

ARTICLE

Episodic Memory and Unrestricted Learning

Simon Alexander Burns Brown 

William H. Miller III Department of Philosophy, Johns Hopkins University, Baltimore, MD, USA
Email: simonabbrown@gmail.com

(Received 05 April 2022; revised 12 July 2022; accepted 19 January 2023; first published online 17 February 2023)

Abstract

Our thinking often uses rich memories of particular past events. Yet frequently we would do better to use other forms of memory. I show that existing accounts of the function of episodic memory cannot account for such cases, then develop an account which can. Roughly: rich representations of particular past events are required for Unrestricted Learning, learning which is not limited in how much of the world's complexity it can capture; and episodic memory's selection for Unrestricted Learning could explain its ubiquitous (and often inappropriate) use for other tasks. This proposal suggests many avenues for further empirical and computational research.

1. Introduction

Our thinking often revolves around rich memories of particular past events. Yet in many uses of such memories, it is unclear why. Consider some representative cases:

RECIPE: You want to make that cauliflower curry — the one you first made three years ago. Your problem: you cannot find the recipe. Your solution: trying to recreate it by recalling the particular occasion on which you made it — including where you were, who was there, and even conversations you had and how you felt.

GRENOBLE: You are considering visiting Grenoble. Part of the decision rests on what Grenoble is like. You have been to Grenoble multiple times, but you turn to memories of a particular day, including details unlikely to recur, or to matter — the shape of the clouds, the smell of the perfume of someone on the train, the shape you noticed of several houses.

Representing all this extraneous detail and tying it to particular events seems a spectacular waste of resources. Only the general patterns — the recipe, the gist of Grenoble — are important to the tasks. We have other forms of memory which

abstract away from specifics to only include such general patterns. *Prima facie*, such memories should be cheaper to store and easier to operate with. So, we frequently expend energy recalling a pointless amount of distracting detail. Why?

Solving this puzzle turns out to have far-reaching implications. There are many accounts of what episodic memory is for, and I will argue that most cannot answer the puzzle. Furthermore, my answer will teach important lessons about the contribution of memory to general, flexible intelligence. According to my answer, rich memories of particular events are useful for what I will call *Unrestricted Learning*. Roughly, this is the ability to continue improving one's model of one's environment without limit, in contrast with forms of learning that are inherently restricted in the amount of complexity they can capture.

§2 clarifies terms, and shows the underappreciated range of potential alternatives to episodic memory, enabling §3 to lay out the puzzle more carefully. §4 shows how various proposals for episodic memory's function fail to solve the puzzle and extracts a set of desiderata for adequate solutions. §§5-6 lay out my positive account, which fulfils these desiderata. §5 explicates *Unrestricted Learning* and shows how it meets some of the desiderata. §6 completes this task by showing how to explain the *ubiquity* of episodic memory, even in cases which are not obviously instances of *Unrestricted Learning*. §7 replies to objections, and §8 concludes.

2. Episodic memory and its alternatives

For the purposes of this paper, "episodic memory" will refer to any rich memory of a particular past event. Clarification of the components of this stipulative definition is in order.

By "richness," I mean a memory's including a large number of details, bound together into a unified representation. When recalling your graduation, you might merely recall the bare fact that it involved a long speech. But you might instead recall a great deal of information in one package — sensory information including how different locations, people, and items looked, sounded, etc., alongside (at least) emotions, associations, and contextual background. And this information might be unified, rather than a series of separate memories of individual features that happen to be of the same event. This is not the place for a full account of this unity. But one salient aspect will be that accessing any individual piece of information in the package makes accessing the rest of the package much more likely.¹

By "particular event" I mean a token event as opposed to an event-type.

It might be supposed that richness already guarantees particularity, given that the richer the representation, the more likely it is to pick out a combination of features unique to just one event. However, it is important to distinguish a *specific* — and hence rarely instantiated — event-type from a *particular event-token*. Tim's tenth birthday party might in fact be the lone instantiation of the kind *an event in which a couple looking precisely like Tim's parents in their forties host a tenth birthday party for their child, and lightning hits the birthday cake*. But, in principle, multiple events could instantiate this kind.

¹ Richness in this sense likely has deep connections to imagery, and both of these likely have deep connections to the sense of "reliving." Pinning down the precise relationship between the three is a task for another paper.

Representing specific event-types and representing event-tokens are rather different enterprises, governed by different norms. If another event is discovered on a distant planet which also involves a lightning strike and doppelgängers of Tim's family, then a representation of the specific event-type enumerated above should be applied to it; but "Tim's tenth birthday party," a representation of the particular event-token, should not. Likewise, representing the particular object Tim — which could be achieved in many ways, including using a definite description ("the boy who attracts lightning"), demonstrative ("that boy"), or name ("Tim") — is a rather different enterprise to representing a specific type (e.g. *boys who attract lightning*).

These distinctions help us get clear on other forms of memory which could be used instead of episodic memory for different tasks. It is common, following Tulving (1972), to introduce episodic memory by contrasting it with *procedural memory* — concerning how to perform activities, such as cycling — and with *semantic memory* — concerning bare facts, like remembering that Bamako is the capital of Mali. It is controversial exactly how to distinguish between different forms of memory, but particularity and richness will play an important role. If procedural memory involves representation at all, it does not represent particular events. Meanwhile, semantic memories — at least, those memories that concern just bare facts — are not rich. However, given that we have distinguished richness from particularity, we can recognize further forms of memory which are either not rich or not particular.

There are rich memories which are of (potentially very specific) event-types rather than tokens, such as remembering your normal morning routine, what lunch was generally like at your school, or a movie clip you have seen multiple times (for related discussions in philosophy, psychology, and neuroscience, see Addis, Moscovitch, Crawley, and McAndrews, 2004; Burge, 2011; J. Campbell, 1994; Franklin, Norman, Ranganath, Zacks, and Gershman, 2020; Ghosh and Gilboa, 2014; Lee, Aly, and Baldassano, 2021).

Additionally, there are non-rich memories about particular events. Examples include remembering facts about historical events where one wasn't even present, remembering one's date of birth, or remembering that one passed an algebra exam at sixteen without remembering the exam itself.

One might worry about how to classify some of these cases. For example, Rubin and Umanath (2015) and Andonovski (2020) have argued in effect that rich, non-particular memories should count as episodic memories, while many would define "episodic memory" by appeal to different features, such as a special phenomenology, or the use of the hippocampus. However, "episodic memory" can be treated *here* as a convenient label for a stipulatively defined class of cases, of interest because they give rise to a certain sort of puzzle. I now turn to that puzzle.

3. The puzzle

As suggested above, many of the tasks for which we seem to use episodic memory are puzzling because, *prima facie*, they would be better tackled with alternative forms of memory, such as rich non-particular memories or semantic memories. Closely related puzzles are expressed by many others, including Lengyel and Dayan (2007, 889), Hoerl and McCormack (2016, 241), and Schulz and Robins (2022, 15). To illustrate

the contours of the puzzle, I will elaborate on the reasons why some existing proposals fail to fully solve it.

The most direct solution is to deny that we do frequently richly remember particular events. One might agree that *rich* memories are frequent, but claim that most of our rich memories are really only *specific*. Andonovski (2020) draws attention to many examples of memories that are often classified as “episodic,” involve a sense of reliving, represent spatial setting and perspective, and involve the medial temporal lobe, but which turn out not to be about particular events. These would include the rich but non-particular memories discussed above such as remembering one’s school dinners; but Andonovski points out that even some examples Tulving gave when originally introducing the term to the literature, like remembering meeting “a retired sea captain who knew more jokes than any other person I have ever met,” might fall into this category. However, while it is true that many do somewhat overestimate the relative frequency of rich, particular memories in this way (as Andonovski details), that frequency is high *enough* to be puzzling. It is not clear how one could ascertain a precise quantitative estimate of episodic memories.² But RECIPE and GRENOBLE are not atypical. Most people immediately recognize such cases, and can readily generate more.

At this juncture, it might be questioned whether RECIPE, GRENOBLE, and similar intuitive cases really are particular, as opposed to misclassified highly specific memories. A full reply to this worry would require defending a substantive account of what makes it the case that a memory represents a particular event rather than a specific event-type. However, it is telling that we do often implicitly respect the particularity/specificity difference in our ordinary memory-based thinking in a way that suggests that many of these memories are particular. Having a rich memory of *the times I used to cook cauliflower curry with Yara and Zara* feels different subjectively to a rich memory of *that time I cooked cauliflower curry with Yara and Zara* — and we do have both sorts of memory. If we find we have mistaken one for the other (for example by rereading old diaries and discovering that we have amalgamated events which happened on different nights with Yara and Zara into one event), we are surprised and disconcerted, perhaps coming to distrust other features of the memory in question. We treat the memory as having *illegitimately* merged distinct events, unlike those memories which wear their status as summaries of specific event-kinds on their faces, applying different norms to the two cases. We treat it as making sense (if sometimes a hopeless endeavour) to search for a particular time at which the events we remember occurred.

Another initially appealing solution would have it that episodic memory is for informing us about things we have experienced. In GRENOBLE and RECIPE, we use episodic memory because it is one way of accessing information about curries and Grenoble.

At some level, this answer is correct. But it fails to tell us why we use episodic memory rather than other forms of memory. What these tasks call for is generic

² Andonovski (2020) cites Barsalou’s (1988) finding that 21% of subjects’ reports when asked to recall events from the previous summer are clearly about particular events; but it is debatable how closely these numbers track number of *memories* available or accessed, or how far we should expect this number to generalise to other tasks and situations.

information — a repeatable recipe or repeatable features of Grenoble. They do not call for rich, unique information about a particular past event. Such information is at best irrelevant, and arguably a costly distraction.

We can get clearer on these costs by understanding learning and cognition as involving algorithms for constructing and operating with statistical models. Such computational accounts of memory and learning are becoming increasingly sophisticated. But to begin, we can consider a very simple model an animal might learn through implementing a very simple algorithm. Suppose you want to collect nectar effectively. A good way to do this is to use a model that predicts where nectar is to be found. A good way to learn this model might be to learn average nectar levels at different locations, throwing away other information about those locations—such as the fact that you once heard a sparrow chirp while you were there—as irrelevant.

One might think remembering particular events would be useful for learning the relevant averages. You could visit a location t times and store each particular occasion i 's nectar level Y_i , summing these and dividing them by t to calculate the mean \bar{Y}_t :

$$\bar{Y}_t = \frac{\sum_{i=1}^t Y_i}{t} \quad (1)$$

However, an alternative method would be to update \bar{Y}_t each visit, throwing away the particular data as soon as it is incorporated into your running average (See e.g. Sutton & Barto 2018 for numerous examples like this). This could be done by setting $\bar{Y}_1 = Y_1$ then updating this initial estimate like so:

$$\bar{Y}_t = \frac{(t-1)\bar{Y}_{t-1} + Y_t}{t} \quad (2)$$

(Eq. 1) and (Eq. 2) give the same results, but (Eq. 2) does not require memory for particular visits. The estimate \bar{Y}_{t-1} already incorporates Y_{t-1} , Y_{t-2} and so on.

This sort of formalism does encapsulates why remembering particular events is unnecessary for many tasks where more generic forms of memory are available. But it also helps us get clear on the costs of storing such episodes individually and in rich detail. Such memory is not simply pointless, but *expensive* in several respects.

Firstly, particularity requires special machinery. The easiest kind of learning to implement in neural networks is simple Hebbian learning. Hebbian learning essentially achieves a more complex version of the above example: it incrementally adjusts connection weights to amalgamate information from all of one's experiences into one representation, from which those individual experiences can no longer be extracted. It is not *impossible* to represent particular experiences uniquely with neural networks: with the right kind of structure they can represent anything (Piantadosi 2021). However, this does require specialized structure and imposes attendant costs.

Richness also carries costs, relative to representing bare facts: actively representing many details presumably requires resources, and these seem wasted if these details are irrelevant.

Episodic memories likely are not just resource-heavy, but carry performance costs, thanks to being worse-suited to many tasks than alternative forms of memory. Predictions based on summary statistics incorporating information from many experiences will typically be better than generalizing from a handful of experiences that

happen to be individually recalled. Indeed, several common failings of human cognition from the classic heuristics and biases literature (Tversky and Kahneman 1974) arguably relate to our overreliance on particulars. For example, even when explicitly asked to estimate the probability of a random sample of men having an average height of greater than 6', we simply give the same answer, irrespective of whether the sample is of size 10, 100, or 1000, suggesting we are insensitive to the appropriate calculations and instead use a rule of thumb, perhaps based on a few particular men that happen to come to mind. Indeed, a tendency to base judgements on how easily particular cases come to mind shows up in other classic results along these lines that relate to events specifically: we have a tendency to assess the probability of events based on the ease with which similar events can be brought to mind, even when that ease is driven by salience. For example, estimates of the probability that a house will burn down will be higher for subjects who have seen a house burn down than for subjects who have merely read about it in the newspaper, even though the relevant statistical information they receive may be identical. This is plausible because they are more likely to access a memory representing that particular house fire and more likely to use general statistical information or other events if they merely read about the fire. Finally, compulsive or addictive behaviour can be caused by undue fixation on a single, unrepresentative event: addicts often seem to be driven by a rich memory of intense, positive experiences the first time they took a drug, which is not representative of their subsequent interactions with the drug (Bornstein and Pickard 2020).

Episodically remembering may impose further costs insofar as episodic memories' richness makes it inefficient to process them. Some inefficiency may arise simply from dealing with the many task-irrelevant details represented qua richness. These problems will be compounded if some of these irrelevant features attract attention, or trigger irrelevant emotional associations. And they will be compounded further if multiple particular memories are considered. Hyperthymesia, where individuals automatically recall abnormally many events, in abnormally great detail, can be *overwhelming* (Parker, Cahill, and McGaugh 2006).

We have refined the puzzle to: why do we use episodic memories in contexts where other forms of memory would seem cheaper and more effective to store, retrieve, and operate on? Before introducing further potential solutions, we can further refine it by asking how it relates to the much-discussed issue of the *function* of episodic memory (Allen and Fortin 2013; Boyer 2008; 2009; Boyle 2019; 2021; Buckner and Carroll 2007; De Brigard 2014; Klein 2014; Klein, Cosmides, Tooby, and Chance 2002; Mahr and Csibra 2018; Michaelian 2016; Rasmussen and Berntsen 2009; Schacter, Addis, and Buckner 2007; Schacter, Guerin, and Jacques 2011; Schulz and Robins 2022; Suddendorf and Corballis 1997; 2007; Templer and Hampton 2013).

Schwartz (2020) argues that this literature is often ambiguous between asking about the *causal role* episodic memory plays in producing specific phenomena, and the selection pressures that shaped the *evolution* of episodic memory. Our puzzle is closer to the evolutionary question: it looks for *advantages* to the way we do things to explain why we do them this way rather than another way, and will not be answered merely by articulating the role episodic memory in fact plays. However, the evolutionary question is often framed in terms of why a system *capable* of episodically remembering *exists* in humans. The question here, by contrast, is about our widespread *use* of episodic remembering. This is shaped by learning and intentional

control as well as natural selection. An answer to our puzzle could appeal to any combination of the past survival value of inherited traits, cultural factors shaping learning, and occasion-specific (albeit possibly only dimly appreciated) reasons individuals have for episodically remembering. Any of the evolutionary accounts of episodic memory in the literature *could* answer our puzzle, and we will consider them in this light, but we will see that they are inadequate. This is partly because they were intended as answers to a slightly different question.

Most of these accounts fail to meet at least one of the following desiderata: reconciling episodic memory's relative expensiveness thanks to its being both (1) rich and (2) about particular events, with (3) its ubiquitous use, even for memories of (4) long past events. Rather than discussing every account in the literature in detail, I will choose a few which illustrate these desiderata.

4. Desiderata for a solution

One prominent view — simulationism — holds that episodic memory should be seen in terms of a broader system for simulating future events and counterfactual scenarios.³ There are many versions of this idea. For our purposes it could be developed in two different ways: claiming that episodic memory per se was not selected for, but was rather a *side effect* of selection for a system that simulates future/counterfactual events; and claiming that episodic memory was selected for some role it plays in *supporting* the simulation of future/counterfactual events.⁴ Each faces a distinct problem if offered as an account of why we so frequently episodically remember.

The side effect view faces a problem in accounting for *ubiquity*. Even if the main reason we have the *capacity* for remembering particular events in rich detail lies in a more general-purpose simulation system, this does not explain why we exercise this capacity so frequently, especially if many of these exercises are costly and inappropriate relative to other processes we could be using. Schulz and Robins (2022) emphasize that suboptimal traits can persist for some time if selection pressures against them are weak enough. However, this point is not enough to solve the problem. Firstly, the reasons outlined above justify suspecting that the selection pressures against our use of episodic memory are relatively strong. Secondly, even if Schulz and Robins can explain the *persistence* of our apparent widespread misuse of episodic memory, they do not have an explanation of its *emergence*. The *capacity* for episodic memory may be a by-product of developing simulated future planning; but this does not explain why we in fact came to exercise this capacity so often, given that many of the costs identified above are incurred by each *use* of episodic memory.

³ Prominent expressions of simulationism include De Brigard (2014); Hassabis, Kumaran, and Maguire (2007); Klein (2015); Michaelian (2016); Schacter, Addis, and Buckner (2007); Schacter, Guerin, and Jacques (2011); Suddendorf and Corballis (2007); and Tulving (2005).

⁴ It is not always obvious which of these views different simulationists are advocating (and exegesis is complicated by different uses of “episodic memory”). A relatively clear example of advocacy of the side-effect view can be found in Schulz and Robins (2022) (although they would resist many claims about what memory is which typically fall under the label “simulationism”), while a relatively clear example of advocacy of the simulation-supporting view is Schacter et al. (2007).

The simulation-supporting view does not straightforwardly face this problem; it can say episodic memory is ubiquitous because simulation is ubiquitously useful, and requires episodic memories to provide “ingredients.” However, it is not clear that simulation really does require any such thing. Indeed, we can look at any uses of episodic memory to provide such ingredients as *instances* of our puzzle. It would seem to make more sense to use amalgamated event-kinds and/or individual components instead, extracting and incrementally updating such raw materials for future simulation *at the time of encoding* without storing particular events.

Side effect views are not the only views to struggle with explaining episodic memory’s ubiquity. Other accounts explain the use of episodic memory in very specific circumstances, but have little to say about most of the occasions on which it is used. For example, Hoerl and McCormack (2016) make the plausible suggestion that some kinds of regret require episodic memory. Plausible, but irrelevant to most uses of episodic memory, including GRENOBLE and RECIPE. Mahr and Csibra (2018), meanwhile, emphasize the use of rich episodic memory for establishing special epistemic authority with respect to past events in certain social contexts. But many of our uses of episodic memory (including RECIPE, plotting your route through a city by thinking about particular occasions you were at certain locations, and reflecting on a particular occasion on which you beat a difficult video game to try to generate more general lessons about how to beat it on other occasions) appear unrelated to such uses. Indeed, Mahr (2019) explicitly claims that representing the past is only useful in social contexts — but in that case, why do we so often use such representations for non-social tasks?

Other accounts do a better job of explaining the ubiquity of the states *they* dub “episodic memory,” but do not explain the ubiquity of states which are about particular events. For example, a number of recent computational models have a role for something called “episodic memory” which on close inspection requires only richness, not particularity. Lengyel and Dayan (2007) in effect suggest that in extremely complicated environments, rather than trying to extract abstract features of scenarios in which certain actions are rewarding, it can make sense instead to store detailed records of successful episodes, and to attempt to reproduce identical sequences of actions in similar scenarios. However, they do not carefully distinguish specificity, being learnable on the basis of a single training example, and representing a particular event; and arguably, their solution to complexity only requires representing specific episode-types. Consider what such a system should do *after* repeating an action based on a single successful past experience and again being rewarded: it ought to strengthen this one memory, not create a memory of a second distinct event. Similar points could arguably be made about, for example, Franklin, Norman, Ranganath, Zacks, and Gershman’s (2020) more complex model, although showing this in detail is beyond the scope of this paper.

Note that these sorts of models *can* explain why we use rich but non-particular memories. For example, they could help explain a different version of RECIPE in which you do not remember a particular occasion on which you made the cauliflower dish, but rather an event-kind, such as the many times you made the dish with Yara and Zara. It is worth emphasizing here that there is no reason to expect one account to cover all the puzzling cases in the vicinity: on the contrary, there are probably many overlapping forces that jointly contribute to episodic memory’s and related

states' uses across different cases. It is just that forces which do not select for particularity only explain some cases, and the account developed later in this paper is needed for others.

Another solution which is instructive to consider is suggested by RECIPE. Episodic memories are often associated with more generic information, and a good way of accessing generic information which is hard to recall can be to recall the associated event. Perhaps episodic memories systematically organize other information in a useful way. Boyle (2021) develops a version of this thought, articulating features of episodic memories which aid recall of non-episodic information. For example, integrating information into a spatiotemporal structure and associating information with irrelevant but distinctive details can aid retrieval. However, as Boyle herself emphasizes, these features also show up when merely *imagined* contexts are associated with pieces of information, as in the well-known method of loci memorization technique, where individuals boost their memory for arbitrary facts by associating them with locations in an imagined space such as a "mind palace." This suggests that any rich representation would do the job. Indeed, fictional events might be *better*, as they could be tailor-made for this purpose, including more unique and distinctive features. So, we are still left wondering why we use episodic memories.

Boyle has an answer available: given the way she thinks generic memories are formed, there will always be an episodic memory ready to hand to associate with any given piece of information. She appeals to McClelland, McNaughton, and O'Reilly's (1995) suggestion that semantic memories are typically formed on the basis of episodic memories being repeatedly replayed, in order to avoid *catastrophic forgetting*. McClelland et al. understand learning as adjusting the weights in a distributed connectionist network. Changing such a network too rapidly in response to new experiences can lead its existing knowledge structure to break down, as changes in any one part of the network require adjustments elsewhere. Such problems can be avoided through adjusting any given weight only a tiny amount at once, and then allowing these changes to ramify through the rest of the network. But if the network only adjusts a tiny amount in response to any new experience, learning requires many experiences. One solution is to repeatedly replay experiences and incrementally update in response to each replay, thereby accruing many incremental updates for each actual experience.

Unfortunately, the idea that forming other memories depends on first storing episodic memories does not solve our problem on its own. One initial worry is that it is debatable whether processes like those posited by McClelland, McNaughton, and O'Reilly (1995) in fact require representations of *particular* events. Another worry brings out a new aspect of our puzzle: We often rely on *old* episodic memories, from months or years ago. In such cases, including GRENOBLE and RECIPE, we *do* already have relevant general representations. We use episodic memories *instead* of available general memories. It cannot even be that episodic memories are still used because they happen to be available for structuring general memories when the latter are formed. This is because McClelland et al.'s two systems do not neatly correspond to episodic and semantic memory as understood here, so much as to (i) rapid formation of *short-lived* episodic memories, and (ii) all longer-lived memories. Keeping rich memories of particular events long term requires specially copying them from hippocampal into neocortical storage. It is unclear why committing episodic memories to

long-term storage would make for a better way of organizing other memories than creating other systems of organization.

In addition to explaining the ubiquitous use of memories for long-past particular events, we also need to explain richness. One proposal that, without supplementation, fails this test is the suggestion that episodic memory is useful in unanticipated tasks. Versions of this view appear in several authors (including S. Campbell 2006; Klein, Cosmides, Tooby, and Chance 2002; Mar and Spreng 2018; Templer and Hampton 2013). The view also helps motivate one paradigm for episodic memory in animals — testing for “incidental encoding,” i.e. accessibility of information which does not seem important at the time of the event (e.g. Fugazza, Pogány, and Miklósi 2016; Fujita, Morisaki, Takaoka, Maeda, and Hori 2012; Zentall, Clement, Bhatt, and Allen 2001; Zentall, Singer, and Stagner 2008; Zhou, Hohmann, and Crystal 2012). One particularly useful articulation of such ideas can be found in Boyle’s (2019) paper. She suggests that episodic memory allows organisms to learn from past events retrospectively. One of her examples is reassessing one’s belief that bees do not sting, upon being unexpectedly stung: retroactively comparing cases of being stung and not being stung can help one form a subtler belief about the *conditions under which* bees sting.

Returning to our model of learning about nectar levels helps to appreciate how particularity is useful for unexpected tasks. Notice that (Eq. 2) is only useful if you have been updating an estimate $\overline{Y_{t-1}}$. If you suddenly need to estimate \bar{Y} for the *first* time, (Eq. 2) will be useless; and if you have been throwing away the data, you will be stuck. But if you *have* been storing the particular nectar levels from each occasion, you could use (Eq. 1).

However, it is not obvious that *rich* memories unifying many disparate details of such events would be needed for this. Why not simply store mere lists of univariate data-points — the levels of nectar on different particular occasions — without connecting these to or even storing other details? Or take Boyle’s bee case: We do use episodic memory in such cases. But do we need to? Boyle’s reason for thinking so seems to rely on assuming that the only possible alternative to episodic memory would be a simple semantic memory “that bees are sometimes harmful after all” (Boyle 2019, 246). Yet *prima facie*, one could instead get by with non-rich semantic memories about particular bee-related-events, specifying unusual and plausibly sting-relevant features of the unusual sting-involving events.

To sum up: several accounts explain *some* uses of episodic memory but far from all, and others explain *some* features of episodic memory but other forms of memory also have those features. We need an account which can reconcile episodic memory’s relative expensiveness thanks to its being both (1) rich and (2) about particular events, with (3) its ubiquitous use, even for memories of (4) long past events. My account will meet these desiderata, partly by combining insights from some of the accounts already discussed.

5. The usefulness of particularity and richness

This section will complicate the example of nectar levels to show that rich memories of particular events, including long-past events, are required for what I call “Unrestricted Learning.” This suggestion combines Boyle’s (2019) emphasis on unanticipated epistemic needs with the idea that certain kinds of learning *systematically*

introduce unanticipated epistemic needs for *rich* representations. Related ideas have been developed before (Gershman and Daw 2017; Nagy and Orbán 2016). However, I will do so while consistently distinguishing particularity and richness, and using a formal framework which renders the point intuitive.⁵ §6 will extend the proposal to explain ubiquity, and will also connect the formalism of this section to more everyday examples of thinking which do not involve consciously and explicitly engaging in statistical learning.

Taking the mean of previously observed values of some variable — like nectar levels at a location — is often a good way of predicting future values of that variable. However, any variable could be interacting in a multitude of ways with any other variable, and simply calculating the means of each variable individually will miss these interactions. One way of capturing interactions is regression analysis. A regression model is an equation giving a predicted value of a target variable Y as a function of n variables X_1, \dots, X_n and at least $n+1$ parameters a_0, \dots, a_n . The simplest case will look like this:

$$Y = a_0 + a_1X_1 \quad (3)$$

Here, Y depends on just one variable, X_1 , and two parameters, a_0 and a_1 . This might be a model that predicts nectar levels solely on the basis of distance from the center of the garden.

We can fit a regression equation to data. That is, observed combinations of values of Y and X_1, \dots, X_n on particular occasions can be used to estimate the values of the parameters $a_0 \dots a_n$, using methods that are similar in spirit to (Eq. 1) or (Eq. 2). Once we have these, we can simply plug in observations of X_1, \dots, X_n to predict Y — e.g. using distance of a previously unvisited location to predict its nectar levels.

We might want to complicate our model in a variety of ways. We might want to add extra variables. Perhaps both location and light levels are relevant to nectar levels:

$$Y = a_0 + a_1X_1 + a_2X_2 \quad (4)$$

We might want to capture non-linear effects. Perhaps the influence of location is not constant, but instead small changes have tiny effects, whilst larger changes have disproportionately larger effects:

$$Y = a_0 + a_1X_1 + a_2X_2 + a_3X_1^2 \quad (5)$$

We might want to include interaction effects capturing how one variable's effect on Y is partly mediated by the value of another. Perhaps the influence of location on nectar level is larger at higher light levels:

$$Y = a_0 + a_1X_1 + a_2X_2 + a_3X_1^2 + a_4X_1X_2 \quad (6)$$

Complicating the model can increase predictive power, assuming that the world is more complex than our current model allows—but at a cost. It requires extra computation; and it requires stretching the limited data we have to estimate more

⁵ This is not to say that my argument depends on the details of the formal framework used here: for example, where I talk in terms of traditional linear regression analysis, the same point would hold if we used a Fourier basis instead, and Nagy and Orbán (2016) make a very similar point using a Mixture of Gaussians framework.

parameters. Adding extra parameters too freely introduces the potential for overfitting—introducing so many parameters that the equation spuriously “finds” patterns in mere noise, patterns which will not generalize beyond the particular observations used to fit the model.

The net benefits of additional model complexity increase with the amount of data available. As we gain the data to fit it properly, a more complex model may capture more of the world’s real complexity, without leading us astray. This means that the form of our optimal model will change as we gather more data. There are a few ways of dealing with this.

We could ignore this change and simply fit one model, using extra data to incrementally improve parameter estimates but not changing the structure of the model itself. This would be relatively simple to implement. But we would be guaranteed to remain eternally blind to any structure in the data that we did not hypothesize from the outset.

We could fit multiple models in parallel—one which takes into account light levels, another with non-linearities etc. This would require a great deal of computation at every stage and so would be extremely costly, especially with a large number of such models. We could easily end up updating values for thousands of parameters in thousands of models. And we would still be limited to the finite set of models we actually chose to fit.

Ideally, we could pursue a different strategy: starting with just one or two simple models and gradually increasing their complexity, flexibly adding and deleting variables and testing the resulting models for improvements in performance over their predecessors, abandoning them if need be. This kind of learning would in principle be capable of capturing indefinite amounts of complexity in the environment. Rather than being limited to fitting pre-specified models, it could in principle learn about any combinations of variables. It would be *unrestricted* learning.

Something like episodic memory would be crucial to Unrestricted Learning. Such learning systematically requires estimating parameters that have not already been estimated. And as we have already seen, this requires access to particular events — original data points which have not been amalgamated into existing models. New data can always be collected to fit new models, of course, but practically this approach would grind to a halt. Each time you tried out a new, fancier model, you would need to collect enough new data to test it properly. So even trying out a new model would require collecting as much data as you have already collected for the current model *again*, plus some more (as this is a more complex model). By contrast, using existing stored data points would mean that you could at least try out potential fancier models, to select some for testing against future data.

Unlike in the case of estimating a mean, remembering rich details about events will also be important to Unrestricted Learning. To fit or even get preliminary evidence for models including the interaction of multiple different variables, we need the values these variables have *simultaneously* taken on during particular events. And, for such learning to be unrestricted, it must be capable of learning any combination of variables, such that we cannot know, at the time of storage, which variables will turn out to be important. Remembering sequences of values of individual variables will not be enough: we need access to the values of multiple variables *at once*.

6. Explaining ubiquity

Providing data for Unrestricted Learning is a potential use for episodic memory over other forms of memory. However, the challenge is not to show a *potential* use, but to explain ubiquitous usage. Talking about Unrestricted Learning may seem like a worse solution to this challenge than some of the proposals rejected above. We rarely consciously think about statistics, and in cases like RECIPE, we are not even trying to form a new generalization. I will argue, however, that (a) in an important sense, there *are* many ordinary uses of episodic memory which should be thought of as contributing to Unrestricted Learning; and (b) we can explain *other* ordinary uses of episodic memory as the natural consequence of a system selected to frequently engage in the kind of activity described in (a).

We do not consciously build complex statistical models of our environment and test them against our episodic memories. However, cognitive psychology often posits computations understood as developing and fitting statistical models which are not consciously accessible. While the *exact* role and status of such explanations is hotly contested, such posits often capture (1) introspectively inaccessible subpersonal processing; and/or (2) a computational level explanation, specifying the computations a process is in some sense implicitly “aiming at.”

It may well be that there is unconscious processing which instantiates algorithms for Unrestricted Learning and which draws on conscious rich remembering of particular events, much as there is likely to be unconscious statistical learning from conscious perception. But defending this claim would require testing detailed computational models with behavioural and neural data, a project beyond the scope of this paper.

What can be done here is to make (2) plausible, by pointing to commonplace cases where our introspectively available cognition is drawing on episodic memories to form new generalizations, and showing how these cases fit the general form of Unrestricted Learning. Such ordinary cases appear abundant once one seeks them.

In nearly any domain where we try to make sense of a complex system we have personally experienced, we find it natural to proceed by trying to generalize from individual experiences with that system, then testing those generalizations against other memories and new experiences. In trying to figure out details of someone’s character (including one’s own), it is common to begin by fixating on particular experiences involving that person and trying to generate potential generalizations from these, before considering these generalizations’ performance across a broader range of cases. Or in getting to know a complex piece of equipment — a new car, musical instrument, or computer — it is natural to go back to the details of how it behaved in some particular situation to try to form new, ever more nuanced, hypotheses about how it behaves in different conditions. There are other ways to learn about a complex piece of equipment or person’s character: we can try to theorize using our existing understanding of the domain, or learn slowly from large amounts of experience. But thinking through a particular, actual case is likely to be especially quick and efficient, as it is more likely to generate genuinely promising hypotheses and allow for preliminary testing of that hypothesis, where these other approaches require either more speculation with less connection to the specifics of the case, or many more experiences.

Something looking very much like Unrestricted Learning also occurs when we are reminiscing or daydreaming about particular past events without any conscious agenda. We can suddenly make a connection to another issue and gain a potential new insight. Perhaps you are daydreaming about your wedding day a year ago and suddenly make a connection to the question of why uncle Bob is reluctant to go to your Christmas gathering — you remember aspects of his behaviour which did not make sense at the time but in hindsight could be seen as suggesting he secretly dislikes aunt Alice. These sorts of cases are particularly suggestive because they are cases where predicting in advance the form of future models, hence exactly which features will become relevant to remember, is particularly difficult. Recalling this particular event may not be the *only* route to understanding uncle Bob's behaviour: perhaps it would also have been possible to guess at based on what you know about his character in general. But the availability of this alternate route will depend on which features of your past experience your current general knowledge has abstracted away from; and in any case, having awareness of particular features of his past behaviour which can be explained by your new hypothesis but which were puzzling on your old understanding of the world can provide extra credence to your new hypothesis.

GRENOBLE can also be understood through the lens of Unrestricted Learning. Having visited Grenoble several times, you will have formed some generalizations. These might be good generalizations. But they will have *some* limits. It is only possible to have generalized about a limited number of aspects of your experiences, in limited ways. It may therefore be worth recalling the original experiences to glean more potential generalizations and to add subtlety to existing generalizations (such as recognizing that certain generalizations only hold for Grenoble *in the summer*). It will be worth having seemingly useless details available about your particular Grenoble-experiences for such revisions, as it will be impossible to predict in advance which details will turn out to be useful.

But why have these details *at the time that you need to make a decision about visiting*, rather than only when your task is updating your model? The answer might lie in focusing on ways of enriching your model which are most relevant to your current decision. There will be *many* potential revisions to the model. When considering the experience afresh *in light of your current purposes*, some of these potential revisions, concerning previously unconsidered dimensions of the experience, might leap out as worth examining. For example, that shape you noticed to many of the houses may, on reflection in light of considering a winter trip, suggest hypotheses about how well Grenoble is designed for snowy weather.

Considering GRENOBLE through this lens not only helps us understand why we often recall irrelevant details during decision-making, but also brings out another noteworthy feature of such cases. We need not, and often do not, simply generalize from a particular recollected experience and directly decide on that basis. We often at least implicitly test any new generalizations against other sources of evidence (including other memories), and perhaps are disposed to remember counterexamples. This is not to say that we are optimal in our use of episodic memories after all: we often are too driven by a meager diet of cases. But not to the extent we might think when first considering the puzzle.

Focusing on Unrestricted Learning also improves on Boyle's (2019) account, although they are superficially similar in this sort of case: while Boyle also pointed to the use of such memories for unexpected revisions to our beliefs, she did not clearly distinguish particularity and richness, and hence underestimated the potential for alternatives to episodic memory. We can now see that such alternatives would not do the job required for Unrestricted Learning. Rich memories might suggest hypotheses, but need to be particular for even preliminary testing, or they will miss out on data that has been smoothed to conform to existing models; and particular memories can only be used for testing *unpredictable* hypotheses about *interactions* between variables if they are rich.

One reason generalization from cases can be too hasty is that not all cases are alike. And it would be too hasty for us to think that all cases can be explained similarly to GRENABLE. RECIPE does not involve a new generalization at all. Rather, episodic memory is used to help recall a specific generalization (the recipe) which has already been formed. However, we *can* shed light on such cases by combining the point about the systematic usefulness of episodic memory for Unrestricted Learning with Boyle's (2021) points about memory organization.

Boyle's insight was that episodic memories can be used to structure access to other forms of information. The main problem with her account was that it lacked a satisfying reason why episodic memory (especially for long-past events) should be used for this rather than other rich representations. We can now solve this problem. It is useful to have a system which is *poised* to use episodic memory for Unrestricted Learning. This means having plentiful *easily available* episodic memories, including some of long-past events. Indeed, it would make sense to design the system so that whenever a task arises which is not solvable by an immediately accessible stored solution, relevant episodic memories are brought to mind *in case Unrestricted Learning is called for* — as it will not be possible to reliably predict exactly when this will be. The system's being so poised changes the relative costs of using episodic memory versus constructing a new fictional event. So memory organization will often make use of the former.

This brings out a broader point: given a system poised for using episodic memories in Unrestricted Learning, such memories will often be readily available, and we are well-practised in using them. And this can make episodic memory convenient to use even for tasks where other forms of memory would otherwise be more appropriate. Therefore, the potential solutions to the puzzle which were dismissed above because other forms of memory are better suited to the tasks in question (e.g. the computational models like Lengyel and Dayan (2007) which only seemed to require richness) might still capture why we use episodic memory for certain tasks, *given that episodic memory is already ubiquitous*. Accounts which do not explain ubiquity on their own, meanwhile, such as Hoerl and McCormack's (2016) regret account, may explain further uses of episodic memory.

Our being poised for Unrestricted Learning may explain aspects of cognitive lives well beyond episodic memory. For instance, we are often more effective when thinking about abstract issues in terms of concrete cases rather than directly using generalizations. The best way to introduce abstract theories to students is often in terms of particular cases, and we often think about complex social patterns in terms of particular historical precedents. This may be because we are well-practiced in

thinking about particular cases. And this, in turn, may be at least in part driven by our being poised to use episodic memory for Unrestricted Learning.

7. Objections and replies

One might worry about various aspects of this account. However, far from undermining it, the most compelling objections to the proposal will turn out to provide reasons for further developing it.

Unrestricted Learning would benefit from remembering all the details of all events ever encountered. Fitting complicated regressions would ideally be done with as much data as possible. And yet we do not seem to episodically remember every detail of every event in our lives. We do throw away a lot of data. Is this a problem for a view which says episodic memory is important because it is used for such processes?

No. I emphasized above that episodic memory is expensive. There are trade-offs between remembering as much as possible for Unrestricted Learning, and not remembering too much given other goals. Further research would explore these trade-offs in detail, and how they might shape what we remember — following an active field of empirical (e.g. Chen, Cook, and Wagner 2015; Rouhani, Norman, and Niv 2018) and modelling (e.g. Benna and Fusi 2021; Lu, Hasson, and Norman 2022; Mattar and Daw 2018) work on related questions.

One might also worry that human episodic memory is too unreliable to be used as data for any sort of useful model-building. Psychologists have found numerous ways (reviewed in Loftus 2005; Roediger 1996) of inducing subjects to make memory errors, and even confabulating entire events. If subjects routinely make such errors, it is hard to see how episodic memory could be useful for even preliminary testing of complicated models.

It is questionable, however, just how routine and serious such mistakes are in normal contexts. Although there are some well-known cases of subjects misremembering details in high-stakes scenarios (Neisser 1981), most of the best-known experimental effects only occur in unnatural conditions, given certain kinds of prompting. It may be best to think of memory errors on the model of perceptual illusions (Roediger 1996); and in neither case does the fact that we can reliably induce mistakes imply that the process is particularly unreliable, let alone too unreliable to be of epistemic use (Michaelian 2016). Again, there is a fruitful question for computational modelling here, namely determining just how much reliability is required for different functions. Computational modelling has already revealed a relevant result: Lu, Hasson, and Norman (2022) show (Appendix 2) that an artificial network can learn to rely on memories less in situations where they are likely to mislead, especially if the costs of error are high.

A more sophisticated version of this objection would point out that the details of experimentally-induced memory confabulation imply reconstruction of a particular kind—reconstruction influenced by one's current model of the environment. Does this pose a problem for using episodic memory to expand on those very models?

Not an insurmountable one. While the details of how reconstruction works in memory are complex and debated, we can be confident of this much: in typical cases, the reconstructed episodic memory at the moment of recall will be based on existing models *in addition to event-specific information*. We do not simply have a model

confirming that its predictions about a scenario are borne out by its own predictions about that scenario. For example, while you might have a general belief that Uncle Clive is grumpy and snobbish, you might also remember a particular occasion when he donned a Hawaiian shirt and let his hair down. Presumably, in remembering this event, your brain does some reconstruction based on *other* generalizations about Uncle Clive, such as a specification of his precise facial features and his typical turns of phrase, but for the unique aspects of the situation, it draws on a memory trace or source of information specifying that he really did those things that one time, overriding the general belief.

A different worry is that we do not really need memories for particular *events* for Unrestricted Learning: why would memory for particular people, places, objects, etc. not provide raw data in our sense — connections to the world beyond those old theories which we are trying to overcome? Recall that in general, generic memories cannot do the job as well as particular memories because they lose the most relevant data in their construction, by smoothing out variance between cases left unexplained by the previous model. This will mean that memories for particular people will be more likely to be useful for Unrestricted Learning — by containing such to-be-explained variance — than memories generalizing across all people. My current model of people in general might be one that paints humans in general as loving flowers, but my uncle David might provide an exception which may prove illuminating. Yet notice also that particular events are in a relevant sense *more* particular than particular people. Memories for particular people will typically be extracted from multiple particular events involving that person, while memories for particular events involving multiple people will not usually average over these different people. Memory for particular events will hence typically be even more granular, even less smoothed out: I might remember not only that Uncle David does hate flowers, but that on his 50th birthday he enjoyed a certain species of orchid. And this might turn out to be key to understanding the exception to our general rule.

We might think that rather than developing one extremely complicated model of the world, the mind uses a collection of overlapping, simpler models (as suggested by Aronowitz 2019; Lu, Hasson, and Norman 2022). However, it is not clear how distinct this approach really is, at least from the zoomed-out perspective at which episodic memory is useful for Unrestricted Learning. If we have multiple models, there will be a pattern to our use of these different models to different extents on different occasions, and we can think of this pattern as (at least implicitly) instantiating a function determining when each model is used. We can treat Unrestricted Learning as determining this function: making decisions about whether to add or modify models to the overall repertoire, and how to use different models. Such decisions will be subject to the same constraints as in regular Unrestricted Learning, requiring rich representation of particular past events for parallel reasons.

8. Conclusion

Let us return to the desiderata from §3. We need an account which can reconcile episodic memory's relative expensiveness thanks to its being both (1) rich *and* (2) about particular events, with (3) its ubiquitous use, even for memories of (4) long past events. Unrestricted Learning points to such an account. (1) Episodic memory

needs to be rich so that it can allow the generation and at least partial testing of new hypotheses about the *interaction* of variables whose interaction has not been considered in general terms before. (2) It needs to be about particular events because it needs to act like raw data: amalgamating information from different events embeds existing generalizations and undermines possibilities for new generalizations. (3) It needs to be ubiquitous if the system is to be constantly poised for the possibility of learning new generalizations, and if it is so ubiquitous and available, it will become the easiest (if often suboptimal) option for many other tasks. (4) It needs to sometimes be about long-past events, because we can never assume that we have absorbed *all* the lessons from some long-past event and that the rest is noise. In light of Unrestricted Learning, it is no longer puzzling that episodic memory is used the way it is. Instead, episodic memory is revealed as a crucial part of radically flexible cognition and general intelligence. While I do not claim to have conclusively established that episodic memory is shaped by the factors I suggest, there are many opportunities for developing and testing these ideas using tools from the burgeoning field of computational modelling of memory and the use of such models for interpreting and guiding empirical experiments.

Acknowledgments. In addition to the anonymous reviewers, I would like to thank Matthew Heeney, John Morrison, Andrew Richmond, Christopher Peacocke, Kate Pendoley, and Ian Phillips for helpful comments on earlier versions of this material, as well as audiences at the Eastern APA, Philadelphia 2020; Issues in Philosophy of Memory 2, Grenoble 2019; XX Taller d'Investigació en Filosofia, Valencia 2019; Southern Society of Philosophy and Psychology, Cincinnati 2019; and PoPRocks Workshop in New York 2019.

References

- Addis, Donna Rose, Morris Moscovitch, Adrian P. Crawley, and Mary Pat McAndrews. 2004. "Recollective Qualities Modulate Hippocampal Activation During Autobiographical Memory Retrieval." *Hippocampus* 14 (6): 752–62. <https://doi.org/10.1002/hipo.10215>.
- Allen, Timothy A., and Norbert J. Fortin. 2013. "The Evolution of Episodic Memory." *Proceedings of the National Academy of Sciences* 110 (supplement_2): 10379–86. <https://doi.org/10.1073/pnas.1301199110>.
- Andonovski, Nikola. 2020. "SINGULARISM about Episodic Memory." *Review of Philosophy and Psychology* 11 (2): 335–65. <https://doi.org/10.1007/s13164-020-00464-y>.
- Aronowitz, Sara. 2019. "Memory Is a Modeling System." *Mind and Language* 34 (4): 483–502. <https://doi.org/10.1111/mila.12220>.
- Barsalou, Lawrence W. 1988. "The Content and Organization of Autobiographical Memories." In *Remembering Reconsidered: Ecological and Traditional Approaches to the Study of Memory*, 193–243. Emory Symposia in Cognition, 2. New York, NY, US: Cambridge University Press. <https://doi.org/10.1017/CBO9780511664014.009>.
- Benna, Marcus K., and Stefano Fusi. 2021. "Place Cells May Simply Be Memory Cells: Memory Compression Leads to Spatial Tuning and History Dependence." *Proceedings of the National Academy of Sciences* 118 (51): e2018422118. <https://doi.org/10.1073/pnas.2018422118>.
- Bornstein, Aaron M., and Hanna Pickard. 2020. "'Chasing the First High': Memory Sampling in Drug Choice." *Neuropsychopharmacology* 45 (6): 907–15. <https://doi.org/10.1038/s41386-019-0594-2>.
- Boyer, Pascal. 2008. "Evolutionary Economics of Mental Time Travel?" *Trends in Cognitive Sciences* 12 (6): 219–24. <https://doi.org/10.1016/j.tics.2008.03.003>.
- Boyer, Pascal. 2009. "What Are Memories For? Functions of Recall in Cognition and Culture." In *Memory in Mind and Culture*, edited by James V. Wertsch and Pascal Boyer, 3–28. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511626999.002>.
- Boyle, Alexandria. 2019. "Learning From the Past: Epistemic Generativity and the Function of Episodic Memory." *Journal of Consciousness Studies* 26 (5–6): 242–51.

- Boyle, Alexandria. 2021. "The Mnemonic Functions of Episodic Memory." *Philosophical Psychology* 35 (3): 327–49. <https://doi.org/10.1080/09515089.2021.1980520>.
- Buckner, Randy L., and Daniel C. Carroll. 2007. "Self-Projection and the Brain." *Trends in Cognitive Sciences* 11 (2): 49–57. <https://doi.org/10.1016/j.tics.2006.11.004>.
- Burge, Tyler. 2011. "Lecture I: Some Origins of Self." *The Journal of Philosophy* 108 (6/7): 287–315. <https://doi.org/10.5840/jphil20111086/715>.
- Campbell, John. 1994. *Past, Space, and Self*. Cambridge, MA: MIT Press.
- Campbell, Sue. 2006. "Our Faithfulness to the Past: Reconstructing Memory Value." *Philosophical Psychology* 19 (3): 361–80. <https://doi.org/10.1080/0951508600690573>.
- Chen, Janice, Paul A. Cook, and Anthony D. Wagner. 2015. "Prediction Strength Modulates Responses in Human Area CA1 to Sequence Violations." *Journal of Neurophysiology* 114 (2): 1227–38. <https://doi.org/10.1152/jn.00149.2015>.
- De Brigard, Felipe. 2014. "Is Memory for Remembering? Recollection as a Form of Episodic Hypothetical Thinking." *Synthese* 191 (2): 155–85. <https://doi.org/10.1007/s11229-013-0247-7>.
- Franklin, Nicholas T., Kenneth A. Norman, Charan Ranganath, Jeffrey M. Zacks, and Samuel J. Gershman. 2020. "Structured Event Memory: A Neuro-Symbolic Model of Event Cognition." *Psychological Review* 127 (3): 327–61. <https://doi.org/10.1037/rev0000177>.
- Fugazza, Claudia, Ákos Pogány, and Ádám Miklósi. 2016. "Recall of Others' Actions after Incidental Encoding Reveals Episodic-like Memory in Dogs." *Current Biology* 26 (23): 3209–13. <https://doi.org/10.1016/j.cub.2016.09.057>.
- Fujita, Kazuo, Ayako Morisaki, Akiko Takaoka, Tomomi Maeda, and Yusuke Hori. 2012. "Incidental Memory in Dogs (*Canis Familiaris*): Adaptive Behavioral Solution at an Unexpected Memory Test." *Animal Cognition* 15 (6): 1055–63. <https://doi.org/10.1007/s10071-012-0529-3>.
- Gershman, Samuel J., and Nathaniel D. Daw. 2017. "Reinforcement Learning and Episodic Memory in Humans and Animals: An Integrative Framework." *Annual Review of Psychology* 68 (January): 101–28. <https://doi.org/10.1146/annurev-psych-122414-033625>.
- Ghosh, Vanessa E., and Asaf Gilboa. 2014. "What Is a Memory Schema? A Historical Perspective on Current Neuroscience Literature." *Neuropsychologia* 53 (January): 104–14. <https://doi.org/10.1016/j.neuropsychologia.2013.11.010>.
- Hassabis, Demis, Dharshan Kumaran, and Eleanor A. Maguire. 2007. "Using Imagination to Understand the Neural Basis of Episodic Memory." *Journal of Neuroscience* 27 (52): 14365–74. <https://doi.org/10.1523/JNEUROSCI.4549-07.2007>.
- Hoerl, Christoph, and Teresa McCormack. 2016. "Making Decisions About the Future: Regret and the Cognitive Function of Episodic Memory." In *Seeing the Future: Theoretical Perspectives on Future-Oriented Mental Time Travel*, edited by Kourken Michaelian, Stanley Klein, and Karl Szpunar, 241–66. Oxford: Oxford University Press.
- Klein, Stanley B. 2014. "Autonoesis and Belief in a Personal Past: An Evolutionary Theory of Episodic Memory Indices." *Review of Philosophy and Psychology* 5 (3): 427–47. <https://doi.org/10.1007/s13164-014-0181-8>.
- Klein, Stanley B. 2015. "What Memory Is." *WIREs Cognitive Science* 6 (1): 1–38. <https://doi.org/10.1002/wcs.1333>.
- Klein, Stanley B., Leda Cosmides, John Tooby, and Sarah Chance. 2002. "Decisions and the Evolution of Memory: Multiple Systems, Multiple Functions." *Psychological Review* 109 (2): 306–29. <https://doi.org/10.1037/0033-295x.109.2.306>.
- Lee, Caroline S, Mariam Aly, and Christopher Baldassano. 2021. "Anticipation of Temporally Structured Events in the Brain." *ELife* 10: e64972. <https://doi.org/10.7554/eLife.64972>.
- Lengyel, Máté, and Peter Dayan. 2007. "Hippocampal Contributions to Control: The Third Way." *Advances in Neural Information Processing Systems* 20: 889–96.
- Loftus, Elizabeth F. 2005. "Planting Misinformation in the Human Mind: A 30-Year Investigation of the Malleability of Memory." *Learning & Memory* 12 (4): 361–66. <https://doi.org/10.1101/lm.94705>.
- Lu, Qihong, Uri Hasson, and Kenneth A. Norman. 2022. "A Neural Network Model of When to Retrieve and Encode Episodic Memories." *ELife* 11: e74445. <https://doi.org/10.7554/eLife.74445>.
- Mahr, Johannes B. 2019. "Thinking About the Past as the Past for the Past's Sake: Why Did Temporal Reasoning Evolve?" *Behavioral and Brain Sciences* 42, e262. <https://doi.org/10.1017/S0140525X19000402>.

- Mahr, Johannes B., and Gergely Csibra. 2018. "Why Do We Remember? The Communicative Function of Episodic Memory." *Behavioral and Brain Sciences* 41, e1. <https://doi.org/10.1017/S0140525X17000012>.
- Mar, Raymond A., and R. Nathan Spreng. 2018. "Episodic Memory Solves Both Social and Nonsocial Problems, and Evolved to Fulfill Many Different Functions." *Behavioral and Brain Sciences* 41, e20. <https://doi.org/10.1017/S0140525X17001418>.
- Mattar, Marcelo G., and Nathaniel D. Daw. 2018. "Prioritized Memory Access Explains Planning and Hippocampal Replay." *Nature Neuroscience* 21 (11): 1609–17. <https://doi.org/10.1038/s41593-018-0232-z>.
- McClelland, James L., Bruce L. McNaughton, and Randall C. O'Reilly. 1995. "Why There Are Complementary Learning Systems in the Hippocampus and Neocortex: Insights From the Successes and Failures of Connectionist Models of Learning and Memory." *Psychological Review* 102 (3): 419–57. <https://doi.org/10.1037/0033-295X.102.3.419>.
- Michaelian, Kourken. 2016. *Mental Time Travel: Episodic Memory and Our Knowledge of the Personal Past*. Cambridge, MA: MIT Press.
- Nagy, D., and Gergő Orbán. 2016. "Episodic Memory as a Prerequisite for Online Updates of Model Structure." *Proceedings of the 38th Annual Conference of the Cognitive Science Society*. <https://cogsci.mindmodeling.org/2016/papers/0465/paper0465.pdf>.
- Neisser, Ulric. 1981. "John Dean's Memory: A Case Study." *Cognition* 9 (1): 1–22. [https://doi.org/10.1016/0010-0277\(81\)90011-1](https://doi.org/10.1016/0010-0277(81)90011-1).
- Parker, Elizabeth S., Larry Cahill, and James L. McGaugh. 2006. "A Case of Unusual Autobiographical Remembering." *Neurocase* 12 (1): 35–49. <https://doi.org/10.1080/13554790500473680>.
- Piantadosi, Steven T. 2021. "The Computational Origin of Representation." *Minds and Machines* 31 (1): 1–58. <https://doi.org/10.1007/s11023-020-09540-9>.
- Rasmussen, Anne S., and Dorthe Berntsen. 2009. "The Possible Functions of Involuntary Autobiographical Memories." *Applied Cognitive Psychology* 23 (8): 1137–52. <https://doi.org/10.1002/acp.1615>.
- Roediger III, Henry L. 1996. "Memory Illusions." *Journal of Memory and Language* 35 (2): 76–100. <https://doi.org/10.1006/jmla.1996.0005>.
- Rouhani, Nina, Kenneth A. Norman, and Yael Niv. 2018. "Dissociable Effects of Surprising Rewards on Learning and Memory." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 44 (9): 1430–43. <https://doi.org/10.1037/xlm0000518>.
- Rubin, David C., and Sharda Umanath. 2015. "Event Memory: A Theory of Memory for Laboratory, Autobiographical, and Fictional Events." *Psychological Review* 122 (1): 1–23. <https://doi.org/10.1037/a0037907>.
- Schacter, Daniel L., Donna Rose Addis, and Randy L. Buckner. 2007. "Remembering the Past to Imagine the Future: The Prospective Brain." *Nature Reviews Neuroscience* 8 (9): 657–61. <https://doi.org/10.1038/nrn2213>.
- Schacter, Daniel L., Scott A. Guerin, and Peggy L. St Jacques. 2011. "Memory Distortion: An Adaptive Perspective." *Trends in Cognitive Sciences* 15 (10): 467–74. <https://doi.org/10.1016/j.tics.2011.08.004>.
- Schulz, Armin W., and Sarah Robins. 2022. "Episodic Memory, Simulated Future Planning, and Their Evolution." *Review of Philosophy and Psychology*, 1–22. <https://doi.org/10.1007/s13164-021-00601-1>.
- Schwartz, Arieh. 2020. "Simulationism and the Function(s) of Episodic Memory." *Review of Philosophy and Psychology* 11 (2): 487–505. <https://doi.org/10.1007/s13164-020-00461-1>.
- Suddendorf, Thomas, and Michael C. Corballis. 1997. "Mental Time Travel and the Evolution of the Human Mind." *Genetic Social and General Psychology Monographs* 123: 133–67.
- Suddendorf, Thomas, and Michael C. Corballis. 2007. "The Evolution of Foresight: What Is Mental Time Travel, and Is It Unique to Humans?" *Behavioral and Brain Sciences* 30 (3): 299–313. <https://doi.org/10.1017/S0140525X07001975>.
- Sutton, Richard, and Andrew Barto. 2018. *Reinforcement Learning: An Introduction*. Second edition. Cambridge, MA: MIT Press.
- Templer, Victoria L., and Robert R. Hampton. 2013. "Episodic Memory in Nonhuman Animals." *Current Biology* 23 (17): R801–6. <https://doi.org/10.1016/j.cub.2013.07.016>.
- Tulving, Endel. 1972. "Episodic and Semantic Memory." In *Organization of Memory*, edited by Endel Tulving and Wayne Donaldson, 381–403. New York, NY, US: Academic Press.
- Tulving, Endel. 2005. "Episodic Memory and Autoeosis: Uniquely Human?" In *The Missing Link in Cognition: Origins of Self-Reflective Consciousness*, edited by Herbert S. Terrace and Janet Metcalfe,

- 3–56. New York: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195161564.003.0001>.
- Tversky, Amos, and Daniel Kahneman. 1974. “Judgment under Uncertainty: Heuristics and Biases.” *Science* 185 (4157): 1124–31. <https://doi.org/10.1126/science.185.4157.1124>.
- Zentall, Thomas R., Tricia S. Clement, Ramesh S. Bhatt, and Jessica Allen. 2001. “Episodic-like Memory in Pigeons.” *Psychonomic Bulletin & Review* 8 (4): 685–90. <https://doi.org/10.3758/BF03196204>.
- Zentall, Thomas R., Rebecca A. Singer, and Jessica P. Stagner. 2008. “Episodic-like Memory: Pigeons Can Report Location Pecked When Unexpectedly Asked.” *Behavioural Processes* 79 (2): 93–98. <https://doi.org/10.1016/j.beproc.2008.05.003>.
- Zhou, Wenyi, Andrea G. Hohmann, and Jonathon D. Crystal. 2012. “Rats Answer an Unexpected Question after Incidental Encoding.” *Current Biology* 22 (12): 1149–53. <https://doi.org/10.1016/j.cub.2012.04.040>.