Memory Engrams and Vehicle Indeterminacy

Caitlin Mace (caitlin.mace@pitt.edu)[1]
[1]Department of History and Philosophy of Science, University of Pittsburgh

Abstract

Memory engrams are hypothetical neural representations that explain the persistence of memory. Finding memory engrams requires the identification and individuation of engram vehicles, which are determinate neural entities or processes bearing memory content. However, not only are there many candidate vehicles for memory engrams, but each candidate is ill-suited for explaining memory persistence due to instability via degradation, turnover, and dynamics. I argue that memory engram models presume the truth of a mere heuristic: that changes in neural activities correspond to changes in cognitive content. As such, explanations of persisting memories are thought to require persisting vehicles. This presumption in turn drives issues with determining exactly what remains stable in the brain to explain how memory persists. But to explain the persistence of memory, all that needs to persist is the capacity of the system to recall a memory. I conceive of engrams according to their functional role as causal motifs for memory recall.

Keywords: memory engrams; neural representations; vehicle indeterminacy

Introduction

Neural representations are central posits in neuroscience. These representations are patterns of neural activity that perform some function for the brain. In theorizing about representations, and no less for neural representations, it is helpful to distinguish the representational vehicle from its content. Doing so allows us to theorize about what is doing the representing and how. For neural representations, representational content might be some environmental stimulus that is tracked by an organism, such as the orientation of a line, the trajectory of a moving limb, or fruit in some foraging patch. The vehicle of neural representational content is some pattern of neural activity, such as firing rates in a single neuron, molecular mechanisms of protein modification, or co-activity patterns in a neuronal ensemble. A major goal of neuroscientific inquiry is to identify and individuate vehicles of neural representations to understand how they represent what they do.

Consider the memory engram—a hypothetical neural representation of a memory. The engram's representational role is in producing memory retrieval such that the representational content is the particular memory. I take the representational role of engrams to provide an explanation of memory in terms of memory persistence. This conception of engrams accommodates the full taxonomy of memory, such that engrams could be neural representations of remembered experiences, places, facts, or skills. While memories of experiences and places are liable to error, lending to a view in which memory is reconstructed rather than stored, I take it that reconstructing a memory of a past experience still requires storing content. As such, I talk throughout this paper of memory being stored. As for the engram's vehicle, it is unknown what precisely the vehicle is, meaning that there is uncertainty about the physical constitution of memory engrams. Dudai (2002), echoing Lashley's (1950) sentiments many decades earlier, queries: "Where is [the engram] within that area [of the brain]? In a dedicated circuit? In identified neurons? In synapses? To what size of circuit do we hope to nail it down? What answer will finally satisfy our urge to localize things in the world?" Debates about what exactly the engram is and where it is to be found are ongoing, with interlocutors presenting evidence that engrams are molecular, synaptic, or in supracellular populations of neurons (Colaço and Najenson, 2024; Mace, O'Sullivan, & Wilson, 2026). Settling these debates requires understanding how vehicles of neural representations should be identified.

A theory of vehicles should tell us what they are and how they bear content. Hypotheses about vehicle-content relations constrain the pool of properties and processes that are worthy vehicle candidates. For example, to explain memory storage, one hypothesis is that the vehicle candidate should persist in the duration between the encoding event and later retrieval. This is one clue for determining which properties of the vehicle are semantic—that is, which properties bear content for later memory recall. A complete theory will have these properties right so that we can explain how they bear content. An analogy can be made to the vehicles of language representations. Should we choose to identify and individuate language symbols by the color of ink used in their inscriptions rather than by the marks that conventionally represent some word or phoneme in a certain context, we would have an unsuitable basis for theorizing about how words get their meaning. (Which, in this case, is by convention.) While we want to be sure we identify vehicles according to semantic rather than non-semantic properties to understand how the vehicle bears or evokes content, what these properties are in the brain remains unknown.

In this paper, I problematize the appeal to engram vehicles in explanations of memory. I first introduce memory engrams and their vehicle candidates in the section **Memory Engrams**, then I introduce vehicle indeterminacy issues concerning them in the section **Vehicle Indeterminacy**. In the section **Persisting and Dynamic Engrams**, I argue that the candidate vehicles do not appear to fulfill their explanatory role, as none of them persist for the duration of memory. I develop an account in **Revisioning Engrams** that reconciles neural dynamics with persisting memories to support theorizing about vehicles and their contents. Researchers identify engrams by their functional role in

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

causing memory recall, which lends to a conception of engrams as causal motifs.

Memory Engrams

Memory engrams were originally conceived of by Richard Semon (1921) as units of memory storage (i.e., engraphy) and retrieval (i.e., ecphory). After a century of searching for such engrams was not met with success, the search was recently reinvigorated by the development of tools that allow manipulation of putative memory engrams (Robins, 2023). Today, memory engrams are localized to neuronal ensembles, which are mesoscale structures composed of populations of neurons. Insofar as neuronal ensembles are or contain vehicles of memories, it remains an open empirical question as to what features or properties of ensembles afford memory engrams their representational role. The main criterion for engram status that researchers are concerned with is stability for the length of the memory, that is, that the relevant property persists for as long as the memory persists. While we would expect some changes in the vehicle, which would explain memory errors and memory degradation over time, an explanation of memory storage, in memory researchers' lights, requires something that persists.

There are roughly three postulated vehicles of memory for neuronal ensembles: molecular, synaptic, and population dynamics. The *synaptic* hypothesis is the most accepted in neuroscience as it has promoted a successful research program that bridges Hebbian theorizing about synaptic function (Hebb, 1949) with discoveries of synaptic processes that result from learning (Bliss & Lømo, 1973). The rough idea is that learning produces changes in neural connectivity via modified synaptic structures and processes such that the same neurons tend to be activated together. Such connectivity creates patterns of activity that are more likely to recur, thereby providing a means of storing information that can later be retrieved. While the synaptic hypothesis is the most widely accepted explanation of memory storage, persistence, and retrieval, a challenge for the synaptic hypothesis is that synaptic modifications are unstable and do not last long enough to explain long term memory storage (Trettenbrein, 2016; Gershman, 2023).

The *population* hypothesis has gained increased acceptance with optogenetic research and the rise of the population doctrine in neuroscience. Ensembles are conceived of as populations of neurons that exhibit recurring coordinated activity patterns (Yuste, Cossart, & Yaksi, 2024). Optogenetic research has shown that populations of neurons active during a learning episode can be reactivated to cause stimulus-independent, memory-related behavior (e.g., Liu et al., 2012; Ramirez et al., 2014). Multivariate statistical analyses have similarly targeted populations of neurons and provide additional correlational evidence for the ensemble hypothesis (Bergstrom, McDonald, & Johnson, 2011). Some have advocated for an ensemble hypothesis in defending the synaptic theory of memory, such as Langille and Brown (2018) who argue that the locus of memory is a population of neurons whose activities are formed, modified, and expressed through plasticity of the synapse and of internal molecular mechanisms. One challenge with the ensemble hypothesis, discussed in further detail below, is representational drift, a phenomenon in which neurons change their activity profiles in relations to a particular stimulus over time (Schoonover et al., 2021). Here, again, ensembles do not seem to persist long enough to explain the persistence of memory.

Finally, the *molecular* hypothesis has been a contentious contender for finding and manipulating engrams. The roots of the hypothesis go back to early experiments in which researchers attempted to transfer memory from one organism to another—either by reproduction across generations (McDougall, 1927, 1930, and 1938; Rhine & McDougall, 1933) or across individuals (McConnell, 1962; Babich, Jacobson, & Bubash, 1965; Babich et al., 1965). These experiments purported to show that trained organisms created memory engrams that could be passed on genetically, by injection, or, quite crudely, as food. (The planaria that McConnell trained were ground up and fed to other planaria.) The sign of successful transfer was that training was much quicker in the organism that received the putative engram. Explaining such results involved appeals to genetic material like RNA that could serve as the medium of transfer. Theoretical considerations support a role for genetic material as engrams since it is well accepted that such genetic material encodes information. Indeed, 'encoding information' seems to be a more meaningful metaphor if genetic material is the substrate of memory (see Gallistel, 2017). Moreover, given advances in understanding epigenetics, which posits mechanisms for environmental-genetic relations, there is a plausible mechanism for encoding experience (Gold & Glanzman, 2021), which offers a potential explanation for how engrams bear content. However, it remains an open question what exactly the molecular substrate of memory is, and experimental evidence for this hypothesis is scarce compared to evidence supporting the synaptic and population hypotheses. Additionally, token RNA strands have short half-lives and their sequences often change due to editing and degradation, again calling to question whether this engram candidate can explain memory persistence.

Importantly, there are many other kinds of putative vehicles for neural representations that can be realized by molecules, individual neurons, or populations of neurons. Some candidate vehicles are spikes (with debate about whether rate, timing, or synchronization is important), manifolds (which are abstracted mathematical structures describing population activity that only contentiously can serve as vehicles), and ensemble structure. I focus on the vehicle candidates here that have been most successful for studying memory storage. What is particularly striking about the vehicle candidates is that they are such disparate entities that it seems we do not even have the first step for determining the vehicles of memory engrams. In this next section, I argue that the first step is made challenging by vehicle indeterminacy issues.

Vehicle Indeterminacy

An important consideration for identifying neural vehicles is that many structures and processes may perform a merely supportive role in memory. For the candidate vehicles, whether they are part of the vehicle or play a supportive role in providing a means for the representation to be realized is the subject of debate. The difference between vehicles and realizers is not always clear and will depend on how each notion is cashed out. Drayson (2018) takes realizers to be physical structures that fulfill computational roles whereas vehicles are abstract symbolic structures. In cases where those physical structures fulfill a representation role and can be treated as standing in for certain contents—as a symbol would—I think it is legitimate to treat those physical structures as vehicles. In this case, vehicles are also realizers. But the distinction between realizers and vehicles does not collapse because there are mechanisms for the vehicle's formation, maintenance, and expression that are realizers but not vehicles. The issue we are addressing is in determining that candidate vehicles are not mere realizers. In either case, parsing out realizers—their componentry and the conditions for their formation and termination—matters for determining their function (Sullivan, 2008), and thus, for determining whether they fulfill their explanatory role.

In the previous section, I claimed that the primary vehicle candidates for memory engrams are each unbefitting of their given explanatory role, which is to explain the persistence of memory. This is the issue on which interlocutors focus in neuroscientific debates about memory engrams. But vehicle indeterminacy is a more pervasive problem each position faces. This problem is most readily demonstrated with populations of neurons. Ebitz and Hayden (2021) write, "The limits of a neuron are obvious—it has cell walls—but what are the limits of a population? Are its boundaries the set of recorded neurons? The tissue surrounding the electrodes? The edges of the Broadman [sic] area? The skull?" (p. 3063). In a popular neuroscience blog, The Transmitter, Mark Humphries (2025) argued that populations are either arbitrarily defined by recording tools or given physical bounds that ignore the rich connectivity neurons have with most other neurons. Far from a mere object indeterminacy, these issues raise questions about the relevant vehicle properties of neural populations. Without knowing what the relevant population is, we cannot hope to determine whether one can approximate the activity of the population by randomly sampling neurons across various trials (as some researchers do, see Ebitz and Hayden, 2021) or whether the relevant property of population activity requires consideration of the simultaneous activities of all (or some substantial subset) of its individual neurons.

Now memory engram research tends to define ensembles with labeling techniques involving immediate early gene expression. Researchers first label neurons using a gene that is expressed immediately upon stimulus onset called *c-Fos*. The protein product of *c-Fos*, also called c-Fos, activates the transcription of other genes. Researchers track *c-Fos* gene expression by tagging the c-Fos proteins. Both *c-Fos* and c-Fos are responsive to various stimuli, ranging from neurotransmitters and ions to mechanical injury and noxious stimuli to light and stress (Lara Aparicio et al., 2022). Because *c-Fos* and c-Fos activation occurs with a wide array of stimuli, tagging c-Fos proteins allows researchers to use *c-Fos* as a marker of neural activity. In particular, one immunohistochemical strategy uses two antibodies to label *c-Fos*-activated cells. One antibody binds specifically to c-Fos proteins. A second antibody containing a fluorescent marker binds specifically to the first antibody, which makes the tagged neurons visible to researchers. Notice that this technique is four steps removed from the population of active neurons—the population is marked by a gene whose expression is tracked by its protein product whose presence is tagged by antibodies whose presence is made visible to researchers by another antibody. Ideally, the population visible to the researchers at the end of all of this tagging is the same as the one that was activated. And after much engineering of this population with viral vectors and genetic transduction, researchers using optogenetics also bank on the *c-Fos*-tagged population of active neurons being the same one that was optogenetically treated and activated. Yet optogenetic intervention on the population of neurons is inexact because the light scatters, likely activating other neurons. The problem being illustrated here is determining the population of neurons that serve as an engram.

Vehicle indeterminacy for molecules and synapses is less obvious but just as indeterminate. Consider long-term potentiation, the primary synaptic process posited in explanations of memory. Long term-potentiation is the modification of pre- and post-synaptic structures and processes that permit increased interaction between neurons. In particular, pre-synaptic vesicles—the structures that release neurotransmitters into the synapse—are more readily released and recycled, and the post-synaptic membrane has more receptors for those neurotransmitters. Other structural synaptic changes occur as a result of learning, such as the growth of dendritic spines and other alterations in their shape and size, weakened synaptic connections in the target area or other areas, enzyme activation, neurotrophin release, and change in gene expression. Now, one might think that we can get an empirical handle on which of these structures and processes are necessary for the synaptic mechanism to serve as a memory engram. It is likely all of them. In treating synapses as vehicles rather than mere realizers, however, determining which synaptic processes or structures compose the vehicle is a question that, as far as I know, has not been addressed. This is largely because experiments will focus on one or a few synaptic processes at a time. Far from being an error on neuroscientists' part, the limited focus is due to a desire to understand synaptic processes rather than to determine all of the synaptic processes or structures that are part of a particular vehicle. While some labeling techniques can tag synapses that express activity at a particular time, this labeling does not discriminate between increased or reduced activity (Eom, Kim, and Ho Hyun, 2025) and glosses over numerous synaptic processes. Thus, it remains an open

question where the vehicle is and what its boundaries are, answers to which are pre-requisite for determining vehicles and vehicle properties, at least without a theory of vehicles. Determining which synapses are part of the vehicle, in turn, will raise issues similar to those faced by researchers individuating populations of neurons, namely which properties and processes are relevant for serving as a vehicle.

As for molecules, consider RNA, the primary molecular posit in explanations of memory. Experiments have shown that a transfer of RNA between organisms appears to facilitate training outcomes in the recipient organisms. Moore et al. (2021) conducted transfer experiments that involved having nematodes ingest homogenates—a suspension of cellular fragments that releases intracellular material like RNA—that were processed from the cells of nematodes who learned to avoid a particular kind of bacteria. Not only did the nematodes who ingested the homogenates produce the avoidance behavior that was learned in the donor nematodes, but the ingesting nematodes' offspring also exhibited this avoidance behavior. It is, of course, an open question whether the RNA is a vehicle in these experiments. Whether the RNA is a representational vehicle partly depends on which kind of RNA is injected—while many RNA molecules code for proteins, some RNA molecules perform other roles such as regulate genes, coordinate epigenetic modifications, inhibit protein activity, and others. It is possible, then, that what is being transferred is instructions for creating vehicles. I set this worry aside for the sake of argument, since the hypothesis here is that the RNA is the vehicle. The ensuing problem is that we are left unsure which particular sequence or structure of the RNA has a representational role. Indeed, the relevant structures may be particular polynucleotides (see Gallistel, 2017, 2021), and which polynucleotide sequence is the memory engram is far from being determined. The upshot is that, without an understanding of what RNA does to perform its role as an engram, we cannot determine what part of the RNA is a vehicle. Thus, here again, we do not have a definite vehicle.

One might reasonably wonder at this point why vehicles need definite boundaries. Why can't we just accept that vehicles are fuzzy objects with blurry or dynamic boundaries, like shadows of clouds or the population of New Mexico? Here I want to press that the indeterminacy issues I'm raising are epistemic. Having a boundary, even one that is coarse or ill-defined, shows that the vehicle has been individuated and thus identified. But the boundaries are currently so coarse or ill-defined that neuroscientific research on engram vehicles does not rule out other candidate vehicles. Thus, an epistemic indeterminacy remains about which details matter and are relevant to explaining memory.

Now it may be the case that all of the details matter, and that experiences and learning leave many kinds of traces in the brain (Eom, Kim, and Ho Hyun, 2025). There is some debate over whether and in what ways these vehicle kinds might all be part of an engram. Some may prefer a multi-level approach that accommodates each vehicle candidate (e.g., Craver 2007), and I suspect that there will be a role for each candidate in a mechanistic model of memory, which will include components and operations with various supportive roles. But to explain memory via the postulation of vehicles, the candidate ought to have a representational role, and we do not yet know which candidate, if any, has a representational role. Some states, processes, or mechanisms may play a realization role, such as by providing an index or pointer to vehicles in other areas of the brain. (See pointer or index views of hippocampal function (Teyler and DiScenna, 1986; Goode et al., 2021) and minimal trace theories (Werning, 2020) that place vehicles for memory content in the cortex). Even if we were to accept that there are many kinds of memory engram vehicles, which is a common view following analyses of scientific practice, such a pluralism should be principled; we may let a thousand flowers bloom yet believe that only some of those flowers are well-suited for the various roles to which we subject them. As science progresses, we can expect much pruning. I offer criteria for ascribing engram status in the next section and diagnosis the problem of vehicle indeterminacy.

Persisting and Dynamic Engrams

Criteria have been offered for memory engram status that may help to parse out which candidates are worthy of memory engram status. The following criteria were developed in previous work (Mace, 2021, 86f) based on desiderata put forward by (Martin, Grimwood, and Morris, 2000; Bickle, 2008; Silva, Landreth, & Bickle, 2013; Mayford, 2014):

1. *Observation*: Occurrences of learning-induced changes in a population of neurons must be strongly correlated with occurrences of the behaviors used as experimental measures of a memory.

2. *Negative alteration*: Intervening directly to decrease, block, or impair the learning-induced changes of the population of neurons must reliably decrease the behaviors used as experimental measures of the memory or learned association.

3. *Positive alteration*: Intervening directly to mimic the learning-induced changes of the population of neurons must reliably elicit the behaviors used as experimental measures of the memory, independent of behavioral training.

4. *Integration*: The hypothesis that the learning-induced changes in the population of neurons is a key component of the causal nexus that produces the behaviors used as experimental measures of memory must be connected up with as much experimental data as is available about both the hypothesized neuronal vehicle and memory.

Aside from the appeal to experimental measures of memory, these criteria are ones that researchers use to find neural representations generally. (But see Baker, Lansdell, & Kording (2022) and Pohl et al. (2026) for criteria that more specifically qualify the correlations and alterations that suffice for representation status. Such qualifications can easily be adopted here as nothing hangs on them.)

Another criterion that is implicit in all of the models above is that the vehicle persists. The reason for this criterion is a

common working assumption in cognitive neuroscience: changes in neural representation should correspond with changes in psychological representations (Moscovitch & Gilboa, 2022). Figuring out which changes matter for this correspondence is, of course, the project of cognitive neuroscience. Thus I offer the fifth criterion as tentative pending further investigation:

5. *Persistence*: The learning-induced changes must result in a structure or process that persists between learning and the behavioral measure of memory recall.

While there may be many candidate vehicles for neural representations generally, I have shown above that there appear to be no candidate vehicles for engrams given this fifth criterion. The main contention for interlocutors is a dissatisfaction with how long any candidate property lasts, as, to explain memory, an intuitive suggestion is to look for properties that last as long as the memory is stored. But no candidate property of neuronal ensembles seems to last long enough to explain the persistence of memory. Given 'representational drift', by which tuning properties of neurons change over time in relation to the same environment or stimulus, even the ensembles themselves appear to change while the memory persists.

While the issue of reconciling plasticity and stability in synapses is a classic issue in memory neuroscience (Abraham & Robins, 2005), acceptance of ensemble drift has led to a focus on neural dynamics in neuroscientific research on memory. This focus seems counter-intuitive: after all, it seems that to be able to recall something that was learned in the past, such as the year the declaration of independence was signed or the layout of your childhood home, there should be something that persists from when I learn the fact or develop the model and use it at some later time. In practice, however, researchers do not actually determine that a particular entity persists. (One prominent exception is studies of working memory, see Constantindis & Klingberg, 2016.) Instead, researchers look for a state or activity that can be reactivated that resembles the state or process that occurred when the past event took place. That is, neuroscientific explanations of memory appeal to activities during onset and maintenance of the memory that are similar to those during later recall and take that resemblance as evidence of persistence. The resemblance or similarity relation generally appeals to coarse-grained features of the vehicle type, such as the percentage of neurons active in an ensemble (Tayler et al., 2013; Jung et al., 2023) or, for synaptic structures, by comparing the length and volume of the dendritic spine, diameter of the head, and volume of the neck (Sorra and Harris, 2000). These measures of stability only questionably track properties of engrams that are representational vehicle properties, so we are left unsure that the vehicles persist.

Whether engrams are stable or not is surely an empirical question. But studying the dynamics of memory is made challenging by the fact that engrams are defined by their stability (Guskjolen & Cembrowski, 2023). Indeed, neuroscience's best current methods for studying memory engrams rely on tracking stable features of ensembles over time. Researchers may use cell labeling techniques to show that the activity changes in the same population over learning (e.g., Monasterio et al., 2025 [preprint]) or show that ensemble reorganization underlying memory persistence involves the stability of multiple ensembles (e.g., Kveim et al., 2024). Some have shown that certain properties of spike trains, namely timing, remain stable, while others, namely rate of spikes, exhibit drift (Sotomayor-Gómez, Battaglia, and Vinck, 2025; Zhu et al., 2025). Indeed, stability in various forms is one cue for researchers implying that there is a persisting vehicle—i.e., a memory engram.

Other researchers have argued that memory engrams are much more dynamic. Tome et al. (2024) created a simulation of memory engrams showing that, when accounting for various plasticities that occur with neural activity and learning, the cells that constitute an engram change over time. They argued in further research using cell labeling, optogenetics, chemogenetics, and calcium imaging that changes over time are part of the consolidation process that increases the engram's specificity for a particular stimulus. That is, early in the engram's formation in prefrontal cortex, many stimuli may activate the engram, but after days and weeks, the engram is only activated by the specific memory stimulus. This specificity is explained by changes in which neurons form the engram. Whether the engram is stable after consolidation is not addressed by the researchers, but that representational drift has been observed over a number of weeks indicates that changes continue. In line with this hypothesis, Abbaspoor, Aljishi, & Hoffman (2025 [preprint]) recorded population activity from macaques and found that population activity correlating with specific memory content drifts if it is not integrated and stabilized within populations for older memories with general content. These researchers aim to understand memory consolidation through dynamics, but whether dynamics generally leads to further consolidation, maintenance, or forgetting is underdetermined by the evidence. That memories often survive drastic changes in synaptic structure and ensemble location seems to be only explainable with the existence of some underlying stable structures or persisting processes.

The ubiquitous success of our memory capacity appears to be a miracle without knowing how dynamics give rise to stable representations. Some hypotheses explain away the dynamics by considering them as a bug of the system, a feature that allows for learning, a means to timestamp memory events, or indicative of variations in behavior (Zaki & Cai, 2024). Some have argued that dynamics underlies our forgetfulness, or the loss of details in our memories, which is a feature rather than a bug of memory systems (Eom, Kim, and Ho Hyun, 2025). The most accepted explanation is that dynamics and drift are just processes of consolidation, by which memories are prepared for long-term storage via increased synaptic coupling and representation in prefrontal cortex. Neural dynamics underlying consolidation further explain how memory degrades and changes over time. But the consolidation picture just pushes the question back. Once the memory is consolidated in prefrontal cortex, is there a

persisting vehicle of that memory? It seems that, as shown in the section on vehicle indeterminacy, every vehicle candidate is subject to instability via molecular degradation, synaptic turnover, neural dynamics, and drift. This means that, even if the consolidation picture is right, dynamic vehicles are underlying memory representations. In the next section, I take seriously the idea that nothing persists in the brain to underly memory representation.

Revisioning Engrams

The primary issue researchers face in resolving memory engram vehicle indeterminacy is this: in studying engrams, researchers want to account for dynamics but find stability. Robins (2020) has argued that neural dynamics is not inconsistent with stable memory. She calls for a revision of how we think memory representations are implemented or realized by neural entities and activities to accommodate neural dynamics in explanations of memory. On her view, the mistakes are in thinking that something that persists cannot change and in thinking that there is some isomorphic mapping between neural entities or activities (i.e., vehicles) and representational content. (Which changes are relevant is a matter of empirical inquiry.) Researchers suppose this simple vehicle-content relation when theorizing that changes in the vehicle explain changes in memory content, as in cases of memory errors or the process of consolidating memories (e.g., Mau et al., 2020; Takamiya et al., 2020). But we should be skeptical that vehicle-content relations are so simple. I take Robins' account as correct and draw out implications for determining representational vehicles of neural representations like memory engrams.

Recall that researchers tend to infer persistence of a vehicle candidate if it is present during each instance of engram detection (and, one might add, missing when the engram fails to elicit memory retrieval or is itself missing). The revisioning of engrams that I offer is a rational reconstruction of neuroscientific practices targeting engrams and makes explicit the engram concept that researchers are committed to when studying neural dynamics. I argue that memory storage is not achieved through persisting vehicles but via a diachronic capacity of the system to recall a memory. In other words, it is functional continuity not structural persistence that necessarily underlies memory recall. The diachronic capacity is a historically grounded disposition, that is, a disposition of the system to produce memory recall that is grounded in the causal history of the system rather than a persisting, stable feature. To explain memory storage, one needs only to explain how a particular capacity persists through various states of the system. This can be done by finding categorical bases for dispositions in persisting properties or in dynamical ones. Engrams, on this approach, are causal motifs (Mace, O'Sullivan, & Wilson, 2026). The idea is that engrams are a recurring cause of a cognitive or behavioral response. As causal motifs, memory engrams elicit memory responses and are individuated by their content. This means that memory recall is the measure of stability and that engrams persist only at the level of content. This is fine for what we need engrams to do, which is explain our capacity to remember past experiences and learned facts or skills. The link to the past need not be a physical link that remains the same for the duration of memory storage.

There are a number of ways that this view can be cashed out. De Brigard (2014) treats neural networks in the brain as having a dispositional property to recreate the representational vehicle. De Brigard seems to have in mind pattern completion properties of neurons in neuronal ensembles that can reactivate a particular pattern of activity. Whether such patterns are engrams remains to be seen, but I take it that focusing on what causes retrieval in each instance of memory recall can help to determine what the vehicle is in each case and whether there is a vehicle that is instantiated across those cases. It may still be worthwhile, that is, to look for something that is functionally stable across physical changes to the system (De Brigard, 2017). But we must also be open to the fact that there is no such structure or activity that persists or is instantiated across instances of memory retrieval. Some systems in the brain lack topographical organization yet retain the capacity for recognition or familiarity (see, for example, Barwich and Severino, 2023). Treating the engram as a causal motif allows it to be an empirical question whether the memory system is dynamical in this way. A third option is to appeal more explicitly to environmental or psychological factors in eliciting recall in the system, such as background information, context, and cues that aid in recall (Robins, 2012). A vehicle candidate is a worthy contender if it has a causal role in eliciting behavior indicative of memory use, but it need not have this causal role exclusively.

Rather than get bogged down in determinations of whether, in principle, the particular entities can persist for the length of the memory, debates about memory engrams should focus on causal efficacy of particular vehicle candidates and whether that particular causal role is best described as bearing content about something learned in the past or as some alternative (e.g., providing instructions for creating engrams or providing an index to cortically stored engrams). An implication of this view is that persisting memory content is compatible with behavioral variation and, importantly, with neural dynamics. Now, memory content famously contains errors and can change over time. Explaining how this happens is certainly a goal in neuroscience. But researchers are also interested in explaining successful, veridical remembering, which we seem to do quite often with memories that persist for an extended period of time. Given neural dynamics, explaining this capacity appears to be the more difficult task. My framework invites a conception of engrams as a causal motif whereby the explanatory potential of engrams comes from their causal profile, not from the postulation of stable neural entities that persist during the course of storage and remembering. What will be explained, in the end, is the capacity to remember.

Acknowledgments

Discussions of memory engrams with Scott Wilson and Fionn O'Sullivan have greatly inspired my understanding of engrams as causal motifs. I am grateful to them for those discussions as well as to Jordan Dopkins, Sarah Robins, and Edouard Machery for helpful discussion and comments on earlier drafts of this paper. I thank two anonymous reviewers who provided feedback that greatly improved this paper. Finally, I am grateful for ChatGPT, which was a useful sounding board (and at times echo chamber) during the development of the ideas presented here; all arguments and formulations are my own.

References

Abbaspoor, S., Aljishi, A., & Hoffman, K.L. (2025). Experience reorganizes content-specific memory traces in macaques. *bioRxiv* [Preprint] doi: 10.1101/2025.04.08.647787. PMID: 40832227; PMCID: PMC12363849.

Abraham, W. C., & Robins, A. (2005). Memory retention – the synaptic stability versus plasticity dilemma. *Trends in Neurosciences, 28*(2), 73–78. https://doi.org/10.1016/j.tins.2004.12.003

Babich, F., Jacobson, A., & Bubash, S. (1965). Cross-species transfer of learning: effect of ribonucleic acid from hamsters on rat behavior. *Proceedings of the National Academy of Sciences of the United States of America, 54*, 1299.

Babich, F., Jacobson, A., Bubash, S., & Jacobson, A. (1965). Transfer of a response to naïve rats by injection of ribonucleic acid extracted from trained rats. *Science, 149*, 656–657.

Baker, B., Lange, R., Richmond, A., Kriegeskorte, N., Cao, R., Pitkow, X., & Schwartz, O. [forthcoming]. Use and usability: concepts of representation in philosophy, neuroscience, cognitive science, and computer science. *Neurons, Brains, Data Analysis, and Theory.*

Barwich, A. S., & Severino, G.J. (2026). The Wire Is Not the Territory: Understanding Representational Drift in Olfaction With Dynamical Systems Theory. *Topics in Cognitive Science* https://doi.org/10.1111/tops.12689

Bergstrom, H., McDonald, C., & Johnson, L. (2011). Pavlovian Fear Conditioning Activates a Common Pattern of Neurons in the Lateral Amygdala of Individual Brains. *PLoS ONE, 6*(1), e15698. https://doi.org/10.1371/journal.pone.0015698

Bickle, J. (2008). Real Reduction in Real Neuroscience: Metascience, Not Philosophy of Science (and Certainly Not Metaphysics!). In J. Hohwy (Ed.), *Being Reduced: New Essays on Reduction, Explanation, and Causation* (pp. 34–51). Oxford University Press.

Bliss, T., & Lømo, T. (1973). Long-lasting Potentiation of Synaptic Transmission in the Dentate Area of the Anaesthetized Rabbit Following Stimulation of the Perforant Path. *Journal of Physiology, 232*(2), 331–356.

Colaço, D., & Najenson, J. (2024). Where memory resides: Is there a rivalry between molecular and synaptic models of memory? *Philosophy of Science, 91*, 1382–1392.

Constantinidis, C., & Klingberg, T. (2016). The neuroscience of working memory capacity and training. *Nature Reviews Neuroscience, 17*, 438–449. https://doi.org/10.1038/nrn.2016.43

Craver, C. (2007). *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford University Press.

De Brigard, F. (2014). The Nature of Memory Traces. *Philosophy Compass, 9*(6), 402–414.

De Brigard, F. (2017). Cognitive systems and the changing brain. *Philosophical Explorations, 20*(2), 224–241. https://doi.org/10.1080/13869795.2017.1312503

Drayson, Z. (2018). The realizers and vehicles of mental representation. Studies in History and Philosophy of Science, 68, 80–87.Dudai, Y. (2002). *Memory from A to Z: Keywords, Concepts, and Beyond*. Oxford University Press.

Ebitz, R. B., & Hayden, B. (2021). The population doctrine in cognitive neuroscience. *Neuron 109*, 3055–3068.

Gallistel, C. R. (2017). The Coding Question. *Trends in Cognitive Sciences, 21*(7), 498–508.

Gallistel, C. R. (2021). The Physical Basis of Memory. *Cognition, 213*, 104533.

Gershman, S. (2023). The molecular memory code and synaptic plasticity: A synthesis. *Biosystems, 224*, 104825.

Gold, A. R., & Glanzman, D. L. (2021). The central importance of nuclear mechanisms in the storage of memory. *Biochemical and biophysical research communications, 564*, 103–113. https://doi.org/10.1016/j.bbrc.2021.04.125

Goode, T. D., Tanaka, K. Z., Sahay, A., McHugh, T. J. (2020). An Integrated Index: Engrams, Place Cells, and Hippocampal Memory. *Neuron, 107*(5), 805–820. doi: 10.1016/j.neuron.2020.07.011.

Guskjolen, A., & Cembrowski, M. S. (2023). Engram neurons: Encoding, consolidation, retrieval, and forgetting of memory. *Molecular Psychiatry, 28*(8), 3207–3219. https://doi.org/10.1038/s41380-023-02137-5

Hebb, D. (1949). *The Organization of Behavior: A Neuropsychological Theory*. L. Erlbaum Associates.

Humphries, M. (2025). The challenge of defining a neural population. *The Transmitter* https://doi.org/10.53053/FZKM5346

Jung, J. H., Wang, Y., Mocle, A. J., Zhang, T., Köhler, S., Frankland, P. W., Josselyn, S. A. (2023). Examining the engram encoding specificity hypothesis in mice. *Neuron, 111*, 1830–1845.

Kveim, V. A., Salm, L., Ulmer, T., Lahr, M., Kandler, S., Imhof, F., & Donato, F. (2024). Divergent recruitment of developmentally defined neuronal ensembles supports memory dynamics. *Science (American Association for the Advancement of Science), 385*(6710), eadk0997. https://doi.org/10.1126/science.adk0997

Langille, J., & Brown, R. (2018). The Synaptic Theory of Memory: A Historical Survey and Reconciliation of

Recent Opposition. *Frontiers in Systems Neuroscience, 12*(52), 1–15.

Lara Aparicio, S.Y., Laureani Fierro, Á.d.J., Aranda Abreu, G.E., Toledo Cárdenas, R., García Hernández, L.I., Coria Ávila, G.A., Rojas Durán, F., Aguilar, M.E.H., Manzo Denes, J., Chi-Castañeda, L.D.; et al. (2022). Current Opinion on the Use of c-Fos in Neuroscience. *NeuroSci 3*, 687–702. https://doi.org/10.3390/neurosci3040050

Lashley, K. (1950). In Search of the Engram. *Physiological Mechanisms in Animal Behavior, 4*(1), 454–482.

Liu, X., Ramirez, S., Pang, P., Puryear, C., Govindarajan, A., Deisseroth, K., & Tonegawa, S. (2012). Optogenetic Stimulation of a Hippocampal Engram Activates Fear Memory Recall. *Nature, 484*(7394), 381–385.

McConnell, J. (1962). Memory transfer through cannibalism in planarians. *Journal of Neuropsychiatiatry, 3*, 542–548.

McDougall, W. (1927). An Experiment for the Testing of the Hypothesis of Lamarck. *British Journal of Psychology, 17*(4), 267–304.

McDougall, W. (1930). Second Report on a Lamarckian Experiment. *British Journal of Psychology, 20*(3), 201–218.

McDougall, W. (1938). Fourth Report on a Lamarckian Experiment. *British Journal of Psychology, 28*(4), 321–345.

Mace, C. (2021). Casting Light on the Search for Engrams: On the Reductionism-Mechanism Debate. *ProQuest Dissertations & Theses, 28644486*, 1–149.

Mace, C., O'Sullivan, F., & Wilson, S. (2026). Reductionism in Engram Neuroscience. *European Journal of Neuroscience, 63*(8), 1–18.

Martin, S., Grimwood, p., & Morris, R. (2000). Synaptic Plasticity and Memory: An Evaluation of the Hypothesis. Annual Reviews *Neuroscience, 23*(1), 649–711.

Mau, W., Hasselmo, M. E., & Cai, D. J. (2020). The brain in motion: How ensemble fluidity drives memory-updating and flexibility. *eLife, 9*, e63550. https://doi.org/10.7554/eLife.63550

Mayford, M. (2014). The Search for a Hippocampal Engram. *Philosophical Transactions of the Royal Society B, 369*(1), 1–10.

Monasterio, A., Lienkaemper, C., Coello, S., Ocker, G. K., Ramirez, S., & Scott, B. B. (2024). CA1 Engram Cell Dynamics Before and After Learning. *bioRxiv* doi: https://doi.org/10.1101/2024.04.16.589790

Moore, R., Kaletsky, R., Lesnik, C., Cota, V., Blackman, E., Parsons, L., Gitai, Z., & Murphy, C. (2021). The role of the Cer1 transposon in horizontal transfer of transgenerational memory. *Cell, 184*, 4697–4712.

Moscovitch, M., & Gilboa, A. (2022). Has the concept of systems consolidation outlived its usefulness? Identification and evaluation of premises underlying systems consolidation. *Faculty Reviews, 11*(33). doi: 10.12703/r/11-33. PMID: 36532709; PMCID: PMC9720899.

Pohl, S., Walker, E., Barack, D., Lee, J., Denison, R., Block, N., Meyniel, F., & Ji Ma, W. (2026). Clarifying the conceptual dimensions of representation in neuroscience. *Nature Reviews Neuroscience* https://doi.org/10.1038/s41583-026-01030-8

Ramirez, S., Tonegawa, S., & Liu, X. (2014). Identification and Optogenetic Manipulation of Memory Engrams in the Hippocampus. *Frontiers in Behavioral Neuroscience, 7*(226), 1–9.

Rhine, J., & McDougall, W. (1933). Third Report on a Lamarckian Experiment. *British Journal of Psychology, 24*(2), 213–235.

Robins, S. K. (2012). The Constructive Trace Theory of Memory. *Arts & Sciences Theses and Dissertations, 282*.

Robins, S. K. (2020). Stable Engrams and Neural Dynamics. *Philosophy of Science, 87*(5), 1130–1139. doi:10.1086/710624

Robins, S. K. (2023). The 21st Century Engram. *WIREs Cognitive Science, 14*(5), 1–15.

Schoonover, C., Ohashi, S., Axel, R., & Fink, A. (2021). Representational drift in primary olfactory cortex. *Nature, 594*, 541–574.

Semon, R. (1921). *The Mneme*. (L. Simon, Trans.) Allen & Unwin.

Silva, A., Landreth, A., Bickle, J. (2014). *Engineering the Next Revolution in Neuroscience: The New Science of Experiment Planning*. Oxford University Press.

Sorra, K., & Harris, K. (2000). Overview of the Structure, Composition, Function, Development, and Plasticity of Hippocampal Dendritic Spines. *Hippocampus, 10*, 501–511.

Sotomayor-Gómez, Boris, Francesco P. Battaglia, and Martin Vinck. (2025). Firing rates in visual cortex show representational drift, while temporal spike sequences remain stable. *Cell Reports, 44*(115547), 1–33.

Sullivan, J. (2008). Memory Consolidation, Multiple Realizations, and Modest Reductions. *Philosophy of Science, 75*, 501–513.

Takamiya, S., Yuki, S., Hirokawa, J., Manabe, H., & Sakurai, Y. (2020). Dynamics of memory engrams. *Neuroscience Research, 153*, 22–26. https://doi.org/10.1016/j.neures.2019.03.005

Tayler, K., Tanaka, K. Z., Reijmers, L. G., & Wiltgen, B. J. 2013. Reactivation of Neural Ensembles during the Retrieval of Recent and Remote Memory. *Current Biology, 23*(2), 99–106.

Teyler, T. J., & DiScenna, P. (1986). The hippocampal memory indexing theory. *Behavioral neuroscience, 100*(2), 147–154. https://doi.org/10.1037//0735-7044.100.2.147

Tomé, D.F., Zhang, Y., Aida, T. et al. Dynamic and selective engrams emerge with memory consolidation. *Nature Neuroscience 27*, 561–572 (2024). https://doi.org/10.1038/s41593-023-01551-w

Trettenbrein, P. C. (2016). The demise of the synapse as the locus of memory: a looming paradigm shift? *Frontiers in Systems Neuroscience, 10*(88), 1–7. doi: 10.3389/fnsys.2016.00088

Werning, M. (2020). Predicting the Past from Minimal Traces: Episodic Memory and its Distinction from Imagination and Preservation. *Review of Philosophy and Psychology, 11*, 301–333.

Yuste R., Cossart, R., & Yaksi, E. Neuronal ensembles: Building blocks of neural circuits. *Neuron, 112*(6), 875-892. doi: 10.1016/j.neuron.2023.12.008. Epub 2024 Jan 22. PMID: 38262413; PMCID: PMC10957317.

Zaki, Y., & Cai, D. J. (2024). Memory engram stability and flexibility. *Neuropsychopharmacology, 50*(1), 285–293. https://doi.org/10.1038/s41386-024-01979-z

Zhu, Hanlin, Fei He, Pavlo Zolotavin, Saumil Patel, Andreas S. Tolias, Lan Luan, and Chong Xie. (2025). Temporal coding carries more stable cortical visual representations than firing rate over time. *Nature Communications, 16*(7162), 1–20.