Stable Engrams and Neural Dynamics

Sarah K. Robins*†

The idea that remembering involves an engram, becoming stable and permanent via consolidation, has guided the neuroscience of memory since its inception. The shift to thinking of memory as continuous and dynamic, as part of a trend toward neural dynamics, has challenged this commitment, with some, such as Lynn Nadel, calling for "the demise of the fixed trace" and others, such as Alcino J. Silva, urging rejection of the "consolidation dogma." Does consideration of neural dynamics offer reasons to reject engram theory? No. I argue that they are compatible. At most, shifting to a dynamic view of neural processes compels revision of the implementational details.

1. Introduction. Neuroscientists and philosophers of neuroscience are increasingly interested in neural dynamics. This includes interest in thinking about the brain's entities and activities as fluid and probabilistic (Poldrack et al. 2015), as well as research projects modeling the brain as a dynamic system (Chialvo 2010). Philosophers of neuroscience have attended to this dynamic shift, asking about the ways it challenges our current explanatory and investigatory practices, most notably its compatibility with mechanistic explanation (Kaplan and Craver 2011) and its implications for our cognitive ontology (Burnston 2017; De Brigard 2017).

In this article, I focus on how neural dynamics are thought to challenge *engram theory*, the long-standing theoretical framework that guides the neurobiology of memory. The idea that remembering involves a fixed engram, which becomes stable and permanent as a result of consolidation, has been a guiding assumption of the neuroscientific study of memory since its inception. Increased attention to neural dynamics has led many to reconceive of memory as a continuous and dynamic process, challenging the commitment

*To contact the author, please write to: Philosophy Department, University of Kansas, 3090 Wescoe Hall, 1445 Jayhawk Blvd., Lawrence, KS 66045; e-mail: skrobins@ku.edu.

†Many thanks to Felipe De Brigard and Colin Klein for organizing the PSA 2018 symposium Neural Dynamics and Cognitive Ontology, where this article was originally presented. Thanks, too, for comments and feedback from the audience and from my fellow symposium participants: Jessica Cohen, Felipe De Brigard, Bryce Gessell, and Colin Klein.

Philosophy of Science, 87 (December 2020) pp. 1130–1139. 0031-8248/2020/8705-0030$10.00
Copyright 2020 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/710624 Published online by Cambridge University Press

to engrams and their consolidation that were previously understood as foundational. Importantly, this version of the dynamics challenge comes from neuroscientists themselves. A central theme in twenty-first-century neuroscience of memory is resistance to engram theory, with some calling for "the demise of the fixed trace" (Nadel 2007, 181) and others urging us to reject the "consolidation dogma" (Silva 2007, 168).

Does consideration of neural dynamics offer reasons to reject engram theory? No, I will argue: engram theory is compatible with neural dynamics. At most, shifting to a more dynamic view of neural processes will compel a revision in our understanding of some of engram theory's implementational details. Engram theory could of course still be rejected (e.g., one could decide to endorse an alternative, reconstructive theory of memory). However, these reasons for abandoning engram theory should be kept distinct from those stemming from attention to neural dynamics.

I begin in the next section with a review of the basic commitments of engram theory and then, in the section 3, turn to the challenges to engram theory raised in light of neural dynamics. In section 4, I respond to the challenge from neural dynamics.

2. Engram Theory. The neurobiology of memory is guided by engram theory, which involves two complementary commitments that frame inquiry into the cellular and molecular mechanisms and processes that support memory. First, remembering requires an *engram*. Second, this engram becomes stable and permanents as a result of *consolidation*. I discuss these two commitments in turn.

Engram.—The engram is a refashioning of the age-old memory trace: the entity responsible for forming, storing, and retrieving memories. The term was coined by the German biologist Richard Semon (1921). He postulated that there must be some such mechanism(s)—that is, a physical or chemical change to the brain as a result of learning, which when sustained over time supports the latter activity of remembering. Semon himself was agnostic as to what kind of thing the engram might be; in proposing that the engram must exist in order to explain remembering, Semon characterized it only as a change to the neural system, leaving open what kind of change or changes it might be. In arguing against the equipotentiality view of neural organization, Hebb (1949) advanced the more specific idea that the engram must be a structural change to the neural system. Hebb's theory of neural communication, coupled with the discovery from Milner and colleagues (Scoville and Milner 1957) that the hippocampus was the neural structure supporting long-term memory, led neurobiologists of memory to look for these structural changes in the synaptic connectivity between hippocampal neurons. Bliss and Lømo's (1973) discovery of long-term potentiation between these neurons provided the first concrete proposal for the engram's mechanistic basis (see Craver

https://doi.org/10.1086/710624 Published online by Cambridge University Press

2003). Our understanding of how hippocampal neurons communicate and change so as to support retention has been refined and revised over time, but nonetheless it remains the case that "the prevailing view is that the formation of an engram involves strengthening of synaptic connections between populations of neurons that are active during encoding" (Josselyn, Köhler, and Frankland 2015, 521).

Consolidation.—Semon himself did not make any claims about consolidation, but this second feature of contemporary engram theory also has a long history. The observation that newly acquired information takes a while to solidify into a long-term memory has been observed by many, from the Roman rhetorician Quintilian (1921) to Théodule Ribot (1882/1977), a doctor working with amnesic patients. Within contemporary neurobiology, observations of this facet of memory are paired with the commitment to engrams to yield a view of consolidation as the process by which the engram is stabilized and preserved to make possible long-term retention and recollection. There are two forms of consolidation, each with a distinct timescale. First, there is *synaptic consolidation*, which is the initial stabilization of the engram that occurs directly after learning (i.e., within the first few minutes or hours). Second, there is *systems consolidation*, where the engram moves from the hippocampus to frontal cortex to make room for the formation of new memories. The precise timescale for systems consolidation is a matter of some debate; it is thought to occur anywhere from a few weeks or months to possibly even years or decades after the initial engram formation (Nadel et al. 2007).

My focus will be on synaptic consolidation primarily, as it is the form of consolidation most closely tied to engram formation and thus most directly implicated in dynamic challenges to engram theory. Synaptic consolidation is considered to be a one-time occurrence, whereby the synthesis of proteins allows the synaptic connections involved in the act of initial learning event to become stabilized and hence more permanent. When such protein synthesis does not occur (e.g., if a protein synthesis inhibitor is administered), then the changes in synaptic connectivity are not maintained and the effects of learning disappear (i.e., forgetting occurs). As Dudai explains: "The current central dogma of synaptic consolidation is that it involves stimulus-induced activation of intracellular signaling cascades, resulting in posttranslational modifications, modulation of gene expression, and synthesis of gene products that alter synaptic efficacy" (2012, 228).

In characterizing the view of consolidation as a "dogma," Dudai is not alone. Earlier I quoted Silva (2007) as using the same term, and other examples are easily found. This reflects the depth to which this view of memory is embedded in the theory and practice of neurobiology. In short, those who study the cellular and molecular mechanisms of memory believe not only that there is an engram that is stable and persistent over time but that this engram is held in the changed strength of the synaptic connections that occur as the

https://doi.org/10.1086/710624 Published online by Cambridge University Press

result learning and are solidified through changes to the protein structure of the involved neurons.

3. The Challenge from Neural Dynamics. Engram theory has its basis in the preservative nature of memory. Memory is a capacity for retaining information, ideas, experiences, and skills over time. The study of memory, in cellular and molecular neuroscience, is a search for the mechanisms and processes that make such continuity and preservation possible. In short, engram theory requires neural stability. The engram is supposed to be a fixed neural structure, whose stability is made possible by the process of consolidation. Thinking of neural processes in dynamical terms is presumed to be a direct threat to this characterization, as it involves denying that any neural processes or structures have the needed stability. Insofar as engram theory requires discrete, isolable, and static neural mechanisms, then it looks like there is no place for it in a view of neural processes as fluid, dynamic, and probabilistic. And so, despite the dogmatic stature of engram theory, many neurobiologists of memory now believe that the framework is inescapably challenged by the shift to neural dynamics.

The challenge from neural dynamics is thought to affect both components of engram theory—the engram and its consolidation. Increased attention to neural dynamics encourages a rethinking of the engram itself. The implementational details belie the structural change that was meant to be characteristic of engrams. Insofar as there is broad agreement that the engram resides in synaptic connections between hippocampal neurons, the "entity" that is the centerpiece of engram theory is not really an entity or structured process in any strong sense. What is more, these changes in synaptic connectivity are instances of synaptic plasticity more generally. Instead of identifying a mechanism that is the marker of stability, engram theory appears to be reliant on processes characterized by their malleable nature: "The activation of αCaMKII, CREB, and all of the other molecules implicated in memory, the induction and stability of synaptic plasticity, as well as the circuit processes that support the representations of information in the brain are all highly dynamic and probabilistic in nature and they do not reflect the determinism and 'solidity' of the consolidation dogma" (Silva 2007, 168).

Attention to neural dynamics further threatens the one-time-and-for-all-time view of synaptic consolidation. Over the past 2 decades, a steady stream of research into the mechanisms of consolidation has revealed that engrams can undergo a period of vulnerability when they are reactivated, allowing the established synaptic connections to change (Sara 2000). Even though this later process is not identical to the initial consolidation, it has become known as *reconsolidation* (Nader, Schafe, and LeDoux 2000). The precise circumstances in which reconsolidation occurs are still a matter of some debate; it does not appear to be the case that the engram becomes labile during every

https://doi.org/10.1086/710624 Published online by Cambridge University Press

instance of reactivation (e.g., Eisenberg et al. 2003). Nonetheless, there is now enough evidence of reconsolidation for most neuroscientists to be wary of the standard interpretation of synaptic consolidation, claiming instead that "consolidations never end" (Dudai 2012, 227). Together, these ways of reconceptualizing the engram and its consolidation introduce fluidity and variability into the formation and retention of engrams. Reflecting on this, Silva asks, "how could something that was supposed to be so solid and reliable be fluid every time it is used?" (2007, 169).

It may strike the reader as curious why neuroscientists whose work is grounded in engram theory would advocate for upending the framework that guides inquiry into the memory systems they study. Although the rejection of this dogma is admittedly destabilizing, it is thought to be a move in the right direction, not only because it aligns with the general trend toward neural dynamics but also because of the way in which it seems to parallel developments in thinking about the nature of memory in other areas of cognitive science, most notably cognitive neuroscience and psychology. In these fields, the understanding of memory has shifted quite radically in recent decades, from a preservative model to a reconstructive one.

The reconstructive view portrays memory not as a preservative capacity for retaining information but as a generative and transformative capacity for creating representations of past events as needed for future purposes. In psychology, versions of this view have been around at least since Bartlett (1932), but it is only more recently that the view has become more popular and arguably the dominant view among cognitive scientists. The shift is due to two large bodies of evidence. First, there is extensive evidence of the malleability of memory. Recollections of past events and information are easily and systematically distorted, which many believe puts pressure on the view of memory as preservative. Rather than characterize memory as a capacity that repeatedly and systematically malfunctions, many now prefer an alternative characterization of memory's function, one that highlights its dynamical and constructive nature (e.g., De Brigard 2014). Second, the last decade has seen an accumulation of neuroimaging evidence that shows significant overlap in the brain regions involved in remembering and those involved in imagination and other forms of personal, episodic simulation (e.g., Schacter et al. 2012). Many assume that these shared neural structures indicate a stronger set of similarities than differences between these capacities and, thus, that considerations of memory as a distinct capacity dedicated to retention should be discarded.

By changing how we think about the neurobiological mechanisms of memory, from fixed and permanent structural changes to fluid systems that are updated dynamically and continuously, these scientists see themselves as aligning their work on memory with the more general reconstructive approach found in psychology and cognitive neuroscience. Dudai, for example,

https://doi.org/10.1086/710624 Published online by Cambridge University Press

claims that the move away from engram theory and its consolidation dogma allows that "the neuroscience of memory reconciles with the intuitive, dynamic view of memory that dominates the cognitive sciences" (2012, 241). Others advocate for a stronger relationship between the two projects, arguing that the dynamical processes of reconsolidation are the mechanisms of reconstructive memory. As Nadel puts the point, "recurrent 're-encoding' or 'reconsolidation' and the underlying instability they reveal, are signatures of this transformative capacity of memory systems" (2007, 180).

4. Evaluating the Challenge from Neural Dynamics. Does the neural dynamics framework pose a threat to engram theory? Answering this question definitively is beyond the scope of this article. Instead, what I want to do is suggest that the challenge is not as straightforward or as necessary as its advocates have supposed. As a first point, it is worth noting that the challenge from neural dynamics is poorly matched to engram theory. It confuses distinct levels of explanation to assume that all features of explanation at one level must be mirrored, copied, or respected at others. The neural system can have features that the cognitive processes it supports do not (and vice versa). In other words, neural dynamics are compatible with functional stability. If neural dynamics were a challenge to the functional stability of engrams, then they would present similar problems for a range of other mental processes—focused attention, perceptual vigilance, or even thinking the same thought on two different days. Remembering need not be dynamic simply because the underlying neural systems that support it are dynamic.

Even in light of this general observation, the concern about the compatibility of engrams and neural dynamics might persist. After all, the claim about stability at the heart of engram theory is not only a claim about the preservative function of memory; it is a claim about the features the underlying mechanism needs to have in order to support remembering. In order to carry out this particular function (i.e., long-term retention of information) neural mechanisms must be such that the influence of learning can be retained over an extended period of time. The motivation for engram theory is not a simple assumption that functions and structures should match but a more particular claim that this *particular* type of function requires a *particular* type of structure. Engram theory makes demands on the neural structure; proponents of neural dynamics are pointing out that neural systems do not have the requisite structure.

This form of the challenge does present a more serious test to engram theory, so it is worth pausing to consider whether it adequately reflects the demands that engram theory makes on the underlying neural structure. As discussed in section 2, the work of Semon (1921), Hebb (1949), and others shows that engram theory is a commitment to the claim that remembering requires a persistent, structural change to the underlying neural system. A similar commitment can be found in more general accounts of memory traces (Robins

https://doi.org/10.1086/710624 Published online by Cambridge University Press

2017). The commitment to an engram as a persistent structural change is compatible with many if not all of the claims about neural dynamics. Consider two central claims about the dynamic nature of engrams:

a) Sustaining the engram's persistent structure requires continual activity, upkeep, and dynamic activity, and
b) The structural basis of the engram can change over time (i.e., the same engram can be supported by distinct structures at different points in time).

Engram theory is compatible with both. On point *a*, the engram can persist and undergo changes simultaneously. Engram theory need not be interpreted as the claim that whatever the engram is it must remain unchanged in order to continue serving as the engram. More strongly, it should not be interpreted in this way. The engram must persist in order to support remembering, but since it is a biological mechanism it is to be expected that its persistence will require constant regeneration. And further, even if certain activities, like reconsolidation, are shown to alter the engram, this is still compatible with their being an engram. The engram may change in some ways (e.g., be altered or weakened) but nonetheless persist in being an engram.

On point *b*, it is worth pointing out that engram theory, in its current nondynamically informed state, already allows for significant changes in the particular physical structure that supports the engram over time. Systems consolidation, as introduced in section 2, is the process by which a single engram is transferred from the hippocampus to frontal cortex. Such transfer does not involve the migration of neurons but the transfer of information. If such transfer is compatible with the understanding of 'persistent structure' in engram theory, then surely more local changes—however dynamic—will be too. Neuroscientists who remain committed to engram theory are happy to make this point explicit: "although the engram is a moving target over time, this characteristic does not preclude tractability and success in capturing the engram at any given moment in time" (Josselyn et al. 2015, 522).

The commitment to engrams seems compatible with neural dynamics. The other component of engram theory—consolidation—may not be as fortunate. Insofar as consolidation "dogma" requires thinking of the process of stabilizing the engram to occur only once and in a particular way, then this component of the framework may be effectively displaced by discoveries about neural dynamics. Even if the view managed to reach the status of dogma, it is unclear that it should have been allowed to do so, given how its implementational details have been understood (even before work on reconsolidation). After all, nothing in the initial observations of how memory stabilizes over time implies that the process by which this stabilization occurs must only happen once. It is also unclear why there would be emphasis on consolidation as a one-and-done

https://doi.org/10.1086/710624 Published online by Cambridge University Press

process, given the widespread recognition of systems consolidation. Many may have thought that synaptic consolidation occurs only once, but this seems to be a claim about the particular features of the mechanism we discovered to be involved in consolidation, not a general claim about how consolidation must go in order to be the process of stabilizing a memory. The claim that there are engrams that are made more persistent or permanent over time can remain in place, even as our understanding of the implementational details of how this is achieved change.

Furthermore, it is wrong to pair the study of reconsolidation with the recent interest in neural dynamics too closely, as evidence of reconsolidation processes has been available and studied for much longer—a point nicely illustrated by Lewis's (1979) review of reconsolidation evidence from the prior few decades.[1] This broader perspective on the study of consolidation reveals far less dogma than a focus on current debates would suggest. Our understanding of consolidation processes has long been in flux. And importantly, up until now, most have been happy to see those changes as ones that can happen from within engram theory. Shifts to the implementational details would be just that: shifts, not reasons to overhaul engram theory in its entirety. Neural dynamics may provide additional reasons to challenge this dogmatic view of consolidation, but several independent reasons to be suspicious were already available.

If the challenge from neural dynamics does not overturn engram theory, then the anticipated link between the mechanisms of memory and its reconstructive nature is disconnected. This may be disappointing to those who were looking for reasons to move the neurobiology of memory away from the preservative model of memory that has kept it out of pace with this broader revision in how memory is portrayed. But there are a number of reasons, beyond this disconnect, for why these two projects were unlikely to align. First, there are significant mismatches between reconsolidation and reconstruction, which make it difficult to support the view that the former is a mechanism for producing the latter. Reconsolidation is a widespread phenomenon, occurring across a range of neuron types and in a range of species. Reconstruction, however, is supposed to be not only unique to memory but, on many views, unique to humans who possess a particularly robust form of episodic memory (Schacter et al. 2012). In addition, as mentioned in the previous section, reconsolidation does not occur with every reactivation of an engram. But reconstruction, as a claim about the general process of memory retrieval, must occur every time there is reactivation or an attempt at recollection. Second, if

1. Dudai (2012, 229–30) makes an interesting bibliometric observation on this point: from 1993 to 1999 only 6 of the 27,061 papers on memory in Thomson Reuters Science Web of Knowledge mentioned "reconsolidation," while 413 of the 61,950 such papers between 2001 and 2010 did.

https://doi.org/10.1086/710624 Published online by Cambridge University Press

one wants to endorse a strongly reconstructive view of memory—according to which no particular trace or engram is maintained from past experience, only general patterns of information amassed from multiple experiences—then it seems it would be more profitable to abandon talk of consolidation and reconsolidation, as both are processes that occur to an engram over time. If this form of reconstruction is right, there may not be any such engram to reconsolidate.

There are also more general reasons to keep these two projects distinct: claims about the neural mechanisms of memory and the overall nature of memory processes are orthogonal. Any neuroscientist who wants to shift from engram theory to a more reconstructive model of memory can do just that. There is no need to take the argumentative route through neural dynamics. Just as engram theory can contort as needed to accommodate whatever details we discover about the brain, dynamic or otherwise, so too can reconstructive views. The framework one uses to investigate the neural processes involved in remembering influences the eventual interpretation of what is discovered. One such theoretical perspective may turn out to be more empirically plausible than another, but there is no straightforward empirical test to determine how memory's function should be characterized.

Increased attention to neural dynamics poses a range of interesting questions for neuroscientists and philosophers of science. And recent discoveries about memory, everything from neural reconsolidation through to the overlap in neural structures used in remembering and imagining, provide plenty of challenges to traditional conceptions of this capacity. But the attempt to link these two projects may not be of much use, at least not for providing reasons to reject engram theory in favor of reconstruction.

5. Conclusion. Neurobiologists of memory who hope to jettison engram theory and its consolidation dogma in light of neural dynamics make two claims. First, that engram theory—in both its general structure and particular implementational details—is in conflict with current understanding of neural dynamics. Second, rejecting engram theory in light of considerations from neural dynamics will make possible an understanding of memory in cellular and molecular neuroscience that is more compatible with the reconstructive view of memory held by most cognitive scientists. In this article I have argued both claims. On the first point, I have shown that a more thorough understanding of engram theory makes it seem unlikely that neural dynamics compels such a rejection. At most, some changes to the implementational details, particularly of consolidation, may be required. But otherwise persistent engrams are compatible with neural change. Second, the relationship between neural dynamics and reconstructive memory is not as well supported as many have hoped. Questions about the dynamic nature of memory mechanisms and questions about memory's reconstructive nature should be pursued independently.

https://doi.org/10.1086/710624 Published online by Cambridge University Press

REFERENCES

Bartlett, Frederick. 1932. *Remembering: A Study in Experimental and Social Psychology*. New York: Cambridge University Press.
Bliss, T. V., and T. Lømo. 1973. "Long-Lasting Potentiation of Synaptic Transmission in the Dentate Area of the Anesthetized Rabbit following Stimulation of the Performant Path." *Journal of Physiology* 232:331–56.
Burnston, Daniel C. 2017. "Interface Problems in the Explanation of Action." *Philosophical Explorations* 20:242–58.
Chialvo, Dante R. 2010. "Emergent Complex Neural Dynamics." *Nature Physics* 6:744–50.
Craver, Carl F. 2003. "The Making of a Memory Mechanism." *Journal of the History of Biology* 36:153–95.
De Brigard, Felipe. 2014. "Is Memory for Remembering? Recollection as a Form of Episodic Hypothetical Thinking." *Synthese* 191:155–85.
———. 2017. "Cognitive Systems and the Changing Brain." *Philosophical Explorations* 20:224–41.
Dudai, Yadin. 2012. "The Restless Engram: Consolidations Never End." *Annual Review of Neuroscience* 35:227–47.
Eisenberg, M., T. Kobilo, D. E. Berman, and Y. Dudai. 2003. "Stability of Retrieved Memory: Inverse Correlation with Trace Dominance." *Science* 301:1102–4.
Hebb, Donald. 1949. *The Organization of Behavior: A Neuropsychological Theory*. Oxford: Wiley.
Josselyn, Sheena A., Stefan Köhler, and Paul W. Frankland. 2015. "Finding the Engram." *Nature Reviews Neuroscience* 16:521–34.
Kaplan, David, and Carl F. Craver. 2011. "The Explanatory Force of Dynamical and Mathematical Models in Neuroscience." *Philosophy of Science* 78:601–27.
Lewis, D. J. 1979. "Psychobiology of Active and Inactive Memory." *Psychological Bulletin* 86:1054–83.
Nadel, L. 2007. "Consolidation: The Demise of the Fixed Trace." In *Science of Memory Concepts*, ed. H. L. Roediger, Y. Dudai, and S. M. Fitzpatrick, 177–82. Oxford: Oxford University Press.
Nadel, L., G. Winocur, L. Ryan, and M. Moscovitch. 2007. "Systems Consolidation and Hippocampus: Two Views." *Debates in Neuroscience* 1:55–66.
Nader, K., G. E. Schafe, and J. E. LeDoux. 2000. "Fear Memories Require Protein Synthesis in the Amygdala for Reconsolidation after Retrieval." *Nature* 406:722–26.
Poldrack, Russ, et al. 2015. "Long-Term Neural and Physiological Phenotyping of a Single Human." *Nature Communications* 6:1–15.
Quintilian. 1921. *Institutio Oratoria*. London: Loeb Classic.
Ribot, T. A. 1882/1977. *Diseases of Memory*. Washington, DC: University.
Robins, Sarah K. 2017. "Memory Traces." In *Routledge Handbook of the Philosophy of Memory*, ed. S. Bernecker and K. Michaelian, 76–87. London: Routledge.
Sara, S. J. 2000. "Retrieval and Reconsolidation: Toward a Neurobiology of Remembering." *Learning and Memory* 7:73–84.
Schacter, D. L., D. R. Addis, D. Hassabis, V. C. Martin, R. N. Spreng, and K. K. Szpunar. 2012. "The Future of Memory: Remembering, Imagining, and the Brain." *Neuron* 76:677–94.
Scoville, W. B., and B. Milner. 1957. "Loss of Recent Memory after Bilateral Hippocampal Lesions." *Journal of Neurology, Neurosurgery, and Psychiatry* 20:11–21.
Semon, R. 1921. *The Mneme*. London: Allen & Unwin.
Silva, A. J. 2007. "Consolidation: Molecular Restlessness." In *Science of Memory Concepts*, ed. H. L. Roediger, Y. Dudai, and S. M. Fitzpatrick, 167–70. Oxford: Oxford University Press.

https://doi.org/10.1086/710624 Published online by Cambridge University Press