purely constructive view, this provides additional reason to favor our proposal.

By taking seriously the possibility that EM is simply a byproduct of SFP, our account illustrates how it is possible to retain the commitment to EM as a system involving informational memory traces, while avoiding worries as to why such a system of EM storage could have been selected for.

5.3 EM Could Be a Separate Trait

Our account leaves open the possibility that EM is a separate trait from SFP. That is, we do not require EM to be a *part* of SFP; a byproduct can be a separate trait. This marks an important distinction between our proposal and others, which have worked to subsume EM and SFP under the same overall ability of episodic simulation.

Leaving it open whether EM and SFP are the same trait allows for, and even encourages, further work in this area.[13] This strikes us as especially important given that ongoing research into the neural overlap between the brain networks involved in EM and SFP is increasingly dedicated to identifying subtle but important differences between these cognitive activities, particularly as more forms of SFP are added to the list (Szpunar, Shrikanth, & Schacter, 2018).

[13] Note that the individuation of biological and psychological traits is difficult theoretically, too (Baum, 2013). Fortunately—and for the same reasons set out in the text—settling this is not necessary here either.

For example, both activities can vary in the amount of detail they involve, which impact performance in generating representations of either kind. Moreover, researchers are also investigating differences in how EM and SFP are engaged at different points in the lifespan (Madore, Jing, & Schacter, 2016), as well as individual differences in the reliance on SFP (Beaty et al. 2018; Beaty et al., 2019).

While it is not yet clear whether these differences between SFP and EM will prove consequential for the ultimate consideration of the two as a single trait, given the range of differences already documented, it seems prudent to leave the options open.

5.4 Animal EM

Finally, our account provides novel inroads into the investigation of the existence of EM in at least some non-human animals.[14] Much of the existing work in this area begins from the assumption that EM capacities are selectively advantageous. However, this work has struggled to establish which animals have EM and why (e.g., Allen & Fortin, 2013). Our account can explain these problems: the grounding assumption of this argument is false. Determining whether an organism has EM capacities should be done without taking these capacities to be selectively advantageous.

A more compelling approach to this issue starts from the assumption that, since EM is tied to the workings of the SFP, any organism that has evolved the latter is likely to have evolved EM as well (see also Hasselmo 2012). Given that, as noted earlier, the evolution of an SFP is favored in complex social environments, we would thus expect the evolution of something resembling EM in animals with larger social groups or amongst those that seem to engage in

[14] This goes against the suggestion of Tulving (2002, p. 7) and Suddendorf & Corballis (2007).

more planning skills for other reasons. While this is a prediction that it is difficult to confirm at present, we think it is something that deserves taken very seriously.

On top of this, our view offers ways to mitigate a range of further challenges which have plagued the exploration of EM in non-human animals. For instance, the characterization of EM as involving autonoetic consciousness has stymied research because of the inability to tie this form of consciousness to particular animal behaviors or objective characteristics. While SFP shares this autonoetic character with EM—and so, in some respects, continues to be susceptible to this concern—it is an easier capacity to investigate in non-human animals. SFP can occur and be useful in a more specific range of contexts in comparison to EM. Decision-making experimental frameworks are much more easily converted to animal models than many of the existing frameworks used for testing EM (which, for instance, are often based on verbal commands). In this way, our account promises to make the further exploration of EM in non-human animals easier.

6. Conclusion

We have argued that four scenarios surrounding the evolution of EM—the ability to produce autonoetic representations of past events—and SFP—the ability to produce autonoetic representations of ways the world might be—should be distinguished. EM and SFP could have independent selective histories, EM could be an unselected by-product of a selected-for SFP, SFP could be an unselected by-product of a selected-for EM, or they could both be unselected traits or byproducts of another trait. We have further argued that these four options have not been clearly distinguished in the literature thus far, and that the second scenario, according to which EM is just an unselected by-product of a selected-for SFP is the most plausible one: at least for the kinds of social organisms that humans are, the SFP plausibly is selectively advantageous, but the extreme specificity and representational richness of EM make it unlikely to have a selective value. We have then noted that this account (a) provides an explanation for why EM is frequently unreliable and inaccurate, (b) still allows for EM's to not be fully constructed on the fly, but at least sometimes be based on stored trace information from the past, and (c) allows EM to be a separate trait of its own. Our account also (d) predicts and explains the "reminiscence bump", and (e) predicts that EM may also be found in social non-human animals. All in all, we thus hope to have clarified the evolutionary relationships between EM and SFP—and provided a stepping stone towards the better understanding of both of these traits.

References

Addis, D.R., A.T. Wong, and D.L. Schacter. 2007. Remembering the past and imagining the future: Common and distinct neural substrates during event construction and elaboration. *Neuropsychologia* 45: 1363– 1377.

Allen, T.A., & Fortin, N.J. (2013). The evolution of episodic memory. *Proceedings of the National Academy of Sciences*, 110, 10379-10386.

Al-Shawaf, L., Conroy-Beam, D., Asao, K., & Buss, D. M. (2015). Human Emotions: An Evolutionary Psychological Perspective. *Emotion Review*.

Baum, D. A. (2013). Developmental Causation and the Problem of Homology. *Philosophy and Theory in Biology, 5*. doi:http://dx.doi.org/10.3998/ptb.6959004.0005.003

Beaty, R.E., Chen, Q., Christensen, A., Jiang, Q., Silvia, P., & Schacter, D.L. (2018). Brain networks of the imaginative mind: Dynamic functional connectivity of default and cognitive control networks relates to openness to experience. *Human Brain Mapping, 39,* 811-821.

Beaty, R.E., Seli, P., & Schacter, D.L. (2019). Thinking about the past and future in daily life: An experience sampling study of individual differences in mental time travel. *Psychological Research*, *83*, 805-816.

Benoit, R.G., and D.L. Schacter. 2015. Specifying the core network supporting episodic simulation and episodic memory by activation likelihood estimation. *Neuropsychologia* 75: 450–457.

Boyer, P. (2008). Evolutionary economics of mental time travel? *Trends in Cognitive Science, 12,* 219-224.

Boyle, A. (2019). Learning from the past: Epistemic generativity and the function of episodic memory. *Journal of Consciousness Studies, 26*, 242-251.

Whiten, A., & Byrne, R. W. (Eds.). (1997). *Machiavellian Intelligence II: Extensions and Evaluations*. Cambridge: Cambridge University Press.

Carvalho, F.M. (2018). Situating mental time travel in the broad context of temporal cognitive: A neural systems approach. *Filosfia Unisinos, 19,* 81-88.

Cheng, S. & Werning, M. (2016). What is episodic memory if it is a natural kind? *Synthese, 193*, 1345-1385.

Chua, E.F., Hannula, D.E., & Ranganath, C. (2012). Distinguishing highly confident accurate and inaccurate memory: insights about relevant and irrelevant influences on memory confidence. *Memory, 20*, 48–62.

Conway, M.A., Wang, Q., Hanyu, Z, & Haque, S. (2005). A cross-cultural investigation of autobiographical memory: on the universality and cultural variation of the reminiscence bump. *Journal of Cross-Cultural Psychology,* 36: 739-749.

Craver, C.F. (2020). Remembering: epistemic and empirical. *Review of Philosophy of Psychology, 11,* 261-281.

De Brigard, F. (2014). Is memory for remembering? Recollection as a form of episodic hypothetical thinking. *Synthese*. 191(2): 155-185.

De Brigard, F., Addis, D., Ford, J.H., Schacter, D.L., & Giovanello, K.S. (2013). Remembering what could have happened: Neural correlates of episodic counterfactual thinking. *Neuropsychologia* 51: 2401–2414.

Dewhurst, S., & Farrand, P. (2004). Investigating the phenomenological characteristics of false recognition for categorized words. *European Journal of Cognitive Psychology, 16*, 403-

416.

Dokic, J. (2014). Feeling the past: A two-tiered account of episodic memory. *Review of Philosophy and Psychology, 5,* 413-426.

Ekman, P. (1989). The Argument and Evidence About Universals in Facial Expressions of Emotion. In H. Wagner & A. Manstead (Eds.), *Handbook of Social Psychophysiology* (pp. 143-164). New York: Wiley.

Fernandez, J. (2020). Self-referential memory and mental time travel. *Review of Philosophy and Psychology, 11,* 283-300.

Fessler, D. M. T. (2010). Madmen: An Evolutionary Perspective on Anger and Men's Violent Responses to Transgression. In M. Potegal, G. Stemmler & C. Spielberger (Eds.), *International Handbook of Anger* (pp. 361-381): Springer New York.

Fridlund, A. J. (1994). *Human Facial Expression: An Evolutionary View*. San Diego: Academic Press.

Gallo, D. A. (2006). *Associative Illusions of Memory: False Memory Research in DRM and Related Tasks*. New York, NY: Taylor & Francis.

Gigerenzer, G., & Selten, R. (Eds.). (2001). *Bounded Rationality: The Adaptive Toolbox*. Cambridge, MA: MIT Press.

Glimcher, P. W., Dorris, M. C., & Bayer, H. M. (2005). Physiological utility theory and the neuroeconomics of choice. *Games Econ Behav, 52*(2), 213-256.

Gluck, J., & Bluck, S. (2007). Looking back across the life span: a life story account of the reminiscence bump. *Memory & Cognition, 35*, 1928-1939.

Godfrey-Smith, P. (2001). Three kinds of adaptationism. In S. H. Orzack & E. Sober (Eds.), *Adaptationism and Optimality* (pp. 335-357). Cambridge: Cambridge University Press.

Harvey, J.P. (2013). Sensory perception: lessons from synesthesia. *Yale Journal of Biology and Medicine, 86*, 203-216.

Hasker, W. (1999). *The Emergent Self*. Ithaca: Cornell University Press.

Hasselmo, M.E. (2012). *How We Remember: Brain Mechanisms of Episodic Memory*. MIT Press.

Henrich, J. (2015). *The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter*. Princeton, NJ: Princeton University Press.

Hoerl, C. (2001). The phenomenology of episodic recall. In *Time and memory. Issues in philosophy and psychology*, eds. Hoerl, C., and T. McCormack, 315–335. Oxford: Oxford University Press.

Hopkins, R. (2014). Episodic memory as representing the past to oneself. *Review of Philosophy and Psychology, 5,* 313-331.

Izard, C. (2011). Forms and Functions of Emotions: Matters of Emotion-Cognition Interactions. *Emotion Review, 3*(4), 371-378.

Kwan, D., Craver, C.F., Green, L., Myerson, J., Rosenbaum, S. R. (2013). Dissociations in Future Thinking Following Hippocampal Damage: Evidence from Discounting and Time Perspective in Episodic Amnesia. *Journal of Experimental Psychology, General. 142*: 1355.

Kwan, D., Craver, C.F., Green, L., Myerson, J., Boyer, P., Rosenbaum, R.S. (2012). Future Decision-Making without Mental Time Travel. *Hippocampus* 22: 1215-19.

LeDoux, J. E. (2012). Evolution of human emotion: a view through fear. *Progress in brain research, 195*, 431-442.

Loftus, E.F. 1997. Creating false memories. *Scientific American* 277: 70–75.

Loftus, E.F., and J.E. Pickrell. 1995. The formation of false memories. *Psychiatric Annals* 25: 720–725.

Madore, K.P, Szpunar, K.K., & Schacter, D.L. (2016). Divergent creative thinking in young and older adults: Extending the effects of an episodic specificity induction. *Memory & Cognition, 44,* 974-988.

Mahr, J.B. & Csibra, G. (2018). Why do we remember? The communicative function of episodic memory, *Behavioral and Brain Sciences, 41*, e1.

McCarroll, C.J. (2019). Navigating intertemporal choices: Mental time travel, perspectival imagery, and prudent decision-making. *Psychology of Consciousness, Theory, Research, and Practice, 6,* 200-213.

Michaelian, K (2016). *Mental Time Travel: Episodic Memory and our knowledge of the personal past.* Cambridge, MA: MIT Press.

Michaelian, K. (2020). Confabulating as Unreliable Imagining: In Defence of the Simulationist Account of Unsuccessful Remembering. *Topoi, 39*, 133-148.

Millikan, R. (1984). *Language, Thought, and Other Biological Categories*. Cambridge, MA: MIT Press.

Niccolai, V., Jennes, J., Stoerig, P., & Van Leeuwen, T.M. (2012). Modality and Variability of Synesthetic Experience. *American Journal of Psychology, 125,* 81-94.

Nichols, S., & Stich, S. (2003). *Mindreading. An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds* (Vol. 114): Oxford University Press.

Peretz, I., Vuvan, D., Lagrois, M.-É., & Armony, J. L. (2015). Neural overlap in processing music and speech. *Philos Trans R Soc Lond B Biol Sci, 370*(1664), 20140090-20140090.

Pertseva, M. (1991). The evolution of hormonal signalling systems. *Comparative Biochemistry and Physiology Part A: Physiology, 100*(4), 775-787.

Picciuto, E., & Carruthers, P. (2016). Imagination and Pretense. In A. Kind & P. Kung (Eds.), *Knowledge Through Imagination* (pp. 314-325). Oxford: Oxford University Press.

Renoult, L. & Rugg, M.D. (2020). An historical perspective on Endel Tulving's episodic-semantic distinction. *Neuropsychologia, 139,* doi: 10.1016/j.neuropsychologia.2020.107366.

Robins, S. K. (2016). Misremembering. *Philosophical Psychology, 29,* 432–447.

Robins, S.K. (2018). Confabulation and Epistemic Authority. *Behavioral and Brain Sciences*.

Robins, S.K. (2019). Confabulation and Constructive Memory. *Synthese,* 196, 2135–2151.

Roediger, H.L., and K.B. McDermott. (1995). Creating false memories: Remembering words that were not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition 21*: 803–814.

Rubin, D.C. & Umanath, S. (2015). Event memory: A theory of memory for laboratory, autobiographical, and fictional events. *Psychological Review, 122,* 1-23.

Sant'Anna, A. (2020). The hybrid contents of memory. *Synthese, 197,* 1263-1290.

Schacter, D. L. & Addis, D. R. (2007). On the constructive episodic simulation of past and future events. *Behavioral & Brain Sciences, 30*, 299-351.

Schacter, D.L., Addis, D.R., Hassabis, D., Martin, V.C., Spreng, R.N. & Szpunar, K.K. (2012). The future of memory: Remembering, imagining, and the brain. *Neuron, 76*, 677-694.

Schacter, D.L., Benoit, R.G., De Brigard, F., & Szpunar, K.K. (2015). Episodic future thinking and episodic counterfactual thinking: Intersections between memory and decisions. *Neurobiology of Learning and Memory.*

Schulz, A. (2011). Simulation, simplicity, and selection: an evolutionary perspective on high-level mindreading. *Philosophical Studies, 152*(2), 271-285.

Schulz, A. (2018). *Efficient Cognition: The Evolution of Representational Decision Making*. Cambridge, MA: MIT Press.

Schulz, A. (2020). Enhancing Thoughts: Culture, Technology, and the Evolution of Human Cognitive Uniqueness. *Mind & Language*, DOI: 10.1111/mila.12320

Schwartz, A. (2020) “Simulationism and the Function(s) of Episodic Memory” *Review of Philosophy and Psychology*, DOI: 10.1007/s13164-020-00461-1

Soman, D., Ainslie, G., Frederick, S., Li, X., Lynch, J., Moreau, P., . . . Zauberman, G. (2005). The Psychology of Intertemporal Discounting: Why are Distant Events Valued Differently from Proximal Ones? *Marketing Letters, 16*(3), 347-360.

Sterelny, K. (2003). *Thought in a Hostile World: The Evolution of Human Cognition*. Oxford: Wiley-Blackwell.

Szpunar, K.K., Shrikanth, S., & Schacter, D.L. (2018). Varieties of future thinking. In G. Oettingen, A.T. Sevincer, & P.M. Gollwitzer (eds.) *The Psychology of Thinking About the Future*. New York: Guilford Press.

Szpunar, K.K., Spreng, R.N., & Schacter, D.L. (2014). A taxonomy of prospection: Introducing an organizational framework for future-oriented cognition. *Proceedings of the National Academy of Sciences USA, 111,* 18414-18421.

Szpunar, K.K., Watson, J.M., & McDermott, K.B. (2007). Neural Substrates of Envisioning the Future. *Proceedings of the National Academy of Sciences, 104*, 642-647

Suddendorf T., Corballis M.C. (2007). The evolution of foresight: what is mental time travel and is it unique to humans? *Behavioral and Brain Sciences*, 30, 299–313.

Tooby, J., & Cosmides, L. (2008). The Evolutionary Psychology of the Emotions and Their Relationship to Internal Regulatory Variables. In M. Lewis, J. M. Haviland-Jones & L. F. Barrett (Eds.), *Handbook of Emotions* (pp. 114-137). New York: The Guilford Press.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of Memory* (pp. 381–403). New York: Academic Press.

Tulving, E. (1983). *Elements of Episodic Memory.* Oxford: Oxford University Press.

Tulving, E. (1986). What kind of hypothesis is the distinction between episodic and semantic memory? *Journal of Experimental Psychology: Learning, Memory, & Cognition, 12*, 307-311.

Tulving, E. 2002. Episodic memory: From mind to brain. *Annual Review of Psychology* 53: 1–25.

Williams, H.L., Conway, M.A., & Cohen, G. (2008). Autobiographical memory. In G. Cohen & M. A. Conway (Eds.), *Memory in the real world* (3rd ed., pp. 21–90). Hove: Psychology Press.