If engrams are the answer, what is the question?

Fionn M. O'Sullivan[1-2] & Tomás J. Ryan[1-4]

Affiliations

[1] School of Biochemistry and Immunology, Trinity College Dublin, Dublin, Ireland.

[2] Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland.

[3] Florey Institute of Neuroscience and Mental Health, Melbourne Brain Centre, University of Melbourne, Melbourne, Victoria, Australia.

[4] Child & Brain Development Program, Canadian Institute for Advanced Research (CIFAR), Toronto, Ontario, Canada.

*Corresponding author. Email: tomas.ryan@tcd.ie

This is a penultimate draft, please cite the final version.

Abstract

Engram labelling and manipulation methodologies are now a staple of contemporary neuroscientific practice, giving the impression that the physical basis of engrams has been discovered. Despite enormous progress, engrams have not been clearly identified, and it is unclear what they should look like. There is an epistemic bias in engram neuroscience towards characterising biological changes, while neglecting the development of theory. However, the tools of engram biology are exciting precisely because they are not just an incremental step forward in understanding the mechanisms of plasticity and learning, but because they can be leveraged to inform theory on one of the fundamental mysteries in neuroscience—how and in what format the brain stores information. We do not propose such a theory here, as we first require an appreciation for what is lacking. We outline a selection of issues in four sections from theoretical biology and philosophy that engram biology and systems neuroscience generally should engage with in order to construct useful future theoretical frameworks. Specifically, what is it that engrams are supposed to explain? How do the different building blocks of the brain-wide engram come together? What exactly are these component parts? And what information do they carry, if they carry anything at all? Asking these questions is not purely the privilege of philosophy, but a key to informing scientific hypotheses that make the most of the experimental tools at our disposal. The risk for not engaging with these issues is high. Without a theory of what engrams are, what they do, and the wider computational processes they fit into, we may never know when they have been found.

This is a preprint of the following chapter: O'Sullivan & Ryan, "If engrams are the answer, what is the question?" published in Engram Biology, edited by S. Ramirez & J. Gräff, 2023, Springer Nature, reproduced with permission of Springer Nature.

Introduction

An outside observer might be forgiven for thinking that hidden in memory engram biology papers, or implicitly known among the systems neuroscience community, there is a clear understanding of what engrams are and how they relate to memory. Memory is defined at a behavioural level as changes resulting from past experience, and may also be defined as reconstructed internal representations (Dudai, 2007). In contrast, engram biology as a research program is dedicated to defining the biological basis of memory. Engrams are not memories, instead, they are what provide "the necessary (physical) conditions for a memory to emerge" (Moscovitch, 2007). Although not always, within neuroscience the terms memory trace and engram are synonymous, with the term engram introduced in Semon's biological theory of memory (1904/21; 1909/1923). In contemporary neuroscience, engrams are typically operationalised as biological constructs of specific cell ensembles. These cell ensembles can be experimentally labelled through co-opting activity-dependent immediate-early genes (IEGs) to express fluorescent and other proteins in those neurons activated during a particular learning experience. Because reactivating these specific labelled cell ensembles (but not others) often reproduces the original learned behaviour, and their silencing or ablation fails to reproduce that exact same behaviour, scientists have claimed these cells fulfil the function of the memory trace or engram and have appropriated and operationalised Semon's theory to help make sense of these experimental findings (Josslyn, Köhler & Franklin, 2015; Tonegawa *et al*., 2015a, Tonegawa *et al*., 2015b).

However, for all the pragmatic contemporary usage of the word, the engram remains an open-ended scientific construct. Semon's definition, restated in Tonegawa *et al*. (2015a) as the "*enduring physical and/or chemical changes that were elicited by learning and underlie the newly formed memory associations*" is not tied to any specific biological level. It is telling that we must reach back over one hundred years for a conceptual framework to operationalise for current use in neuroscience[1]. In other words, although Semon's framework has been validated, it has not been elaborated, and there are no fleshed out alternative biological theories that would create healthy competition for the field. In general, the memory engram field remains under-theorised (Robins, 2023). Is this something that can be remedied by more advanced experimentation alone, or is *new theory* also needed?

For some, the concept of an engram may serve as a working abstract placeholder because the engrams that the scientific construct refers to have not been truly found, and the abstraction affords the elbow room required to manoeuvre while figuring out how engrams are realised in biological tissue. Despite experimental limitations regarding the temporal resolution of the cell-tagging protocol, differences in endogenous IEG function, controlling for off-target effects, and detecting exact differences between engram and non-engram cells, referring to the existence of "engram cells" is a pragmatic step forward based on progressive behavioural evidence that cannot be dismissed. Examining the cells involved in learning and recall did not *a priori* suggest that we would find cells labelled at encoding being reactivated at recall, or that these overlapping cells would be functional at the level of behaviour. Rather, this was demonstrated by empirical investigation (Reijmers *et al*., 2007, Liu *et al*., 2012). Meanwhile, the understanding of "how information is stored in an engram" is acknowledged as the major outstanding overarching question in the engram field (Queenan *et al*., 2017; Josselyn & Tonegawa, 2020) and cognitive science generally (Poeppel & Idsari, 2022).

For others, not only have engrams not been satisfactorily found, they may never be because the concept is not well defined (Sossin & Hardt, 2020) and there are views of memory that do not place prime importance on the existence of engrams (Schacter & Addis, 2007). If their existence can neither be proven or disproven, their scientific utility is questionable. Additionally, it is unclear from either theory or empirical work what the information content is and how it is supposed to persist in

[1] This is not to disregard the original theory's insightfulness, proper historical accreditation, or the very real sociological, ideological and experimental limitations that might have previously discouraged further theorising about the nature of the engram during the 20th century (Schacter, 2001).

the form or format of an engram. The language used in discussing existing engram theory can hide this lack of knowledge. To say that physical and/or chemical changes "underlie" memory is either a truism or insinuates that we have a clear theory for *what* the enduring changes are and *how* they do this underlying[2]. We have neither. Performing observational, gain-of-function, loss-of-function, and mimicry experiments has established *that* there is a replicable and specific linking between these unique cell ensembles and behaviour. However, when the theory being tested is ambiguous, such studies are limited in their power to explain what engrams are or *how* this specific linking works (Krakauer *et al*., 2017; Tonegawa *et al*., 2015b). There are now several putative levels for where the persistent changes that function as engrams may be found. These include i) synaptic weight plasticity (Martin, Grimwood & Morris, 2000; Kandel, 2001), ii) intracellular and molecular mechanisms (e.g. protein or polynucleotide modifications) (Sacktor, 2011; Bédécarrats *et al*., 2018; Campbell & Wood, 2019; Gershman 2023), or iii) that the relevant mechanism is supra-cellular, such as changes in the structural connectivity of the connectome (Chklovskii, Mel & Svoboda, 2004; Ryan *et al*., 2021) or non-neuronal structures, like perineuronal nets (Tsien, 2013), astrocytes and oligodendrocytes (Kol & Goshen, 2020), or microglia (Wang *et al*., 2020). However, none of these proposals explain the *how*—how such changes are the appropriate mechanism capable of supporting recall of a given behaviour over any other[3]. This discrepancy—between the establishment of a linking between labelled neuronal ensembles (or any other level) and behaviour, but a dearth of understanding how this works—is a gauntlet that needs to be picked up by a new theoretical framework. However, our objective here is not to outline a concrete "engram theory". Developing new theories first requires understanding what our current framework is missing, including whole types of theory. This chapter therefore discusses current ambiguities in order to identify those areas any future theories must pay attention to.

Section 1 addresses how there are in fact multiple things-to-be-explained when we are concerned with the biological basis of memory, and apparently competing explanations may not actually be in conflict. Section 2 singles out the dominant theory of engrams as operationalised in contemporary neuroscience, focusing on the limitations raised by its assumed modular architecture. Attention is paid to questioning the assumption that differences at the level of cell ensembles actually correspond to differences in the "type of information stored" by said ensembles. Section 3 argues this conflation occurs due to failure to dissociate questions about *information content* from questions about *vehicles of content*, and emphasises the need for more careful terminology regarding differing conceptions of what the engram should be. Section 4 is devoted to considering the limitations of theories of information in neuroscience and biology. For us, one of the most exciting aspects of engram technology is its potential to inform a new conception of information in the brain. Finally, in the Discussion section we summarise the implications of this approach for helping to guide experimental research.

[2] The use of such "filler-verbs" is a kind of interpretative (mal)practice not unique to engram biology but endemic to all disciplines of contemporary neuroscience (Krakauer *et al*., 2017).

[3] Our explanation may involve integration across many of these levels, but a multi-level explanation must still account for the *how*.

Section 1 — Levels, Questions, & Explanations

1.1 — If engram is the answer, what was the question?

The major theoretical hurdle faced by engram biology is a lack of clarity regarding what the problems to be solved are, or not spelling them out in sufficient detail. The engram is assumed to be a good explanation, but what exactly are engrams posited to explain? The existence of something capable of functioning as an engram is usually invoked in order to explain the general phenomena of learning and memory. However, "learning and memory" is not a coherent, singular phenomenon, a*nd when the phenomenon we want to explain is too general, multiple competing explanations will appear equally adequate*. There are different questions that need to be parsed. In this section, three different methods for identifying and parsing questions are presented based on i) biological organisation, ii) question-type, iii) and question-subdivision.

1.2 — Dividing explanations based on levels of biological organisation

Many biological processes and computations involve changes across multiple organismal levels. The changes studied in learning and memory research cover several orders of magnitude of various nested biological processes (Josselyn, Köhler, & Frankland, 2015; Craver, 2007, ch.5), from molecular to systems physiology. However, we are not sure which levels of organisation are most relevant for theorising about how the brain can be said to store information. It may be useful to draw analogies with similar problems in other areas of neuroscience. In their discussion comparing the relevance of neuron-and-circuit-level versus population-manifold-level descriptions when it comes to explaining cognitive phenomena, Barack & Krakauer (2021) illustrate the importance of choosing the best level of explanation with analogy to different answers to the question "why did the window break?" The *first-level explainer* (we denote here as E_1) should be "because the ball was thrown at it" and not explanations based on the physiology of arm-throwing or the Si-O amorphous molecular structure of glass. These *second and third level explainers* (E_2 and E_3) are defined by their ability to explain E_1, but that does not make E_2 and E_3 better answers to the original question[4].

If we are seeking to explain the biological basis of memory, we are not seeking to explain "the engram". Engram may be posited as a first-level explanation (E_1) *for* the biological basis of memory, or may be a given step (E_n) along the way, say E_5. Alternatively, the engram may refer to a given chain of processes, with no one privileged level. Since we do not know what the engram is, our focus shifts to picking out the next level of explanation down (E_{n-1}) to that which might support the engram. Thus, when a biological mechanism or process is conjectured as an explanation, care should be taken to ascertain whether this is an explanation for the biological basis of memory, or actually the-thing-that-explains-the-thing (etc.) that explains the biological basis of memory[5].

Although this seems like a straightforward pitfall to avoid, it is not when it comes to memory. This is because the problem to which memory is a solution to, that of dealing with dynamically changing external conditions that have some statistical dependencies, is not a problem that manifests once at a single scale[6]. It is evolutionarily universal, and solutions or adaptations have manifested across multiple levels of biological organisation and phylogeny (unlike, say, speech production or motor imagery). Bacteria have been said to store information in protein post-translational modifications (Sterling & Laughlin, 2015; Gerhsman, 2023) and slime moulds are capable of anticipating temperature changes (Saigusa *et al.,* 2008). Should any biological mechanism of "storing

---

[4] If the question is not fixed but we vary it, then what is E_2 to one question may become the E_1 for another.

[5] The pitfall to avoid is to take, say, E_{20} for E_2.

[6] This question-answer dialectic of an environment posing problems to which an organism evolves "solutions" is imperfect (see Lewontin, 1983) but illustrative for our purposes here.

information" by modifying some structure be considered an engram? Avoiding a situation of "endless engrams" all the way down, in which vastly different processes all stake claim to the same name, implying some non-trivial equivalence across all, is preferable to us[7,8]. Granted, different biological processes across the evolutionary continuum and in parallel across multiple scales within multicellular organisms involve the creation of specific changes could be said to function as engrams that are relevant for various kinds of phenotypic plasticity[9]. However, the question to be addressed in order to explain behavioural memory is *which aspect of the problem is each mechanism or process a solution to* (see Figure 1) and how these different information storage mechanisms are hierarchically scaffolded together (Simon, 2002; Hoffmeyer, 2007). Experimental data without theory does not reveal this[10].

Given these general considerations, competing proposals for the location of "the engram" suddenly no longer seem to be in competition at all. In recognising that synapses (or synaptic strength at least) may not be the site of the lasting changes that count as an engram (Chen *et al*., 2014; Ryan *et al*., 2015), two theories, one that goes down a level of biological organisation from synapses to nucleotides and protein conformational states (Gallistel, 2017; Gershman, 2023), and one that goes up a level to structural connectivity (Chklovskii, Mel & Svoboda, 2004; Tonegawa *et al*., 2015b; Ryan *et al*., 2021), have seen a resurgence or been proposed. However, the spatial scales of causal influence implicated at each level may differ dramatically. It may be completely accurate to claim that there is information *in* neurons, that it is molecular and that stable distributed molecular states could function as engrams, but these may be engrams *for* the cell informing itself about, say, cytoskeletal rearrangements, and not engrams directly *for* whole-organism behaviour that is retrievable over millisecond timescales[11]. The former is unlikely to be an E_1 for the biological basis of how an organism such as a mammal is capable of creating an episodic memory. It may feature in a future theory of the biological basis of episodic memory at E_{10}, say, and may indeed be necessary, but E_{10} alone is insufficient.

In contrast, it is also possible that the engram will not be the E_1 for the biological basis of episodic memory but will require it being situated in a larger context. For example, a "manifold view of the engram," whereby making a particular previously latent connectivity pattern suddenly relevant when through inhibition, which hypothetically forces neural activity to take on a different dynamical trajectory in a certain context (Langdon, Genkin & Engel, 2023), may be a more relevant level of explanation, or chain of explanations, for the form of engram posited to explain episodic memory. It may in turn be explained by cytoskeletal rearrangements and other molecular level changes, just as motor neuron firing can at a certain level be said to implement throwing a ball at a window. Spelling the problem(s) out at different gradations of complexity will facilitate selecting the appropriate type of engram or information storage mechanism posited in a given hierarchy. Again, our goal here is not to provide a speculative theory of what these different levels or engrams are, but to demonstrate the importance of levels-of-explanation differences in formulating theory in neuroscience.

---

[7] There may be nothing that unites them other than uncertainty reduction very generally defined.

[8] Biological processes are nested and previous solutions are repurposed, but this does not make a solution at one level universal or that it will be used for addressing the same problem in future evolution (Gould & Vrba 1982; Kauffman, 2019). Solutions to general evolutionary problems can be convergent in some cases, but they can also be divergent. When the same problem is faced at different levels of organisation, even within the organism, divergent solutions are found. Both we and our immune cells have to "navigate." We use a musculoskeletal system and the algorithm of walking. Immune cells use cytoskeletal rearrangements and roll along tissue with surface proteins. Like the general problem of navigation, it is sensible to assume the similarly general problem of learning and the solution of storing changes also has many levels.

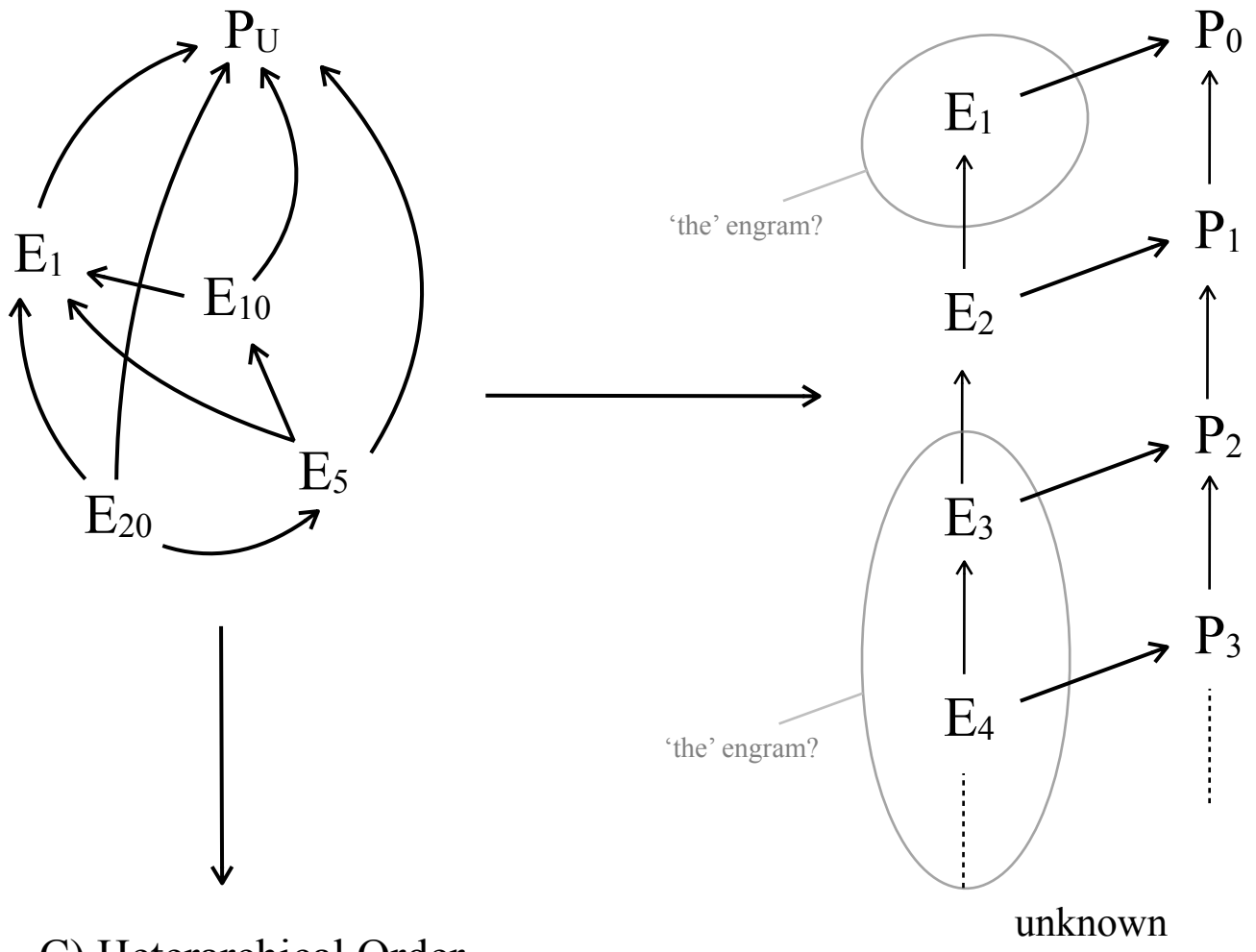
[9] For example, DNA, immune memory, and external memory (Donald, 1991; Clark & Chalmers, 1998)

[10] Although we might claim that the level of learning and memory we study in systems neuroscience is tethered to whole-organism behavioural parameters, and all experimental data is collapsed relative to this, the changes we choose to look at, stain, sequence, or quantify are not determined by the data but by hypotheses of where to look.

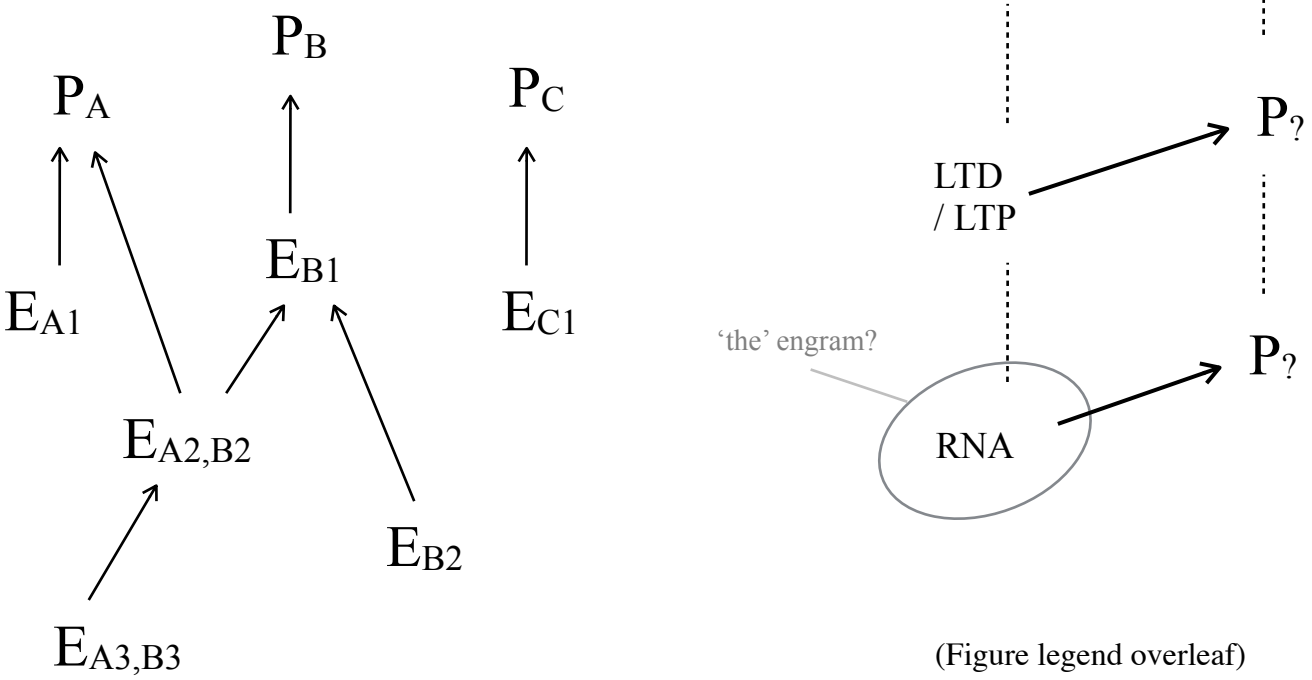
[11] Stable molecular states may be E_{n+2} explanations for cytoskeletal rearrangements, E_{n+1}, which are explanations for E_n, but this does not make E_{n+2} an explanation for E_n.

A) Current Mess

B) Hierarchical Order



C) Heterarchical Order



(Figure legend overleaf)

Figure 1: A) When the problem to which engrams are posited to solve is articulated over-generally as a kind of universal problem, P_U, multiple biological mechanisms may seem to qualify as equally adequate explanations. Either many biological mechanisms get to be engrams or their conglomeration is an engram, neither of which is illuminative. In moving from A) to B) or C) by unfolding P_U into a series of more specific problems that have to be solved, the biological mechanisms can be parsed and scaffolded into a more coherent theory, whether hierarchical or heterarchical. We may choose to define the engram as a particular causal motif (Ross, 2021) or as that mechanism which offers the greatest control in terms of causal manipulation of memory content, or as a chain of biological mechanisms, none of which alone qualify as an engram. Specifying problems discretely rarely works in practice, and here only illustrates a conceptual point.

1.3 — Dividing explanations based on question-type

So far we have discussed choosing explanations with respect to a fixed question type (and by varying the level at which the question manifests), but we can also vary the type of question we ask. For example, classic heuristics like Tinbergen's (1963) "Four Questions" differentiates *how-questions* from *why-questions* for any given behaviour[12]. We used a why-question in asking why the window broke, but if we ask "how did the window break?" then explanations for the brittleness of glass and velocity of the ball are indeed the relevant form of explanation. It could be argued that when it comes to the memory relevant for behaviour, the *what* (the behaviour we are studying) is learning about the social and physical environment, the *why* is recall in the future for adaptive behaviour, and the *how* is the engram and its expression. However, there is no singular answer to the how-question, because *how to form* an engram versus *how to use* an engram we presume are different processes. Formation and reactivation may therefore require different levels of description. For example, an explanation for *how an engram is used* may require situating structural changes in terms of their downstream effects (e.g. funnelling network activity into a subspace regardless of where the network starts at the time of retrieval, but which preserves some high-level invariant properties with the landscape at the time of encoding), before jumping to explaining behaviour.

1.4 — Dividing explanations based on subdividing the problem

Section 1.2 introduced the problem of potentially many different processes being collapsed into an unknown hierarchical relevance with respect to a particular observed behaviour. Section 1.3 demonstrated different kinds of questions. Drawing on both concerns, this section focuses on taking the general problem of "dealing with dynamically changing external conditions…" and decomposing it into sub-problems. For example, we might focus on the problem of efficient information storage (Sterling & Laughlin, 2015), or the problem of making stable changes accessible (Ryan & Frankland, 2022), or the problem of whether a given type of engram is computationally content-addressable or address-addressable (Gallistel, 2017), to speculate on a few. The most relevant level description may differ for each, and there may be multiple versions of each problem across scales. Deciding how to divide a problem like biological information storage into subproblems is also not something that can be approached solely as an engineering problem, but requires consideration of the phylogenetic history of the organism and its ancestral environments (Cisek, 2019).

Failure to decompose the problem hinders inferring which changes in what process are most relevant. If all we are looking for is structured variation in a particular behavioural parameter, there is an endless array of variables across the scales of biological organisation that we can intervene with

---

[12] The proximate or how questions are the development and its mechanism of generation. The ultimate or why questions are the evolutionary history and what specific adaptation it affords.

that will produce such effects. This risk, as raised in Jonas & Kording's (2017) problem of inferring the logic of a microprocessor using standard neuroscientific gain-of-function and loss-of-function causal interventional techniques, is that we will take the wrong level of explanation to be E_1 just because manipulation at that level resulted in structured outcomes in behavioural parameters. Structured variation in behavioural outcomes due to protein synthesis inhibition (e.g. no memory consolidation, see Ryan *et al*., 2015) does necessarily not mean that proteins are going to feature in how our explanation for how an engram instantiates a particular form of information.

1.5 — What would a future engram theory look like? What does it require?

Returning to the question of what features a comprehensive engram theory would have following these approaches, a comprehensive theory would be pluralistic. A classification system or *typology* of the different mechanisms or processes involved different types and levels of engrams could be constructed. Even if the term engram becomes reserved exclusively for the "high-level" traces of declarative memories, the theory must be able to differentiate other forms of information-storage mechanisms and explain how these traces differ. Although classification systems will be an important scaffold in any engram theory, a list of types, mechanisms and processes in isolation is insufficient. Firm principles explaining why these categories are carved the way they are (rather than any other way) are needed to accompany any classification system. This approach would bring together several different avenues of research and wrestle effort away from debating what "the engram" is towards the unique ways competing classification systems propose to unify similarities and explicate differences among trace-supporting mechanisms. These similarities and differences can be grouped along different lines, such as from Tinbergen's Four Questions or Marr's levels of analysis. Pursuing these avenues will lead to a versatile paradigm for engram biology, accounting for the different types of engrams, what they do, and how they evolve over time and are used. Alternatively and importantly, if after pursuing this approach "the engram" remains even more ethereally elusive or seems to have been transformed, evaporated or assimilated into different biological processes, leaving in its place an assemblage of new and distinct problems to be explained, we would consider this exciting progress.

Section 2 — What flavour of model are we using in engram biology?

2.1 — Current engram theory is modular and compositional

The existence of multiple levels of engrams aside, the dominant sense of the engram that has emerged in contemporary systems neuroscience research is that the engram is somehow instantiated as cellular ensembles and a brain-wide distributed complex (Josselyn, Frankland & Köhler, 2015; Tonegawa *et al*., 2015; Josselyn & Tonegawa, 2020; Wheeler *et al*., 2013; Vetere *et al*., 2017; Roy *et al*., 2022). But exactly what kind of model is being proposed? Must the engram necessarily be a global and non-decomposable unitary entity, or can discrete "sub-engrams" have fixed identities while still making up a "super-engram" or "engrome," and is this the most appropriate way to carve up the system? Having a clearer understanding of what assumptions this model entails will aid in refining or replacing the existing paradigm.

The terms engram cells, engram cell pathways, engram components, and engram complexes were defined and proposed in Tonegawa *et al*. (2015a) in a contemporary attempt to operationalise Semon's engram theory in light of the last century of advances in biology. In short, there are ensembles of cells (engram cells) that instantiate different components of a much larger, distributed engram (the engram complex) joined together by connectivity pathways between these engram cell subpopulations (Josselyn, Frankland & Köhler, 2015; Josselyn & Tonegawa, 2020). Each of these different sub-engrams or engram building-blocks contribute different pieces of information, such as hippocampal DG and CA3 for contextual information and CA1 for contextual and temporal information (MacDonald *et al*., 2011; Tonegawa *et al*., 2015; Roy *et al*., 2022). A further level of complexity is added by the existence of inhibitory neuronal ensembles ("inhibitory engrams") contributing to broader engram complexes which have been proposed as a general feature of brain function (Baron *et al*., 2017; Koolschijn *et al*., 2019; Sun *et al*., 2020; Nambu *et al*., 2022). The picture that emerges is one with a strong sense of modularity and compositionality, where parts of specialised function can be viewed as plug-and-playable modules, participating in a larger distributed engram which in sum we presume has sufficient capacity and richness of stored information to reconstitute a memory.

If explicitly articulated, a modular or compositional theory may serve as a flexible and helpful architecture to explore. However, there are different ways of treating the parts of any modular theory. An *essentialist* approach explains the behaviour of a system by appealing to the intrinsic power and casual abilities that inhere within its parts. These elements have necessary properties that make them what they are. In contrast, a *relational* approach explains the behaviour of a system by appealing to the unique way the parts are configured with respect to each other (Buzsaki, 2006, p14). Another way of describing these positions is as follows: *what confers on the trace the properties that is has, the intrinsic qualities of trace, or the system in which it is embedded?* We explore the differing consequences of essentialist and relational approaches to the basic modular theory of engram biology across different levels of granularity[13].

2.2 — Relational and essentialist conceptions of 'the engram'

We start with the foundational construct of the engram and engram cells first before talking of cell-ensembles and ensemble-complexes. An essentialist approach views the engram *as* or distributed *in* the constellation of neurons that were activated during learning and whose reactivation produces a behaviour indicative of recall. Any cells that were active during learning that did not get incorporated

[13] We do not have to limit ourselves to two positions. See Lalpane (2016) for a fourfold classification of "stemness" in stem cell research. Additionally, we also do not focus on hypothesis that the engram is molecular in this section in the interest of space. However, there is nothing that precludes imagining differences in relational or essentialist approaches at a molecular level.

into the ensemble, or neurons not active during learning but were recruited later, are an unexplained nuisance. Information is instantiated in a fixed, solid entity, and any alterations will change or corrupt the information preserved. So-called "representational drift" (Rule, O'Leary & Harvey, 2019) might therefore be taken to be a threat to the integrity of the stored information).

A relational approach to defining the engram does not deny that these neurons must play an important part, but attempts to appreciate their role in a wider context. We summarise this attitude with the maxim *it's not what the changes are, it's what the changes do*. For example, Guskjolen and Cembrowski (2023) recommend that the construct of the engram must expand to include both non-neuronal cells and also those neurons that are actively silenced during learning and must be re-silenced for retrieval. In other words, knowing which cells are inactive is just as important as those that are active for information storage and retrieval, because plasticity in engram cells and their connectivity patterns will change the local connectome around them. Unlike artificial networks with scalar activations, E/I balance is an essential principle of nervous system function, where computationally inhibition can be understood as transiently modifying the morphology of other neurons (Buzsaki, 2006). Whether we consider a necessary unity between excitatory and inhibitory cells to be required for something to count as an engram, or we treat excitatory and inhibitory ensembles to be dissociable structures that cooperate, it is clear that any theory of engrams must account for the relevant context that makes an engram possible. Just as a gene is understood not merely as the sequence of base pairs that code for a protein but a region that includes enhancers, promotors, introns, the presence and absence of general and specific transcription factors, so too is there ample room to speculate about the neurobiological context that makes something an engram, or makes it the engram that it is. This sets us up to consider engrams not as monolithic entities or "data structures" of one format, such as a singular "information molecule" or singular kind of information (e.g. binary states) realised in several different biomolecules or larger structures, but rather complex (that is, non-decomposable) informational units that may have many necessary pieces, none of which on their own are sufficient to count as a piece of latent information. But what exactly are these "pieces of information"? Do differences at a cellular, ensemble, or complex level explain the differences in the information content held by one engram versus another?

2.3 — Ensembles are conflated with pieces of information

A common assumption, introduced in Section 2.1, is that different cell-ensembles each instantiate different pieces of information to a brain-wide engram. This encompasses any statements to the effect that one ensemble is said to provide spatial context, another temporal, perhaps a valence component from the amygdala and the various sensory modality engram components from sensory areas, and so forth. However, it is a strong assumption that differences at the ensemble-level are the *right level of explanation* for differences in the "type of information content" the hypothetical engram is posited as preserving. Relating these is a potential category error (Ryle, 1949) and cannot be assumed. It is not yet clear that cell populations in different areas with different functions somehow neatly instantiate the informational components that make up the full engram. It is a promising hypothesis, but must be articulated as a hypothesis and not an assumption. There are at least two things this hypothesis must be able to account for, i) *Intra-ensemble differences:* What is it *about* one ensemble of a given type that differentiates it from another of the same type? For example, temporal engram component X versus temporal engram component Y in CA1. And ii) *Inter-ensemble differences:* What is it *about* one type of ensemble that makes it different from another type? A dentate gyrus engram ensemble differs from a medial prefrontal cortex engram ensemble, but in what way are they the same and in what ways do they differ? Is it region alone? Ideally, we want a taxonomy of engram types that avoids the mistake of grouping heterogeneous things under a unifying construct (lumping error), while also avoiding unnecessary fragmentation of engram-ensemble types into as many as there are possible ensembles (splitting error). The essentialist-relational distinction can be used to draw out questions from these ambiguities.

2.4 — Essentialist and relational views of ensemble and complex function

At an *inter-ensemble* level, there are three essentialist ways of explaining what makes one type of ensemble instantiate the information that it does. First, perhaps it is the *type of changes* undergone (whether intracellular or membrane localised) that confers the informational specificity of that engrams function (1). Given two approximately identical ensembles in different locations, the unique changes undergone determine whether the engram is a fear engram, because the cells underwent "fear changes." The second is that all engram cells undergo the *same kinds of changes*, but it is the cells in which the changes occur that confers informational specificity (2). Since the region does not itself confer properties, either it is something about the intrinsic properties of the cells in that region (2.i), the connectivity of the cells in that region (2.ii), or a combination of both (2.iii). In this case, standard theories of circuit function (e.g. canonical computations) become our explanations. For example, the hippocampal circuit has often been said to implement an indexing operation (Goode *et al*., 2020), meaning the informational content of hippocampal engrams are addresses, rather than a full representation of the entire experience. The third possibility meets (1) and (2) somewhere in the middle: there are different kinds of changes, but location matters too (3). For example, Shpokayte *et al*. (2022) demonstrated topographic segregation (with some overlap) between ventral hippocampal engram cells for aversive and appetitive experiences, as well as different transcriptomic profiles for each population. Whatever the specific changes are in each case, they denote either the positive or negative aspect of the experience. Regardless of which account of how the type of information is determined at the level of the engram ensemble, an essentialist explanation for how these ensembles work together is straightforward—their place in the brain-wide engram-complex is merely additive[14]. Failure to recruit one ensemble still results in a functional memory, minus, say, knowledge of when an experience happened.

A relational approach would argue that, helpful as the three distinctions are, these changes are not explanations for why one ensemble instantiates a given type of latent information over any other. The changes as described do not explain *how* it is that this ensemble is capable of storing a "spatial type of information" while a given other ensemble is capable of "positive valence information storage." We want to know what confers this stored informational capacity and type, and if changes at the level of the ensemble even are the most relevant level for determining the type of latent information. A relational approach would also be open to the possibility that information is not determined by persistent changes at the level of cell ensembles, but only at the level of the brain-wide engram when activated. In contrast to an essentialist view of engram components as fully formed cogs that are simply "called up" by context with appropriate syntactic synchrony, a relational approach might argue that each component is mutually determined, or has a strong dependence on the other components present based on how they are arranged within the connectome (Vetere *et al*., 2017)[15]. This view informs how we test questions about whether or not the cell ensemble is the best level to locate discrete storage arising, whether there is a minimal number of ensembles needed[16], and whether or not ensemble parsing and their syntactic activation order (Buzski, 2010) must be taken into consideration for determining informational specificity at a brain-wide level for successful recall.

2.5 — Cutting across the levels

---

[14] Ensemble-type$_1$ + Ensemble-type$_2$ + Ensemble-type$_3$ = successful retrieval of memory to guide behaviour.

[15] In linguistics and philosophy, this is called semantic inferentialism or conceptual role semantics. In fusional but especially polysynthetic languages, morphemes (the units of meaning) when split apart often do not make sense on their own, thus communicating involves using sentence-words. A missing morpheme is easily guessed (pattern completion).

[16] Say a protein may be functional only with 8 subunits of different isoforms. If only 7 subunits are transcribed, the protein is not coherent, either it does not form, or forms something else with a different function.

To conclude Section 2, the struggle to flesh out the details of the modular framework go hand in hand with a lack of clear hypotheses regarding the basic concept of the engram itself. It is likely that the way we have divided the problems faced at each level is misplaced. Some problems may apply across all levels, and to solve the problem at one level requires solving them across all. We want to see a vibrant future where different modular frameworks, with clearly staked hypotheses as to which functional units (such as cells, ensembles, and complexes) best explain observed informational specificity, will compete to falsify each other in a healthily adversarial way.

Section 3 — If engrams are vehicles, what are they vehicles of?

3.1 — Representations vs neural representations

In this section we introduce a common if controversial distinction made in philosophy between the *vehicle* and *content* of mental representations (Dennett, 1978; Hurley, 1998). Content or semantic content is what a representation is about, such as a particular scene or scent in the case of an immediate perceptual representation and/or an indirect uncoupled conceptual representation such as an idea, abstract concept, or mental map that we use to stand in and flexibly guide decisions in the absence of direct experience. There is much debate over what conditions must be satisfied for something to count as a representation (Ramsey, 2007; Krakauer, 2022), and whether they map neatly to discrete vehicles or not. The vehicle is often taken to be a physical (e.g. neural) part or process (Shea, 2018; Burnston, 2021). However, this debate on representations is typically agnostic as to neural implementation. Thus, both senses of representation just outlined (immediate and uncoupled) are very different senses of representation than when neuroscientists refer to an activity pattern as a "neural representation" of a face or tuning curves or firing rates as *representing* some features of a stimulus (e.g. Chang & Tsao, 2017). This is a softer concept (that of encoding) and one we will assess in section 4. For now we need only note that neural representations are unlikely to qualify as representations in the sense outlined above. Instead, they may be candidate vehicles of representations, or recording-apparatus-relative proxies of vehicles at best[17]. Nonetheless, employing the vehicle/content distinction to engrams is instructive, as it can help us separate out at least three categories of engram concepts with their own sub-types. These three are *engram-as-content, engram-as-vehicle,* and *engram-as-ensemble*.

3.2 — Engrams-as-content

First is the conception that engrams are "stored content." However, we must be cautious about what we mean by content, as there are two senses that need to be distinguished. First is an engram-as-representation view, whereby engrams are a latent version of representational content. This is a view of memory as a representation transforming from latent trace to active and back again, i.e. of the same thing flip-flopping between two states. This is somewhat oxymoronic, as representations as outlined in the previous paragraph pertains only to first-person conscious use of representations (Krakauer, 2022). Nonetheless, although memories may be referred to as representations (Dudai, 2007), engrams are also sometimes also referred to as "the representation of an encoded event or experience" (Moscovitch, 2007). While we colloquially refer to the contents of an image and the contents of a file using the same word *contents*, we caution against calling engrams representations including the sense of downsampled representations or encoded versions of memories. This is because engrams were never envisioned to be *of the same stuff* as memory, hence their separation from ecphory (Robins, 2018). The second sense of engrams-as-content is that engrams are not representations with rich semantic content comparable to human concepts but some form of simple informational content locally for the neural system itself. For example, this includes the idea that an engram *is* a particular index for constructing a given memory. Talking of information rather than representation suggests that there is a difference in format between engram and memory so large it may require a whole other way of conceiving of engrams other than as the "encoded version" of a memory. This is an idea we will return to, but requires clarification.

---

[17] There are more pragmatic accounts of what it means for something to be a representation in practice, under which the following discussion may qualify as treating engrams as representations (Cao, 2022). We do not adopt this pragmatism here for the purposes of avoiding conflation between memories and engrams.

3.3 — Engrams-as-vehicles

As per Semon and recent definitions, the engram is envisioned as an *implementational* construct that concerns physical and chemical biological changes. However, we often fail to distinguish between the conception of *engram-as-content* and *engram-as-vehicle*. There are at least three possible senses of *engram-as-vehicle*, based on what the engram is a vehicle of or *for*. First is that engrams are not memories but the neural vehicles of memories. However, whatever the neural implementation of memory is, this concerns what happens at time of recall. The engram is a separate neural implementation which persists when a memory is not being created. Engrams, therefore, do not have to be the vehicles of memories[18]. Given this, we may propose a second though similar option—that engrams are not the vehicles of memory but are vehicles instantiating a stored, reformatted, and downsampled encoding of a memory. This does not abet the problems outlined with the *engram-as-representation* view above. Finally, there is the conception that engrams are not vehicles of representations but vehicles of a simpler informational content. The content may not be seen as semantic in the representational sense of being about or standing in for things outside the brain, rather the content of the vehicle is what it does: for an engram theory that places primacy on connectivity, it may be shifting the probability of a given neural activity pattern being expressed through altering the pathways that neural activity is capable of taking. At this point in time, it may no longer be fruitful to look at the engram from separate *vehicle* and *content* perspectives, or talk of content at all (Figure 2). This is the view we are sympathetic to and wish to see clarified further in the context of engram biology.

3.4 — Engrams-as-ensembles

The last category of engram views is that engrams have already been discovered and they the are unique sparse ensembles that we can partially label and manipulate. These cell-ensembles may or may not also qualify as vehicles and could overlap the previous section (see Figure 2). However, if these ensembles are not the vehicles of stored information, then these engrams do not fulfil the objective of explaining the biological basis of memory and we will need to posit some other construct that fulfils this role. If these ensembles are indeed the vehicles, then we need to be able to bring to the foreground those most relevant changes and relegate those not relevant but still picked up by the tagging protocol. However, current experimental methods have not revealed to us what these most relevant changes are (are they molecular changes, connectivity changes, or something else?). Although adopting the engrams-are-ensembles definition may seem pragmatic, it gives the illusion that the research program of engram biology has finished answering its fundamental question and moved on to simply looking for applications to behavioural neuroscience. Although the existing techniques of engram biology are powerful and will undoubtedly lead to more applications where a true understanding of the biological basis of memory is not essential (for example, in understanding the biology of forgetting, addiction, trauma, delusions, etc.), if we want the engram to explain the specificity of memory, then we must admit that satisfying Semon's criteria is only a beginning and is not alone sufficient (Ortega-de San Luis & Ryan, 2022). In other words, we have not yet isolated the engram. Rather, these labelled ensembles may encapsulate, be adjacent to, or partially instantiate the "enduring physical/chemical changes" that constitute the engram[19].

---

[18] This conflation may come from using "memory" to refer ambiguously to both memory and memory traces.

[19] Different IEG expression profiles do not always overlap, leading to the confusion that two disjunct labelled populations may both be called engrams. At best, differences in IEG function may be for implementing preservation of different types or layers of information that ultimately comprise the engram (Nambu *et al*., 2022).

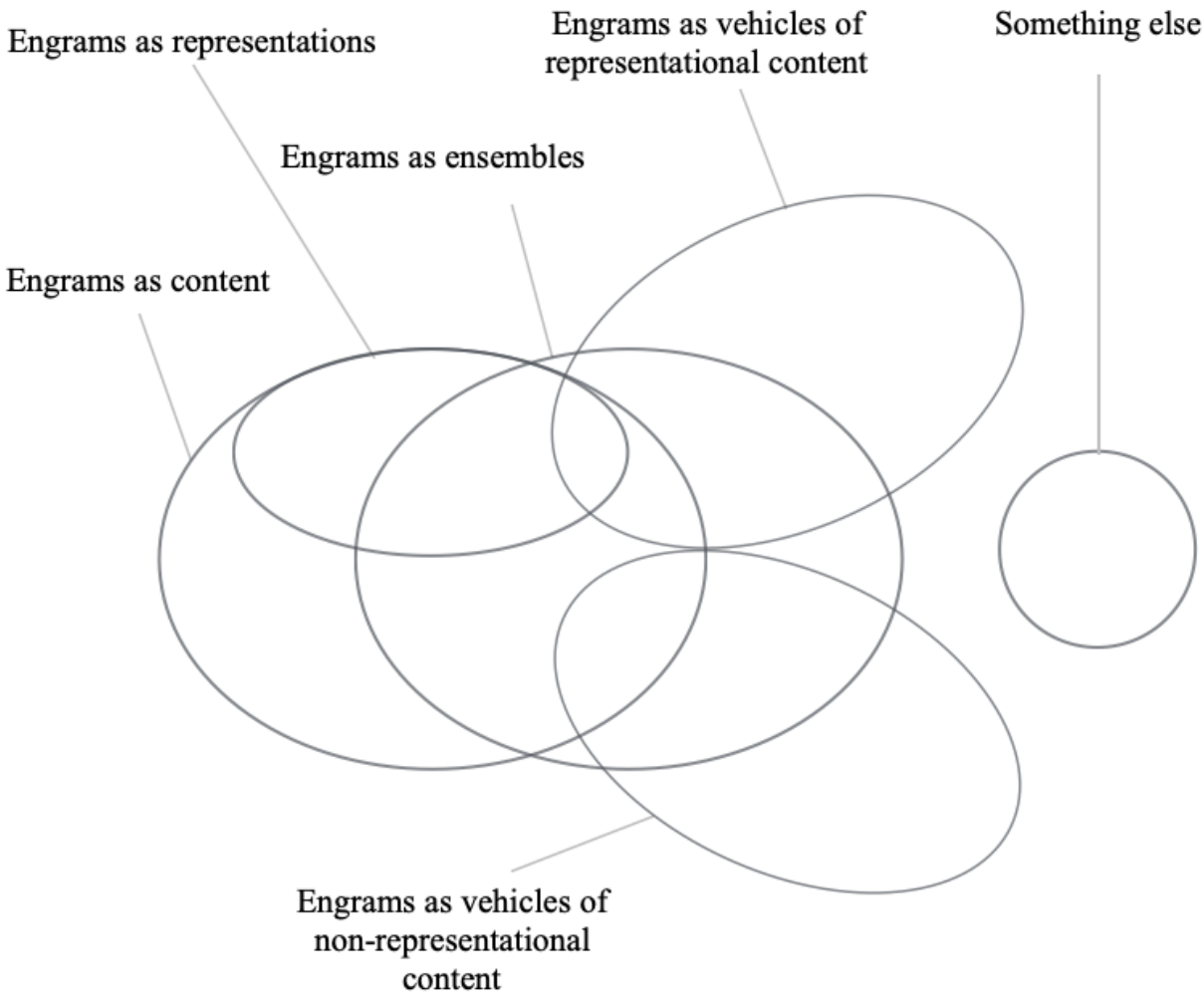


Figure 2: Different overlapping conceptions of what engrams should be.

3.5 — A vehicle-information distinction

These distinctions are controversial[20] and coarsely wielded, but entertaining these differences opens up a space for considering other possibilities. *If engrams do not implement storage of representations, what do they have, or what can they be said to be/function as?* Applying the vehicle-content distinction dissociates different senses of the word engram, but this is not its only utility. Even if engrams do not have representational content, there may be utility in employing a modified *vehicle-information* distinction, whereby "information…persists via a memory trace, which is the vehicle for information preservation" (Mace, 2021). There are two avenues leveraging this distinction helps us consider. Just as we presume that not anything can count as a representational vehicle, we presume that not everything is *competent* to function as an engram. What would it take for a vehicle to function as preserving information and what criteria would exclude a biological structure from qualifying as competent (e.g. scar tissue preserves persistent changes despite cellular turnover)? This means further refining Semon's four criteria for what it takes for something to count as an engram (see Josselyn, Köhler & Frankland, 2015). Second, this distinction makes clear an asymmetrical bias in engram biology research towards characterising the vehicle of stored information. Discoveries about the

[20] To distinguish between latent information and vehicles is not to separate them and treat them as separate entities, but to look at the same phenomena from different perspectives. Nonetheless, it creates a dualistic divide that is not always desirable (Mitchell, 2023).

vehicle, even figuring out the mechanism that makes something an engram, does not necessarily tell us anything about what information is stored. As will be discussed in Section 4, we have a poor understanding of what this latent information may be. At the very least, we can use the vehicle-information distinction to check ourselves, asking "do the questions addressed in this study focus solely on the vehicle, or do they also tell us exactly what information is stored?"

This has precedent in biology, as a clear vehicle-information distinction is seen in genetics. A gene may be given a physicochemical description as a string of nucleic acids with certain molecular properties, or an informational description as a *symbolic* sequence that has a particular meaning in the context of the cell. It can be interpreted though specific cellular processes into a functional peptide or nucleic acid. DNA/genes lead a "double life" as matter and *semiotic symbol* (Gazzaniga, 2018; Pattee 1987/2012). In the case of genetics, an entirely new way of thinking about biology had to be created, that of thinking in terms of a symbolic code system: The eventual cracking of the DNA triplet codon to tRNA amino acid allocation, proved that the mapping is arbitrary and therefore code-like, with redundancy. This shift required a level of abstraction that turned genetics from a form of molecular biology into a kind of biological information science.

Important for our discussion is that information does not fall out of physical knowledge of the vehicle. The physical form of two genes with the same base-pair length may be near-identical, but what the genes code for may differ radically (a plant enzyme or a SARS-CoV-2 cell surface protein). The meaning can be divergent even from the same information in cases of differential splicing and post-translational modifications. Whether we are looking at DNA with the fine resolution of an electron microscope or a chromosomal karyotype, we are not able to see this genetic information. We are characterising its vehicle. When we look at a labelled cell ensemble in a brain tissue slice, we are looking only at one side of the engram. Distinguishing between engram and non- or other-engram cells is not looking at stored information. Double-blinded to a tissue slice of labelled cells, we could not say "this is a fear engram" or "this is an episodic memory from this place and time." Again, *knowledge of the vehicle of information is not enough to tell us what this engram is for* (Eichenbaum, 2016). In the case of engram neuroscience, we may even have the resolution needed, but cannot from the forest pick out i) what the vehicle is, or ii) how it affords the capacity for the specificity of memory and therefore behaviour. These are two different targets of explanation; analogously, explaining the structure of DNA is not the same as explaining the role of DNA in heredity and evolution. Regarding i), identifying the vehicle is not a binary YES/NO achievement, since there are levels of precision. When Avery, MacLeod & McCarty (1944) and Hershy & Chase (1952) discovered DNA to be the molecule responsible for transmitting genetic information and not protein, they had identified the vehicle in a coarse sense. This did not tell us what the genetic information was, or reveal which changes in the physical vehicle were most relevant for storing information. When we find that unique labelled cell-ensembles are capable of reinstating a behaviour, we may (at best) have identified engrams in a similarly coarse sense.

We would not say that a gene stores representations of proteins, organelles or organs. Rather there is a stable molecule that plays a symbolic role (Pattee, 1987/2012) in constraining the creation of functional molecules in a highly reproducible way during the lifetime of an organism. This informational description, that of a 1-dimensional sequence and code, is an abstract and imperfect model we create to understand the semiotic logic implemented in biological systems at a cellular level. Similarly, we are not suggesting we try to "find the information in the brain" in a literal sense, rather we are arguing that engram biology needs a framework that makes a similarly abstract leap to help us understand whatever constructs serve as the biological basis of memory. In summary, we feel a vehicle-information distinction is important, because it illustrates how engram biology is focused exclusively on the vehicle side of the engram and how there is a dark side of the moon to engram biology that is rarely discussed. With this in mind, we now focus on how we might understand this information.

Section 4 — Information and coding

4.1 — The necessity of a theory of information

Talk of information-processing is ubiquitous throughout neuroscience, but what is meant by information is rarely specified. This same state of affairs applies in engram biology, where understanding the biological basis of memory is often framed in terms of discovering how plasticity allows information to be stored in the brain or manages to persist through time (Josselyn & Tonegawa, 2020). But what exactly is this information? The use of information-concepts in biology is controversial (Godfrey-Smith & Sterelny, 2016). However, in engram biology there are behavioural and other phenomena that may be difficult to explain without invoking the existence of stored information. For example, various forms of memory learned during the infantile amnesia window are not lost but rather inaccessible and can be artificially retrieved (Guskjolen et al., 2018; Power et al., 2022), and other apparently lost memories can also be recalled (Ryan *et al*., 2015; Roy *et al*., 2016; Perusini *et al*., 2017; Abdou et al., 2018; Poll *et al*., 2020; Lei *et al*., 2022; Bolsius *et al*., 2023; Autore *et al*., 2023; O'Leary *et al*., 2023). Accounting for these phenomena without positing latent information, or another concept that performs a related function, is problematic. To the extent that it is helpful to posit the existence of information and use information-concepts, a stronger theory of what this information is is needed (Robins, 2023). We therefore survey two different existing general conceptions of information, and conclude that neither are completely satisfactory for use in engram biology. Instead, the techniques of engram biology mandate creating a new conception of information for use in neuroscience.

4.2 — Classical information theory and coding metaphors

The first general class of views on information are inflected by classical information theory and computer science (Stone, 2022). Information is treated as something entering the brain from an external source (e.g. a stimulus) whereupon it is "encoded" (Roediger, 2007), and is passed around between neurons, perhaps travelling in inter-spike intervals as a rate or combinatorial code (Gallistel, 2017). Learning here is the acquisition and consolidation of relevant information extracted from "sense data," with stored information held in something like a weight matrix (Luo, 2020, p.448). However, classical information theory is not a theory of information content but of quantifying how information can be optimally transmitted. It explicitly excludes accounting for the semantics, or meaning, of what is communicated (whether termed a message, signal, or information)[21]. Neuroscientific theories about how information is coded, the computational and therefore metabolic benefits of using coding principles (Friston, 2010), inherit this same problem. Although the details of such views can differ, as not everyone might agree that the brain computes using *bits* (Sterling & Laughlin, 2015), or uses a von-Neumann architecture (Gallistel & King, 2010), they do not provide an account for what the information being coded is. This leaves much to be desired, as neuroscientists often care about the meaning of what brain signals are communicating.

The limitations of this heavily abused coding metaphor are summarised in Brette (2019). Principally, when we look for a correlation between any external variable, whether sensory or higher order task structure, and neural activity (e.g. Ca^{2+} events or Blood Oxygenation Level Dependent patterns as proximate measurements), we have assumed the existence of information, but only in our capacity as a third person observer fixing a relation between two variables, one inside and one outside the brain. The brain does not have this third-person perspective, and so this cannot be assumed to be the information stored in engrams. Instead, for any future engram theory seeking inspiration from

[21] In telecommunications and computer science, this is not a problem, as it is humans who design the hardware and determine the arbitrary conventions for how variables should be encoded for other humans to decode them.

computer science, it may be more fruitful to direct attention to computer hardware engineering principles.

4.3 — Information is relational and momentary

A second view, one that tries to accommodate the meaning of information, is that information is not a noun, property or thing that can ever be localised, but is inherently relational or contextual. There is nothing *in* DNA. Nucleic acids like DNA and RNA do not *carry* information by themselves. Outside the context of the cell, they are merely polynucleotides. There is only information *for* a given system in context. Neurotensin, interleukin-17, insulin, and serotonin only have biological meaning or unique consequences in the context of signalling, and are differentially responded to by cells on the basis of the receptors they express. The meaning of a signal has physical constraints, such as molecular mass and hydrophilicity, but these constraints do not necessarily determine the meaning, which is fixed by evolutionary convention and may be modified by learning. Information in biology is not something that inheres to any physical pattern, it is information only in relation to or for some system in context (Brette, 2019; Buzsaki, 2006; Deacon, 2021). This means that information can never be something held or stored, but is being continually created.

Alternatively, this view may be restricted to apply only to first-person consciousness experience, such that it is only permissible to talk of information when dealing with mental states, such as representations, as in Section 3.1. In the context of theories of memory, this may fit with a view of information only existing at the time of recall (Schacter & Addis, 2007). Under this view, the function of the engram (if posited) is not to store information, but to participate in the process of reproducing information meaningful for action in the relevant context. There is none of this information *in* the unique pattern of persistent changes that make an engram. Although viewing recall as a process of forming transiently existing information *using* engrams fits with Semon's account of ecphory, this view still leaves the biological basis of memory to be explained.

Both of the perspectives in 4.3, which we are more sympathetic to than that of 4.2, needs much more refinement as they are limited to live processes unfolding moment to moment, and therefore do not provide a clear way of thinking about information that is not live, dormant, or potential. This is orthogonal to the objectives of engram biology, which seeks to explain i) a mechanism for how organisms appear to possess dormant information at all, and ii) explain how it is that engrams have the specific potential they have for eliciting a given behaviour. In other words, *how* it is that the reaction of a specific ensemble, *rather than any other,* exerts such strong behavioural specificity.

4.4 — What next?

The computationalist view of information-processing and the information as relational and created view both have appealing characteristics, but neither are adequate for understanding engrams. It is unclear what form of theory would be adequate. This is not a question that would have been asked two decades ago, before the development of engram technology. It is a new problem. Nonetheless, there are several issues that any theory of engrams ought to be able to deal with: i) *arbitrariness*, ii) *coding-strength and universality*, and iii) *reading*.

4.4.i — Arbitrariness

Although decoding the meaning of any purported information in the brain is not possible if an external variable is being used as a reference point by a third person observer, this does not mean that internal codes within the nervous system, rather than between nervous system and world, are not employed.

A key characteristic of a code is the use of a *symbolic* or *arbitrary* relation, such as between a triplet codon "standing for" a particular amino acid (but unlike between the triplet codon of DNA and a triplet of RNA, which is a relation based on similarity or resemblance). Understanding the origins and degree of arbitrariness will be important to deciphering any internal code. For example, the taste or meaning of a chemical compound is *not* determined by the intrinsic qualities of the chemical compound but is fixed instead by anatomical wiring, namely whether the neuron expressing the relevant taste receptor is part of a labelled line to sweet or bitter cortex (Peng *et al*., 2015; Wang *et al*., 2018). Thus, to decode the meaning requires knowledge of evolutionary and developmental history of the organism. If we cannot infer anything about the properties of a chemical compound at a sensory level based on the neuronal activity it elicits, "reading out" what mnemonic features a given ensemble or complex would be capable of reinstating (without activating the ensemble and looking at neural activity) is considerably more conceptually challenging. There is the possibility that the cells that comprise a given engram are entirely arbitrary. For example, if allocation is always stochastic, then there is no other reason why *these* cells, rather than *those* cells, should be involved. At least in the hippocampus, silencing the specific neurons previously labelled during conditioning while being exposed to the same conditioning paradigm simply results in a new and non-overlapping engram to form (Tanaka *et al*., 2014). Therefore an engram is constrained, but not determined by anatomy. However, if engrams are not determined by their anatomy, this does not mean that connectome-level analysis is not relevant. Instead, the kinds of invariants that may form the basis of an internal code may be structural topological invariants, rather than geometric properties of the relevant network[22]. The degree of arbitrariness may also vary regionally and phylogenetically.

4.4.ii — Coding-strength and universality

Related to arbitrariness is how strong the internal code is, namely what degree of determination over recall and ultimately behaviour a given engram is capable of exerting. Cao & Rathkopf (2019) suggest that nervous systems learn their own internal modest codes. This means that there is no one-to-one correspondence between neural signal and function. The function is indeterminate or non-disjunct. For example, Dorst *et al*. (2023) demonstrated that the activation of the same ensemble while varying the size of the environment results in different behaviours. Thus, it is unlikely there will be a one-ensemble-one-behaviour relationship. The causal contribution of a latent structure may vary as it confluences with the online contextual information of the retrieval environment (Schacter, 2007). In the case of instincts, these may involve internal codes that do not have to be learned, and may be at the stronger end of the spectrum (Ryan *et al*., 2021).

Closely tied to the concept of coding-strength is *coding-universality*, that is, whether there is any conservation in internal codes between individuals (e.g. how conserved one animal's internal code for a simple specific associative fear memory is relative to another's). If there is no selection pressure for there to be any level of similarity, then even if it were possible to "read out" any features based off a labelled ensemble in one animal, this knowledge would not be transferable. However, knowing the internal code may not be necessary to test this hypothesis. As discussed in Section 3.6, before the "hereditary factors" that explained the specificity to heredity were known, transformation experiments were performed. RNA transfer experiments are one kind of transformation experiment, and have been successful in *Aplysia* (Bédécarrats *et al*., 2018), but have yet to be tested in model organisms with larger nervous systems. Even if we do not think that RNA is the source of the "enduring changes" that comprise the engram in mammalian nervous systems, surrogate transformation experiments can still be performed. If a we hypothesise that the topology of the connectome is where the enduring changes are held (Ortega-de San Luis *et al*., 2023), then based off the topological characteristics of animals that have learned the behaviour, experimental manipulation

[22] There is an important distinction to be made here between structural topology and topological properties of high dimensional data.

that implements that same topological invariance should change the behaviour in a way that a topological theory predicts, but simple anatomy would not.

4.4.iii — Reading

Finally, any new theory of information must combat the *reading problem*. It is important to ask how these engram vehicles have their informational contents read. Here is where the standard coding metaphor can creep back in. Gallistel & King (2009) argue that the brain uses von-Neumann addressable read-write mechanisms where information is written into cells in the changes they undergo, and that it is then read by a reader mechanism that knows this code. Others deem it too awkward to apply a classical computer architecture onto the brain. Godfrey-Smith (2014) draws the distinction between two kinds of biological information. Genetic information does indeed have a read mechanism. It has a defined coding-relationship for how to go from the latent informational structure, the gene, to a protein, but no symmetrical write mechanism. Instead mutations, duplications, and deletions occur during the course of cell division and evolution. Therefore, there is an *evolve-read* mechanism. The brain conversely has no clear reader(s), rather a *write-only* (Donahue, 2010) or rather *write-activate* system: "Evolved neural machinery has the function of introducing marks or traces into the brain as a result of experience, but these marks have useful effects on behaviour without being read" (Godfrey-Smith, 2014). While there may be a kind of distributed reading that occurs, locating the processes doing this without turning them into readers is notoriously difficult (Cao, 2012), although attempts to consider partial and multiple observers within a system have been proposed (Kolchinsky & Wolpert, 2018). Again, should it prove less cumbersome to think without the concepts of information, storage, and codes, a robust theory of how memories can be created out of accumulated changes that explains the specificity in behaviour from reactivating labelled cell ensembles is still required.

Discussion

For all the methodological advances in recent years in characterising the biological basis of memory, conceptual advances in understanding whether or not there are memory traces and what their precise functions are, or what they might look like, has not kept pace. There is a multiplicity of unclear views, areas of consideration missing, and much yet to be articulated and clarified before we have a formal theory. On a practical level, how does the above analysis help? First and most straightforward, one can ask "which sense of engram have I been using? Multiple? None of those discussed?" This piece has not aimed to be prescriptive regarding which sense of the engram should be used. A classification system that unifies different levels of engrams or information storage mechanisms has the potential to tie together disparate fields into a research program of its own. Secondly, we can better appraise the explanatory remit of experimental approaches. We can ask "what is this study aiming to address and what do these results inform? Do they further our understanding primarily of the implementational vehicle aspect of the engram, or do they inform the more abstract, informational questions of the engram?" Thirdly, this piece emphases the size and shape of the explanatory gap we face in engram biology and systems neuroscience generally, and the nature of some of the questions we will need to engage with and move beyond in order to develop a deeper understanding. Finally, what exactly is broadly at stake here for engram biology and systems neuroscience? We believe what is at stake is our understanding of how the brain is capable of storing anything. Here engram biology can pursue two paths. As a sub-discipline of neuroscience coming into its own, it can attempt to help resolve the existing conceptual confusions of the wider field by articulating an engram theory within this worldview using existing, poorly defined concepts. Alternatively, the techniques and ambitions of engram neuroscience can be used as an inside job, to challenge dominant metaphors and attempt articulating a wholly different theory of what there is to be found in the brain.

Acknowledgements

The authors would like to thank Caitlin Mace, Clara Ortega-de San Luis and Paul Conway for their helpful comments on a draft of this manuscript, and members of the Ryan Lab and the Trinity College Dublin *Representations: Past, Present, and Future* group for useful discussions.

References

Abdou, K., Shehata, M., Choko, K., Nishizono, H., Matsuo, M., Muramatsu, S. I., & Inokuchi, K. (2018). Synapse-specific representation of the identity of overlapping memory engrams. *Science (New York, N.Y.)*, *360*(6394), 1227–1231. https://doi.org/10.1126/science.aat3810

Autore, L., O'Leary, J. D., Ortega-de San Luis, C., & Ryan, T. J. (2023). Adaptive expression of engrams by retroactive interference. *Cell reports*, *42*(8), 112999. https://doi.org/10.1016/j.celrep.2023.112999

Avery, O.T., MacLeod, C.M., McCarty, M. (1944). Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types: Induction of Transformation by a Deoxyribonucleic Acid Fraction Isolated from Pneumococcus Type III. *Journal of Experimental Medicine*. **79** (2): 137–158. doi:10.1084/jem.79.2.137.

Barack, D.L., Krakauer, J.W. (2021). Two views on the cognitive brain. *Nat Rev Neurosci* **22**, 359–371. https://doi.org/10.1038/s41583-021-00448-6

Barron, H. C., Vogels, T. P., Behrens, T. E., & Ramaswami, M. (2017). Inhibitory engrams in perception and memory. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(26), 6666–6674. https://doi.org/10.1073/pnas.1701812114

Bédécarrats, A., Chen, S., Pearce, K., Cai, D., & Glanzman, D. L. (2018). RNA from Trained *Aplysia* Can Induce an Epigenetic Engram for Long-Term Sensitization in Untrained *Aplysia*. *eNeuro*, *5*(3), ENEURO.0038-18.2018. https://doi.org/10.1523/ENEURO.0038-18.2018

Bolsius, Y. G., Heckman, P. R. A., Paraciani, C., Wilhelm, S., Raven, F., Meijer, E. L., Kas, M. J. H., Ramirez, S., Meerlo, P., & Havekes, R. (2023). Recovering object-location memories after sleep deprivation-induced amnesia. *Current biology : CB*, *33*(2), 298–308.e5. https://doi.org/10.1016/j.cub.2022.12.006

Brette, R. (2019) 'Is coding a relevant metaphor for the brain?', *Behavioral and Brain Sciences*. 2018/07/16. Cambridge University Press, 42, p. e215. DOI: 10.1017/ S0140525X19000049.

Burnston, D.C. (2021). Contents, vehicles, and complex data analysis in neuroscience. *Synthese* 199 (1-2):1617-1639.

Buzsáski G. (2006). *Rhythms of the Brain*. Oxford University Press.

Buzsáki G. (2010). Neural syntax: cell assemblies, synapsembles, and readers. *Neuron*, *68*(3), 362–385. https://doi.org/10.1016/j.neuron.2010.09.023

Campbell, R. R., & Wood, M. A. (2019). How the epigenome integrates information and reshapes the synapse. *Nature reviews. Neuroscience*, *20*(3), 133–147. https://doi.org/10.1038/s41583-019-0121-9

Cao, R. (2012). A Teleosemantic approach to information in the brain. *Biology and Philosophy*.

Cao, R. (2022). Putting representations to use. *Synthese* 200 (2).

Cao, R., & Rathkopf, C. (2019). Modest and immodest neural codes: Can there be modest codes?. *The Behavioral and brain sciences*, *42*, e221. https://doi.org/10.1017/S0140525X19001420

Cisek, P. (2019) 'Resynthesizing behavior through phylogenetic refinement', *Attention, Perception, & Psychophysics*, 81(7), pp. 2265–2287. doi: 10.3758/s13414-019-01760-1.

Chang, L., & Tsao, D. Y. (2017). The Code for Facial Identity in the Primate Brain. *Cell*, *169*(6), 1013–1028.e14. https://doi.org/10.1016/j.cell.2017.05.011

Chen, S., Cai, D., Pearce, K., Sun, P. Y.-W., Roberts, A. C., and Glanzman, D. L. (2014). Reinstatement of long-term memory following erasure of its behavioral and synaptic expression in Aplysia. eLife 3:e03896. doi: 10.7554/eLife.03896

Chklovskii, D. B., Mel, B. W., & Svoboda, K. (2004). Cortical rewiring and information storage. *Nature*, *431*(7010), 782–788. https://doi.org/10.1038/nature03012

Clark, A., & Chalmers, D. (1998). The Extended Mind. *Analysis*, *58*(1), 7–19.

Craver, C. (2007). *Explaining the Brain*. Oxford University Press.

Deacon, T. W. (2021). How molecules became signs. *Biosemiotics*, **14**, 537–559. https://doi.org/10.1007/s12304-021-09453-9

Dennet D. C. (1978). Toward a Cognitive Theory of Consciousness, in C.W. Savage, ed., Minnesota Studies in the Philosophy of Science, Volume 9: Perception and Cognitions, Issues in the Foundation of Psychology. Minneapolis: University of Minnesota Press, pp201-228.

Donald, M. (1991). Origins of the modern mind. Harvard University Press.

Donohoe, J. (2010). Man as Machine. Review of Memory and the Computational Brain, by C. R. Gallistel and A. P. King. *Behavior and Philosophy* 38:83–101.

Dorst, K., Senne, R.A., Diep, A.H., de Boer, A.R., Suthard, R.L., Leblanc, H., Ruesch, E.A. Skelton, S., McKissick, O.P., Bladon, J.H., Ramirez, S. (2023). *bioRxiv*. 2023.02.24.529744; https://doi.org/10.1101/2023.02.24.529744

Dudai, Y. Memory: It's all about representations. In H.L. Roediger, Y. Dudai & S.M. Fitzpatrick (eds.) *Science of Memory: Concepts* (pp. 17-21). Oxford University Press.

Eichenbaum H. (2016). Still searching for the engram. *Learning & behavior*, *44*(3), 209–222. https://doi.org/10.3758/s13420-016-0218-1

Friston, K. (2010) 'The free-energy principle: a unified brain theory?', *Nature Reviews Neuroscience*, 11(2), pp. 127–138. doi: 10.1038/nrn2787.

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. https://doi.org/10.1016/j.neuron.2020.07.011

Gallistel C. R. (2017). The Coding Question. *Trends in cognitive sciences*, *21*(7), 498–508. https://doi.org/10.1016/j.tics.2017.04.012

Gallistel, C. R., & King, A. P. (2009). *Memory and the computational brain: Why cognitive science will transform neuroscience*. Wiley-Blackwell.

Gazzaniga, M. (2018). *The Consciousness Instinct*. Farrar, Straus and Giroux.

Gershman, S.J. (2023). The molecular memory code and synaptic plasticity: a synthesis. *BioSystems, 224*, 104825.

Godfrey-Smith, P. (2014). Sender-Receiver Systems within and between Organisms. *Philosophy of Science*, *81*(5), 866–878. https://doi.org/10.1086/677686

Gould, S., & Vrba, E. (1982). Exaptation—a Missing Term in the Science of Form. *Paleobiology, 8*(1), 4-15. doi:10.1017/S0094837300004310

Guskjolen, A., & Cembrowski, M. S. (2023). Engram neurons: Encoding, consolidation, retrieval, and forgetting of memory. *Molecular psychiatry*, 10.1038/s41380-023-02137-5. Advance online publication. https://doi.org/10.1038/s41380-023-02137-5

Hardt, O., & Sossin, W. S. (2020). Terminological and Epistemological Issues in Current Memory Research. *Frontiers in molecular neuroscience*, *12*, 336. https://doi.org/10.3389/fnmol.2019.00336

Hershey A, Chase M (1952). Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J Gen Physiol*. 36 (1): 39–56. doi:10.1085/jgp.36.1.39.

Hoffmeyer, J. (2008). Semiotic Scaffolding Of Living Systems. In: Barbieri, M. (eds) Introduction to Biosemiotics. Springer, Dordrecht.

Hurley, S.L. (1998). *Consciousness in Action*. Harvard University Press.

Jonas, E., & Kording, K. P. (2017). Could a Neuroscientist Understand a Microprocessor?. *PLoS computational biology*, *13*(1), e1005268. https://doi.org/10.1371/journal.pcbi.1005268

Josselyn, S. A., Köhler, S., & Frankland, P. W. (2015). Finding the engram. *Nature reviews. Neuroscience*, *16*(9), 521–534. https://doi.org/10.1038/nrn4000

Josselyn, S. A., & Tonegawa, S. (2020). Memory engrams: Recalling the past and imagining the future. *Science (New York, N.Y.)*, *367*(6473), eaaw4325. https://doi.org/10.1126/science.aaw4325

Kandel E. R. (2001). The molecular biology of memory storage: a dialogue between genes and synapses. *Science (New York, N.Y.)*, *294*(5544), 1030–1038. https://doi.org/10.1126/science.1067020

Kauffman, S. (2019). *A World Beyond Physics*. Oxford University Press.

Kol, A., & Goshen, I. (2021). The memory orchestra: the role of astrocytes and oligodendrocytes in parallel to neurons. *Current opinion in neurobiology*, *67*, 131–137. https://doi.org/10.1016/j.conb.2020.10.022

Kolchinsky, A., & Wolpert, D. H. (2018). Semantic information, autonomous agency and non-equilibrium statistical physics. *Interface focus*, *8*(6), 20180041. https://doi.org/10.1098/rsfs.2018.0041

Koolschijn, R. S., Emir, U. E., Pantelides, A. C., Nili, H., Behrens, T. E. J., & Barron, H. C. (2019). The Hippocampus and Neocortical Inhibitory Engrams Protect against Memory Interference. *Neuron*, *101*(3), 528–541.e6. https://doi.org/10.1016/j.neuron.2018.11.042

Krakauer, J.W. (2022). Representation in Cognitive Science by Nicholas Shea: But Is It Thinking? The Philosophy of Representation Meets Systems Neuroscience. *Studies in History and Philosophy of Science*, *92*, 267-269. https://doi.org/10.1016/j.shpsa.2021.05.014.

Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., & Poeppel, D. (2017). Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron*, *93*(3), 480–490. https://doi.org/10.1016/j.neuron.2016.12.041

Laplane, L. (2016). *Cancer Stem Cells: Philosophy and Therapies*. Harvard University Press.

Langdon, C., Genkin, M., & Engel, T. A. (2023). A unifying perspective on neural manifolds and circuits for cognition. *Nature reviews. Neuroscience*, *24*(6), 363–377. https://doi.org/10.1038/s41583-023-00693-x

Lei, B., Lv, L., Hu, S., Tang, Y., & Zhong, Y. (2022). Social experiences switch states of memory engrams through regulating hippocampal Rac1 activity. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(15), e2116844119. https://doi.org/10.1073/pnas.2116844119

Lewontin, R. C. (1983). The Organism as the Subject and Object of Evolution. In R. Levins and R.C. Lewontin (eds.), *The Dialectical Biologist*. Harvard University Press.

Liu, X., Ramirez, S., Pang, P. T., Puryear, C. B., Govindarajan, A., Deisseroth, K., & Tonegawa, S. (2012). Optogenetic stimulation of a hippocampal engram activates fear memory recall. *Nature*, *484*(7394), 381–385. https://doi.org/10.1038/nature11028

Luo, L. (2020). *Principles of Neurobiology*. (2nd ed). Taylor & Francis.

MacDonald, C. J., Lepage, K. Q., Eden, U. T., & Eichenbaum, H. (2011). Hippocampal "time cells" bridge the gap in memory for discontiguous events. *Neuron*, *71*(4), 737–749. https://doi.org/10.1016/j.neuron.2011.07.012

Mace, C. B. (2021). *Casting light on the search for engrams: On the reductionism-mechanism debate* (Order No. 28644486). Available from ProQuest Dissertations & Theses A&I; Publicly Available Content Database. (2597766427).

Martin, S. J., Grimwood, P. D., & Morris, R. G. (2000). Synaptic plasticity and memory: an evaluation of the hypothesis. *Annual review of neuroscience*, *23*, 649–711. https://doi.org/10.1146/annurev.neuro.23.1.649

Mitchell, K. (2023). The origins of meaning – from pragmatic control signals to semantic representations. *PsyArXiv*. Doi: 10.31234/osf.io/dfkrv

Moskovitch, M. (2007). Memory: Why the engram is elusive. In H.L. Roediger, Y. Dudai & S.M. Fitzpatrick (eds.) *Science of Memory: Concepts* (pp. 17-21). Oxford University Press.

Nambu, M. F., Lin, Y. J., Reuschenbach, J., & Tanaka, K. Z. (2022). What does engram encode?: Heterogeneous memory engrams for different aspects of experience. *Current opinion in neurobiology*, *75*, 102568. https://doi.org/10.1016/j.conb.2022.102568

Ortega-de San Luis, C., & Ryan, T. J. (2022). Understanding the physical basis of memory: Molecular mechanisms of the engram. *The Journal of biological chemistry*, *298*(5), 101866. https://doi.org/10.1016/j.jbc.2022.101866

Ortega-de San Luis, C., Pezzoli, M., Urrieta, E., Ryan, T.J. (2023). Engram cell connectivity as a mechanism for information encoding and memory function. *bioRxiv*. 2023.09.21.558774; doi: https://doi.org/10.1101/2023.09.21.558774

O'Leary J., Bruckner, R., Autore, L., Ryan, T.J. (2023). *bioRxiv*. 2023.04.15.537013; https://doi.org/10.1101/2023.04.15.537013

Pattee, H.H. (2012). Cell Psychology: an evolutionary approach to the symbol-matter problem. In H.H. Pattee & Rączaszek-Leonardi, J. (Ed.), *LAWS, LANGUAGE and LIFE: Howard Pattee's classic papers on the physics of symbols with contemporary commentary*. (Original work published 1987)

Perusini, J. N., Cajigas, S. A., Cohensedgh, O., Lim, S. C., Pavlova, I. P., Donaldson, Z. R., & Denny, C. A. (2017). Optogenetic stimulation of dentate gyrus engrams restores memory in Alzheimer's disease mice. *Hippocampus*, *27*(10), 1110–1122. https://doi.org/10.1002/hipo.22756

Poeppel, D., & Idsardi, W. (2022). We don't know how the brain stores anything, let alone words. *Trends in cognitive sciences*, *26*(12), 1054–1055. https://doi.org/10.1016/j.tics.2022.08.010

Poll, S., Mittag, M., Musacchio, F., Justus, L. C., Giovannetti, E. A., Steffen, J., Wagner, J., Zohren, L., Schoch, S., Schmidt, B., Jackson, W. S., Ehninger, D., & Fuhrmann, M. (2020). Memory trace interference impairs recall in a mouse model of Alzheimer's disease. *Nature neuroscience*, *23*(8), 952–958

Power, S., Stewart, E., Zielke, L., Byrne, E., Ortega-de San Luis, L., Lynch, L., Ryan, T. (2022). Immune activation state modulates retrieval of infant engrams. *bioRxiv*. doi: https://doi.org/10.1101/2022.04.24.488172

Queenan, B. N., Ryan, T. J., Gazzaniga, M. S., & Gallistel, C. R. (2017). On the research of time past: the hunt for the substrate of memory. *Annals of the New York Academy of Sciences*, *1396*(1), 108–125. https://doi.org/10.1111/nyas.13348

Ramsey, W.M. (2007). *Representation Reconsidered*. Cambridge University Press.

Ramirez, S., Liu, X., MacDonald, C. J., Moffa, A., Zhou, J., Redondo, R. L., & Tonegawa, S. (2015). Activating positive memory engrams suppresses depression-like behaviour. *Nature*, *522*(7556), 335–339. https://doi.org/10.1038/nature14514

Reijmers, L. G., Perkins, B. L., Matsuo, N., & Mayford, M. (2007). Localization of a stable neural correlate of associative memory. *Science (New York, N.Y.)*, *317*(5842), 1230–1233. https://doi.org/10.1126/science.1143839

Robins, S. K. (2018). Memory and Optogenetic Intervention: Separating the Engram from the Ecphory. *Philosophy of Science* 85 (5):1078-1089.

Robins, S. K. (2023). The 21st century engram. *WIREs Cognitive Science,* e1653. https:// doi.org/10.1002/wcs.1653

Roediger, H.L. (2007) Encoding. In H.L. Roediger, Y. Dudai & S.M. Fitzpatrick (eds.) *Science of Memory: Concepts* (pp. 17-21). Oxford University Press.

Ross, L.N. (2021). Causal Concepts in Biology: How Pathways Differ from Mechanisms and Why It Matters. *British Journal for the Philosophy of Science* 72 (1):131-158.

Roy, D. S., Arons, A., Mitchell, T. I., Pignatelli, M., Ryan, T. J., & Tonegawa, S. (2016). Memory retrieval by activating engram cells in mouse models of early Alzheimer's disease. *Nature*, *531*(7595), 508–512. https://doi.org/10.1038/nature17172

Roy, D. S., Park, Y. G., Kim, M. E., Zhang, Y., Ogawa, S. K., DiNapoli, N., Gu, X., Cho, J. H., Choi, H., Kamentsky, L., Martin, J., Mosto, O., Aida, T., Chung, K., & Tonegawa, S. (2022). Brain-wide mapping reveals that engrams for a single

memory are distributed across multiple brain regions. *Nature communications*, *13*(1), 1799. https://doi.org/10.1038/s41467-022-29384-4

Rule, M. E., O'Leary, T., & Harvey, C. D. (2019). Causes and consequences of representational drift. *Current opinion in neurobiology*, *58*, 141–147. https://doi.org/10.1016/j.conb.2019.08.005

Ryan, T. J., Roy, D. S., Pignatelli, M., Arons, A., & Tonegawa, S. (2015). Memory. Engram cells retain memory under retrograde amnesia. *Science (New York, N.Y.)*, *348*(6238), 1007–1013. https://doi.org/10.1126/science.aaa5542

Ryan, T. J., Ortega-de San Luis, C., Pezzoli, M., & Sen, S. (2021). Engram cell connectivity: an evolving substrate for information storage. *Current opinion in neurobiology*, *67*, 215–225. https://doi.org/10.1016/j.conb.2021.01.006

Ryan, T. J., & Frankland, P. W. (2022). Forgetting as a form of adaptive engram cell plasticity. *Nature reviews. Neuroscience*, *23*(3), 173–186. https://doi.org/10.1038/s41583-021-00548-3

Ryle, G. (1949). *The Concept of Mind: 60th Anniversary Edition*. Hutchinson & Co.

Sacktor T. C. (2011). How does PKMζ maintain long-term memory?. *Nature reviews. Neuroscience*, *12*(1), 9–15. https://doi.org/10.1038/nrn2949

Saigusa, T. *et al*. (2008) 'Amoebae Anticipate Periodic Events', *Physical Review Letters*. American Physical Society, 100(1), p. 18101. doi: 10.1103/PhysRevLett.100.018101.

Schacter, D. L., & Addis, D. R. (2007). The cognitive neuroscience of constructive memory: remembering the past and imagining the future. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *362*(1481), 773–786. https://doi.org/10.1098/rstb.2007.2087

Schacter, D. (2001). Forgotten ideas, neglected pioneers: Richard Semon and the story of memory. Psychology Press

Semon, R. (1921). *The Mneme*. George Allen & Unwin. (Original work published 1904)

Semon, R. (1923). *Mnemic Psychology*. (B. Duffy, trans.) George Allen & Unwin. (Original work published 1909)

Shea, N. (2018). *Representation in Cognitive Science*. Oxford University Press.

Shpokayte, M., McKissick, O., Guan, X., Yuan, B., Rahsepar, B., Fernandez, F. R., Ruesch, E., Grella, S. L., White, J. A., Liu, X. S., & Ramirez, S. (2022). Hippocampal cells segregate positive and negative engrams. *Communications biology*, *5*(1), 1009. https://doi.org/10.1038/s42003-022-03906-8

Simon, H. A. (2002). Near decomposability and the speed of evolution. *Industrial and Corporate Change, 11*(3), 587-600.

Sterling, L. & Laughlin, S. (2015). *Principles of Neural Design*. MIT Press.

Stone, J.V. (2022). *Information Theory: A Tutorial Introduction*. (2nd ed). Sebtel Press.

Sun, X., Bernstein, M. J., Meng, M., Rao, S., Sørensen, A. T., Yao, L., Zhang, X., Anikeeva, P. O., & Lin, Y. (2020). Functionally Distinct Neuronal Ensembles within the Memory Engram. *Cell*, *181*(2), 410–423.e17. https://doi.org/10.1016/j.cell.2020.02.055

Tanaka, K. Z., Pevzner, A., Hamidi, A. B., Nakazawa, Y., Graham, J., & Wiltgen, B. J. (2014). Cortical representations are reinstated by the hippocampus during memory retrieval. *Neuron*, *84*(2), 347–354. https://doi.org/10.1016/j.neuron.2014.09.037

Tinbergen, N. (1963). "On aims and methods of ethology." Zeitschrift für Tierpsychologie 20:410-433.

Tonegawa, S., Liu, X., Ramirez, S., & Redondo, R. (2015a). Memory Engram Cells Have Come of Age. *Neuron*, *87*(5), 918–931. https://doi.org/10.1016/j.neuron.2015.08.002

Tonegawa, S., Pignatelli, M., Roy, D. S., & Ryan, T. J. (2015b). Memory engram storage and retrieval. *Current opinion in neurobiology*, *35*, 101–109. https://doi.org/10.1016/j.conb.2015.07.009

Tsien R. Y. (2013). Very long-term memories may be stored in the pattern of holes in the perineuronal net. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(30), 12456–12461. https://doi.org/10.1073/pnas.1310158110

Vetere, G., Kenney, J. W., Tran, L. M., Xia, F., Steadman, P. E., Parkinson, J., Josselyn, S. A., & Frankland, P. W. (2017). Chemogenetic Interrogation of a Brain-wide Fear Memory Network in Mice. *Neuron*, *94*(2), 363–374.e4. https://doi.org/10.1016/j.neuron.2017.03.037

Wang, C., Yue, H., Hu, Z., Shen, Y., Ma, J., Li, J., Wang, X. D., Wang, L., Sun, B., Shi, P., Wang, L., & Gu, Y. (2020). Microglia mediate forgetting via complement-dependent synaptic elimination. *Science (New York, N.Y.)*, *367*(6478), 688–694. https://doi.org/10.1126/science.aaz2288

Wang, L., Gillis-Smith, S., Peng, Y., Zhang, J., Chen, X., Salzman, C. D., Ryba, N. J. P., & Zuker, C. S. (2018). The coding of valence and identity in the mammalian taste system. *Nature*, *558*(7708), 127–131. https://doi.org/10.1038/s41586-018-0165-4

Wheeler, A. L., Teixeira, C. M., Wang, A. H., Xiong, X., Kovacevic, N., Lerch, J. P., McIntosh, A. R., Parkinson, J., & Frankland, P. W. (2013). Identification of a functional connectome for long-term fear memory in mice. *PLoS computational biology*, *9*(1), e1002853. https://doi.org/10.1371/journal.pcbi.1002853